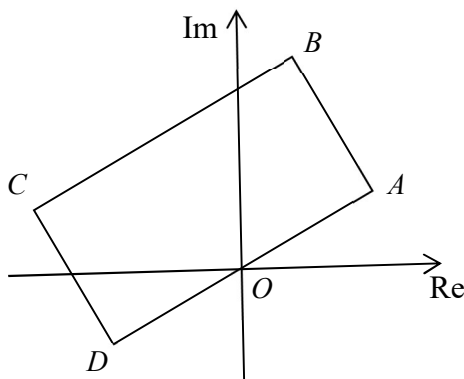




H2 Math Year 6 Preliminary Examination Paper 2: Solutions with comments

- 1 (a) The complex number w is such that $w^2 + ww^* = 6 + 2\sqrt{3}i$. Find exactly the possible values of w , giving your answer in the form $x + iy$ where x and y are real numbers. [3]
- (b) The complex numbers found in part (a) are represented by points A and D in an Argand diagram as shown below. Points B and C represent complex numbers b and c respectively.



Given that $ABCD$ is a rectangle with $AD = 2AB$, find complex numbers b and c , giving your answers in the form $u + iv$, where u and v are real numbers. [3]

- (c) By finding $\arg(b)$, find the exact value of $\tan \frac{5\pi}{12}$. [2]

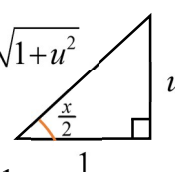
Solution		Comments
(a) [3]	$w^2 + ww^* = 6 + 2\sqrt{3}i$ Let $w = x + iy$, then $(x + iy)^2 + (x + iy)(x - iy) = 6 + 2\sqrt{3}i$ $x^2 + 2ixy - y^2 + x^2 + y^2 = 6 + 2\sqrt{3}i$ $2x^2 + 2ixy = 6 + 2\sqrt{3}i$ Comparing real part: $2x^2 = 6 \Rightarrow x^2 = 3 \Rightarrow x = \pm\sqrt{3}$ Comparing imaginary part: $2xy = 2\sqrt{3} \Rightarrow y = \pm 1$ $\therefore w = \sqrt{3} + i$ or $w = -\sqrt{3} - i$	Most students handled this part well.

(b) [3]	From the Argand diagram, point A represents $\sqrt{3} + i$ and point D represents $-\sqrt{3} - i$. Let the midpoint of BC be E, then E is represented by the complex number $i(\sqrt{3} + i) = -1 + \sqrt{3}i.$ Hence, $\begin{aligned} b &= \sqrt{3} + i + (-1 + \sqrt{3}i) \\ &= (\sqrt{3} - 1) + (\sqrt{3} + 1)i \\ c &= -\sqrt{3} - i + (-1 + \sqrt{3}i) \\ &= (-\sqrt{3} - 1) + (\sqrt{3} - 1)i \end{aligned}$	This part challenges many students. Useful ideas include multiplying a complex number by i and the geometric effects of adding/subtracting complex numbers. Note that we rotate $\overrightarrow{OA}$ about O by $\frac{\pi}{2}$ anticlockwise to obtain $\overrightarrow{OE}$. Also, $\overrightarrow{OB} = \overrightarrow{OA} + \overrightarrow{OE}$ and $\overrightarrow{OC} = \overrightarrow{OD} + \overrightarrow{OE}$
(c) [2]	Since $OA = AB$, $\angle BOA = \frac{\pi}{4}$. Also, $\arg(a) = \tan^{-1}\left(\frac{1}{\sqrt{3}}\right) = \frac{\pi}{6}$. Hence, $\begin{aligned} \arg(b) &= \arg(a) + \angle BOA \\ &= \frac{\pi}{6} + \frac{\pi}{4} \\ &= \frac{5\pi}{12} \end{aligned}$ $\therefore \tan \frac{5\pi}{12} = \frac{\operatorname{Im}(b)}{\operatorname{Re}(b)} = \frac{\sqrt{3}+1}{\sqrt{3}-1} \times \frac{\sqrt{3}+1}{\sqrt{3}+1} = \frac{3+2\sqrt{3}+1}{3-1} = 2+\sqrt{3}$	A key element of this part is explaining why the exact value of $\arg(b)$ is $\frac{5\pi}{12}$, enabling us to determine the exact value of $\tan \frac{5\pi}{12}$ from the real and imaginary parts of b found in part (b). Thus, numerical verification using the GC alone is insufficient.

2 (a) Find $\int \frac{2x+3}{x^2+2x-3} dx$. [3]

(b) Using the substitution $u = \tan \frac{x}{2}$, show that $\int \frac{1}{2+\cos x} dx = \int \frac{2}{3+u^2} du$. Hence find the exact value of $\int_0^{\frac{\pi}{2}} \frac{1}{2+\cos x} dx$. [6]

Solution	Comments
(a) [3] $\int \frac{2x+3}{x^2+2x-3} dx$ $= \int \frac{3}{4} \left(\frac{1}{x+3} \right) + \frac{5}{4} \left(\frac{1}{x-1} \right) dx$ $= \frac{3}{4} \ln x+3 + \frac{5}{4} \ln x-1 + c$ OR $\int \frac{2x+3}{x^2+2x-3} dx$ $= \int \frac{2x+2}{x^2+2x-3} + \frac{1}{x^2+2x-3} dx$ $= \int \frac{2x+2}{x^2+2x-3} + \frac{1}{(x+1)^2-4} dx$ $= \ln x^2+2x-3 + \frac{1}{2(2)} \ln \left \frac{(x+1)-2}{(x+1)+2} \right + c$ $= \ln x^2+2x-3 + \frac{1}{4} \ln \left \frac{x-1}{x+3} \right + c$	Generally well done, except that some missed out the modulus sign for $\ln$ or arbitrary constant, c in the final line of the answer.

(b) [6] <u>Method 1:</u> $\int \frac{1}{2 + \cos x} dx$ $= \int \frac{1}{2 + \left(2 \cos^2 \frac{x}{2} - 1\right)} dx$ $= \int \frac{1}{1 + 2 \cos^2 \frac{x}{2}} dx$ $= \int \frac{1}{1 + 2 \left(\frac{1}{1 + u^2}\right)} \cdot \frac{2}{1 + u^2} du$ $= \int \frac{2}{(1 + u^2) + 2} du$ $= \int \frac{2}{3 + u^2} du \text{ (shown)}$ <u>Method 2:</u> $\int \frac{1}{2 + \cos x} dx$ $= \int \frac{1}{2 + \frac{1 - u^2}{1 + u^2}} \cdot \frac{2}{1 + u^2} du$ $= \int \frac{1}{\frac{2(1 + u^2) + (1 - u^2)}{1 + u^2}} \cdot \frac{2}{1 + u^2} du$ $= \int \frac{2}{3 + u^2} du \text{ (shown)}$ $\int_0^{\frac{\pi}{2}} \frac{1}{2 + \cos x} dx$ $= \int_0^1 \frac{2}{3 + u^2} du$ $= 2 \left[\frac{1}{\sqrt{3}} \tan^{-1} \frac{u}{\sqrt{3}} \right]_0^1$ $= \frac{2}{\sqrt{3}} \tan^{-1} \frac{1}{\sqrt{3}}$ $= \frac{2}{\sqrt{3}} \left(\frac{\pi}{6} \right) = \frac{\pi}{3\sqrt{3}} = \frac{\sqrt{3}\pi}{9}$	Let $u = \tan \frac{x}{2}$ $\frac{du}{dx} = \frac{1}{2} \sec^2 \frac{x}{2}$ $= \frac{1}{2} \left(1 + \tan^2 \frac{x}{2} \right)$ $= \frac{1}{2} (1 + u^2)$ $dx = \frac{2}{1 + u^2} du$  $u = \tan \frac{x}{2}$ $\therefore \cos \frac{x}{2} = \frac{1}{\sqrt{1 + u^2}}$ $\cos x = 2 \cos^2 \frac{x}{2} - 1$ $= \frac{2}{1 + u^2} - 1$ $= \frac{2 - (1 + u^2)}{1 + u^2} = \frac{1 - u^2}{1 + u^2}$ Since $u = \tan \frac{x}{2}$ When $x = \frac{\pi}{2}$, $u = \tan \frac{\pi}{4} = 1$ When $x = 0$, $u = 0$	For the “show” part, most were able to differentiate the given substitution and made use of trigo identities to complete the proof. However, do note that as this is a “show” question, clear working is required to earn full credit. This last part was generally well done, although some did not evaluate $\tan^{-1} \frac{1}{\sqrt{3}}$ which is a special angle.
---	---	---

3 The function f is given by

$$f : x \mapsto a - e^{|x-a|}, \quad x \in \mathbb{R},$$

where a is a positive constant.

It is given that the range of f is $(-\infty, k]$, where $k > 0$.

(a) Find the range of values of a . [2]

(b) Show that f^{-1} does not exist. [2]

(c) If the domain of f is restricted to $(-\infty, m]$, state the largest value of m in terms of a for which f^{-1} exists. [1]

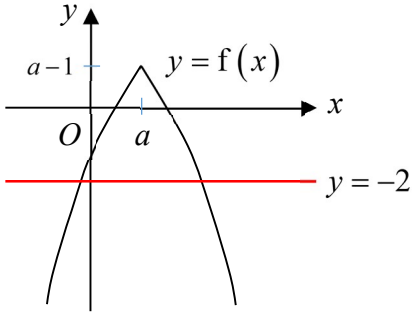
(d) Using your answer found in part (c), find f^{-1} in similar form. [3]

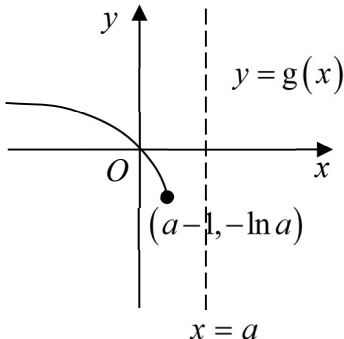
The function g is given by

$$g : x \mapsto \ln\left(1 - \frac{x}{a}\right), \quad x < a.$$

(e) Show that the composite function gf exists and find the range of gf in terms of a . [3]

Solution		Comments
(a) [2]	Since $x - a \geq 0$ for all x, $e^{ x-a } \geq 1$ $-e^{ x-a } \leq -1$ $a - e^{ x-a } \leq a - 1 = k$ Since $k > 0$, $a - 1 > 0$ $a > 1$	The objective of this question is to find the range of values of a given that the maximum y-value is positive. Students may either use inequalities or a graphical approach to determine k in terms of a. Note that this question is not about solving x.
(b) [2]	Method 1: Since $f(a+1) = f(a-1) = a - e$, f is not one-one. Hence, f^{-1} does not exist.	Note that method 1 is more time efficient than method 2!

	Method 2:  Since the horizontal line $y = -2$ cuts the graph of $y = f(x)$ at more than one point, f is not one-one. Hence, f^{-1} does not exist.	
(c) [1]	Largest $m = a$.	
(d) [3]	Let $y = a - e^{ x-a }$. For $x \leq a$, $y = a - e^{-(x-a)}$ $e^{a-x} = a - y$ $a - x = \ln(a - y)$ $x = a - \ln(a - y)$ $\therefore f^{-1} : x \mapsto a - \ln(a - x), x \leq a - 1$	It would be worthwhile to resolve the modulus first by considering the domain of f, before making x the subject. Do note that a function cannot take on 2 y-values for 1 x-value – hence the rule should not include a $\pm$ sign. Students should also be mindful to present the answer in similar form as required by the question.
(e) [3]	$R_f = (-\infty, a - 1]$, $D_g = (-\infty, a)$ Since $R_f \subseteq D_g$, gf exists.	Students should state both R_f and D_g in interval notation before concluding that $R_f \subseteq D_g$. It is also important for students to show clearly that the upper bound of R_f is indeed smaller than that of D_g.

	Sketching g with restricted domain,  $D_f = (-\infty, a] \rightarrow R_f = (-\infty, a-1] \rightarrow R_{gf}$ $g(a-1) = \ln\left(1 - \frac{a-1}{a}\right) = \ln\left(\frac{1}{a}\right) = -\ln a$ $\therefore R_{gf} = [-\ln a, \infty)$ OR $gf(x) = \ln\left(1 - \frac{a - e^{ x-a }}{a}\right)$ $= \ln\left(\frac{e^{ x-a }}{a}\right)$ $= x-a - \ln a, \quad x \in \mathbb{R}$ Since $x-a \geq 0$, $ x-a - \ln a \geq -\ln a$ $\therefore R_{gf} = [-\ln a, \infty)$	Students who attempted mapping method (together with the sketch of $y=g(x)$) tend to achieve greater success in arriving at the correct answer. Students who used the direct method in finding rule of gf need to be mindful that they should use D_{gf} to find R_{gf}, rather than using R_f to find it.
--	---	---

4 Do not use a calculator in answering this question.

- (a)**
- The curve
- $y = f(x)$
- passes through the origin and has gradient given by

$$\frac{dy}{dx} = \frac{x+y}{1+xy}.$$

- (i)**
- Show that
- $(1+xy)\frac{d^2y}{dx^2} + x\left(\frac{dy}{dx}\right)^2 + (y-1)\frac{dy}{dx} = 1$
- . [2]

- (ii)**
- By further differentiation of the result in part
- (a)(i)**
- , show that

$$(1+xy)\frac{d^3y}{dx^3} + p(x)\left(\frac{dy}{dx}\right)\left(\frac{d^2y}{dx^2}\right) + q(y)\frac{d^2y}{dx^2} + r\left(\frac{dy}{dx}\right)^2 = 0,$$

where the functions $p(x)$, $q(y)$ and the constant r are to be determined. [3]

- (iii)**
- Hence find the Maclaurin series for
- y
- up to and including the term in
- x^3
- . [2]

- (b)**
- The equation

$$2\sqrt{2}\cos\left(x - \frac{\pi}{4}\right) + 3\sin 2x - 3 = 0$$

has a root θ , which is close to zero.

- (i)**
- Show that
- $\theta^2 + a\theta + b \approx 0$
- , where
- a
- and
- b
- are integers to be determined. [3]

- (ii)**
- Hence find an approximation for
- θ
- in surd form. [2]

Solution	Comments
(a)(i) [2] $\frac{dy}{dx} = \frac{x+y}{1+xy}$ $(1+xy)\frac{dy}{dx} = x+y$ Differentiating w.r.t x, $(1+xy)\frac{d^2y}{dx^2} + \left(x\frac{dy}{dx} + y\right)\frac{dy}{dx} = 1 + \frac{dy}{dx}$ $(1+xy)\frac{d^2y}{dx^2} + x\left(\frac{dy}{dx}\right)^2 + (y-1)\frac{dy}{dx} = 1 \quad (\text{shown})$	Note that this is a “show” question, so it is important not to skip steps in your working. Students who first changed the equation to $(1+xy)\frac{dy}{dx} = x+y$ before differentiating with respect

		to x were mostly able to show the given result. Students who did not make this change either applied product rule on $(x+y)(1+xy)^{-1}$ or quotient rule on $\frac{x+y}{1+xy}$ and ended up taking longer to show the given result and many gave up eventually.
(a)(ii) [3]	Differentiating again w.r.t x, $(1+xy)\frac{d^3y}{dx^3} + \left(x\frac{dy}{dx} + y\right)\frac{d^2y}{dx^2} + \left(\frac{dy}{dx}\right)^2$ $+ x(2)\left(\frac{dy}{dx}\right)\left(\frac{d^2y}{dx^2}\right) + (y-1)\frac{d^2y}{dx^2} + \left(\frac{dy}{dx}\right)^2 = 0$ $(1+xy)\frac{d^3y}{dx^3} + (3x)\left(\frac{dy}{dx}\right)\left(\frac{d^2y}{dx^2}\right) + (2y-1)\frac{d^2y}{dx^2} + 2\left(\frac{dy}{dx}\right)^2 = 0$ where $p(x) = 3x$, $q(y) = 2y-1$, $r = 2$	Note that $p(x)$, $q(y)$ and r need to be stated explicitly to gain full credit. Slips with differentiation continue to show up here especially with differentiating the $(1+xy)$ term and the constant term on the RHS of the equation. Some students also had difficulty collecting the like terms together due to the many terms involved.
(a) (iii) [2]	When $x = 0$, $y = 0$ (since curve passes through the origin) , $\frac{dy}{dx} = 0, \quad \frac{d^2y}{dx^2} = 1, \quad \frac{d^3y}{dx^3} = 1.$ By Maclaurin expansion, $y = 0 + 0x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$ $= \frac{x^2}{2} + \frac{x^3}{6} + \dots$	Note that $y = 0$ when $x = 0$ since the curve passes through the origin!

(b)(i) [3]	$2\sqrt{2} \cos\left(x - \frac{\pi}{4}\right) + 3 \sin 2x - 3 = 0$ $2\sqrt{2} \left(\cos x \cos \frac{\pi}{4} + \sin x \sin \frac{\pi}{4} \right) + 3 \sin 2x - 3 = 0$ $2\sqrt{2} \left(\frac{1}{\sqrt{2}} \right) (\cos x + \sin x) + 3 \sin 2x - 3 = 0$ $2(\cos x + \sin x) + 3 \sin 2x - 3 = 0$ Since θ is a root of the equation, $2(\cos \theta + \sin \theta) + 3 \sin 2\theta - 3 = 0$ Since θ is close to zero, applying the standard series for sine and cosine from MF27, $2\left(1 - \frac{\theta^2}{2!} + \theta\right) + 3(2\theta) - 3 \approx 0$ $-\theta^2 + 8\theta - 1 \approx 0$ $\theta^2 - 8\theta + 1 \approx 0$ where $a = -8$ and $b = 1$.	The correct method is to use the compound angle formula to separate the small angle θ from $\frac{\pi}{4}$ before using the small angle approximation. Many students were unable to make much progress in this part due to the mistake in assuming that the angle $\left(\theta - \frac{\pi}{4}\right)$ is “small” and wrote $\cos\left(\theta - \frac{\pi}{4}\right) \approx 1 - \frac{\left(\theta - \frac{\pi}{4}\right)^2}{2}$ which is wrong! Note also that $\sin 2\theta \approx 2\theta$ when θ is small, so there is no need to expand the $\sin 2\theta$ term. Integer values of a and b need to be stated explicitly to gain full credit.
(b)(ii) [2]	$\theta^2 - 8\theta + 1 \approx 0$ $\theta \approx \frac{8 \pm \sqrt{8^2 - 4(1)(1)}}{2(1)}$ $= \frac{8 \pm \sqrt{60}}{2}$ $= 4 \pm \sqrt{15}$ Since θ is close to zero, $\theta \approx 4 - \sqrt{15}$.	The two common ways to solve the equation are using the quadratic formula or by completing the square. As the answer is required in surd form, the calculator should not be used. Common mistake: Not rejecting the θ value which is not close to zero. (Note: Both values of θ are positive)

- 5 A random variable X has probability distribution given by

$$P(X = x) = \begin{cases} kx^2 & \text{for } x = 1, 2, 3, \\ k(7 - x) & \text{for } x = 4, 5, 6, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a positive constant.

(a) Show that $k = \frac{1}{20}$. [2]

(b) Find the values of $E(X)$ and $\text{Var}(X)$. [3]

Another random variable Y is defined by $Y = 3X_1 - X_2$, where X_1 and X_2 are independent observations of X .

(c) Find $\text{Var}(Y)$. [2]

Solution		Comments														
(a) [2]	<table border="1"><tr><td>x</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>$P(X = x)$</td><td>k</td><td>$4k$</td><td>$9k$</td><td>$3k$</td><td>$2k$</td><td>k</td></tr></table> Since $\sum P(X = x) = 1$, $k(1 + 4 + 9 + 3 + 2 + 1) = 1$ $20k = 1$ $k = \frac{1}{20}$ (shown)	x	1	2	3	4	5	6	$P(X = x)$	k	$4k$	$9k$	$3k$	$2k$	k	This part was well done.
x	1	2	3	4	5	6										
$P(X = x)$	k	$4k$	$9k$	$3k$	$2k$	k										
(b) [3]	$E(X) = \frac{1}{20}(1 + 8 + 27 + 12 + 10 + 6) = \frac{64}{20} = \frac{16}{5} = 3.2$ $E(X^2) = \frac{1}{20}[1 + 2^2(4) + 3^2(9) + 4^2(3) + 5^2(2) + 6^2(1)]$ $= \frac{232}{20} = \frac{58}{5} = 11.6$ $\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{58}{5} - \left(\frac{16}{5}\right)^2 = \frac{34}{25} = 1.36$	Some students used the GC to obtain the values of $E(X)$ and $\text{Var}(X)$ directly which is permissible here since the question did not ask for “exact” values nor was this a show question.														

		For students who square the standard deviation found in the GC to find $\text{Var}(X)$, do note that a 5 s.f. value should be used to obtain a final 3 s.f. answer.
(c) [2]	$\begin{aligned}\text{Var}(Y) &= \text{Var}(3X_1 - X_2) \\ &= 3^2 \text{Var}(X_1) + \text{Var}(X_2), \\ &\quad \text{since } X_1 \text{ and } X_2 \text{ are independent r.v.} \\ &= 9\text{Var}(X) + \text{Var}(X) \\ &= 10\text{Var}(X) \\ &= \frac{68}{5} \quad (\text{OR } 13.6)\end{aligned}$	Common mistakes include writing $\text{Var}(3X_1 - X_2)$ as $3\text{Var}(X_1) - \text{Var}(X_2)$.

- 6 (a) For a special opening act in a talent show, 8 performers (3 singers, 3 dancers and 2 musicians) are to stand in a circle on stage. Find the number of possible arrangements if
- (i) there are no restrictions, [1]
- (ii) all performers of the same role must stand together. [2]
- (b) For the main performance, 12 performers (3 singers, 5 dancers and 4 musicians) are arranged in a line on stage such that no two musicians are adjacent. Find the number of possible arrangements. [2]
- (c) 7 performers are selected from 15 performers (6 singers, 5 dancers and 4 musicians) for the closing act. Find the number of possible selections if there must be at least 2 performers from each role. [3]

Solution		Comments
(a)(i) [1]	Number of arrangements = $(8-1)! = 7! = 5040$	Many students failed to take circular permutation into account.
(a)(ii) [2]	Treat each group as a single unit: 3 units arranged in a circle = $(3-1)! = 2!$ Internal arrangements within each group: $3! \times 3! \times 2!$	
	Number of arrangements = $2! \times 3! \times 3! \times 2! = 144$	

(b) [2]	First arrange the 8 non-musicians: $8!$ This creates 9 gaps where musicians can be placed. Choose 4 gaps from 9 and arrange musicians: ${}^9C_4 \times 4!$ Number of arrangements = $8! \times {}^9C_4 \times 4! = 121\,927\,680$	Since the musicians are distinct individuals, accounting for their permutations requires multiplying by $4!$.																
(c) [3]	Possible combinations are: <table><tr><th># of Singers</th><th># of Dancers</th><th># of Musicians</th><th># of Ways</th></tr><tr><td>2</td><td>2</td><td>3</td><td>${}^6C_2 \times {}^5C_2 \times {}^4C_3 = 600$</td></tr><tr><td>2</td><td>3</td><td>2</td><td>${}^6C_2 \times {}^5C_3 \times {}^4C_2 = 900$</td></tr><tr><td>3</td><td>2</td><td>2</td><td>${}^6C_3 \times {}^5C_2 \times {}^4C_2 = 1200$</td></tr></table> Total number of ways = $600 + 900 + 1200 = 2700$	# of Singers	# of Dancers	# of Musicians	# of Ways	2	2	3	${}^6C_2 \times {}^5C_2 \times {}^4C_3 = 600$	2	3	2	${}^6C_2 \times {}^5C_3 \times {}^4C_2 = 900$	3	2	2	${}^6C_3 \times {}^5C_2 \times {}^4C_2 = 1200$	Many students considered the cases correctly. However, some calculated the number of ways for each case incorrectly by introducing unnecessary permutations. A large number of students evaluated the expression as ${}^6C_2 \times {}^5C_2 \times {}^4C_2 \times {}^9C_1$, which is incorrect due to double-counting. Please refer to Example 11(ii) in Chapter S1A Lecture Notes for a detailed explanation. Students who attempted the complement method were unsuccessful, as none addressed the double-counting issue correctly. In general, for this type of question, the direct method is employed. We consider the complement method only if there are many cases involved in the direct method.
# of Singers	# of Dancers	# of Musicians	# of Ways															
2	2	3	${}^6C_2 \times {}^5C_2 \times {}^4C_3 = 600$															
2	3	2	${}^6C_2 \times {}^5C_3 \times {}^4C_2 = 900$															
3	2	2	${}^6C_3 \times {}^5C_2 \times {}^4C_2 = 1200$															

7 In this question, you should state the parameters of any distributions you use.

In a running competition, the time taken by members from the Sports Recreational Club to complete a 20 km race is normally distributed with a mean of 105 minutes and a standard deviation of 12 minutes.

- (a) Find the probability that a randomly chosen Sports Recreational Club member finishes the 20 km race in less than 100 minutes. [1]
- (b) Find the time that is exceeded by 10% of the Sports Recreational Club members. [1]

Members from the Swift Runners Club competed in the same 20 km race. After analysing their results, it was found that 36% of them completed the race in less than 95 minutes and 8% of them completed the race in more than 125 minutes. You may assume that the Swift Runners Club members' running times can be modelled using a normal distribution with mean μ and standard deviation σ .

- (c) Determine the values of μ and σ . [4]
- (d) Using the values of the parameters found in part (c), find the probability that the total running time of 2 randomly chosen Swift Runners Club members is less than 1.8 times of the running time of a randomly chosen Sports Recreational Club member.

State an assumption you have made in your calculation. [4]

Solution		Comments
(a) [1]	Let X be the time in minutes taken by a Sports Recreational Club member to complete a 20 km race. $X \sim N(105, 12^2)$ $P(X < 100) = 0.338$ (3 s.f)	This part was well done except that many students did not define the random variable properly.
(b) [1]	Let the required time be t minutes. $P(X > t) = 0.1$ Using GC, $P(X > 120.38) = 0.1$ $\therefore t = 120$ minutes (3 s.f)	This part was also well done except for a few students who misinterpreted the question as $X < t$.
(c) [4]	Let Y be the time in minutes taken by a Swift Runners Club member to complete a 20 km race. $Y \sim N(\mu, \sigma^2)$	Interestingly, most students defined the random variable properly when they reached this part. Sadly, they did not

	$P(Y < 95) = 0.36$ $P\left(Z < \frac{95 - \mu}{\sigma}\right) = 0.36$ $\frac{95 - \mu}{\sigma} = -0.35846$ $95 - \mu = -0.35846\sigma \text{ -----(1)}$ $P(Y > 125) = 0.08$ $P\left(Z > \frac{125 - \mu}{\sigma}\right) = 0.08$ $\frac{125 - \mu}{\sigma} = 1.4051$ $125 - \mu = 1.4051\sigma \text{ -----(2)}$ $(2) - (1)$ $1.76356\sigma = 30$ $\sigma = 17.011 = 17.0 \text{ (3 s.f.)}$ $\mu = 125 - (1.4051)(17.011) = 101.10 = 101 \text{ (3 s.f.)}$	feel the need to amend their answer to part (a). A good number of students were able to standardize the variable and set up the 2 equations correctly. A small number standardize wrongly, writing $\mu - 95$ and $\mu - 125$ or dividing by σ^2. The most efficient and accurate way to solve the equations was via GC as some students made careless addition mistakes or rounding off errors while solving by hand.
(d) [4]	Required Probability = $P(Y_1 + Y_2 < 1.8X)$ Let $W = Y_1 + Y_2 - 1.8X$ $E(W) = 2E(Y) - 1.8E(X) = 2(101.1) - 1.8(105) = 13.2$ $\text{Var}(W) = 2\text{Var}(Y) + 1.8^2 \text{Var}(X)$ $= 2(17.011^2) + 1.8^2(12^2)$ $= 1045.3$ $W \sim N(13.2, 1045.3)$ $P(W < 0) = 0.342 \text{ (3 s.f.)}$ Assumption: The time taken by a club member to complete a 20 km race is independent of the time taken by another club member to complete a 20 km race.	A number of students forgot that values used in intermediate work should be accurate up to 5 s.f. as they misunderstood the question and used the 3 s.f. values from part (c) instead. On the assumption required, many students have forgotten that it is the independence of the different normal variables that is needed here. A good number confused this with the independence of events and came up with phrases like “the event of the timings ...”. Others just assumed that the

		distribution is normal, reminiscent of questions on Hypothesis Testing.
--	--	---

- 8** A light source is placed at varying distances from a sensor. The intensity of light measured by the sensor is recorded. The results are as shown in the table below; the distance, x , is in cm and the intensity of light, y , is in lux.

x	10	15	20	30	40	50	70	100
y	112.2	60.1	42.5	28.8	24.2	22.0	20.1	19.2

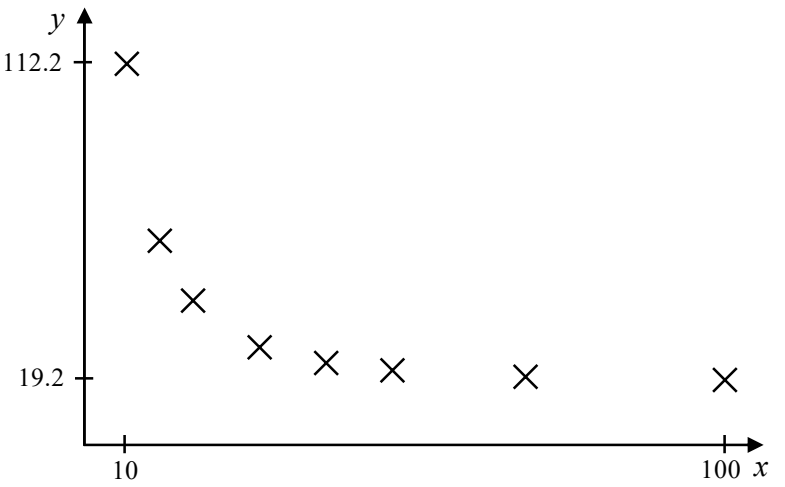
- (a) Sketch the scatter diagram for the data. Hence, explain why a linear model $y = a + bx$ is likely to be inappropriate. [3]

- (b) The following two models are proposed, where c , d , e and f are constants:

$$y = c + \frac{d}{x} \qquad y = e + f \ln x$$

Calculate the product moment correlation coefficient for both models. State, with a reason, why the model $y = c + \frac{d}{x}$ is the more appropriate model. [3]

- (c) Find the regression equation for the model $y = c + \frac{d}{x}$, giving your constants c and d correct to 2 decimal places. [1]
- (d) Use your equation found in part (c) to estimate the light intensity when $x = 120$. Comment on the reliability of this estimate. [2]
- (e) In this context, what does the value of c in your regression line in part (c) represent? [1]
- (f) Suggest a physical interpretation for the constant c . [1]

Solution	Comments
(a) [3]  From the scatter diagram, as x increases, y decreases at a decreasing rate, so using a linear model is likely to be inappropriate.	The points for $x = 40, 50, 70$ and 100 should be close to the horizontal as their difference in the y-value is very small as compared to the distance between 19.2 and 112.2. Some used the product moment correlation coefficient to explain the inappropriateness of the linear model. However, the question stated “Hence”, and so one will need to explain the inappropriateness using the scatter diagram.
(b) [3] Model (y on $\frac{1}{x}$): $r_1 \approx 0.97308$ (5 s.f.) = 0.973 (3 s.f.) Model (y on $\ln x$): $r_2 \approx -0.85879$ (5 s.f.) = -0.859 (3 s.f.) Since the value of r_1 is closer to 1 compared to r_2, indicating a much stronger linear relationship between y and $\frac{1}{x}$, $y = c + \frac{d}{x}$ is the more appropriate model.	A common mistake is using the values of r to comment, instead of using the values of r.
(c) [1] Equation of the regression line is $y \approx 0.86938 + \frac{1008.9551}{x}$ $y = 0.87 + \frac{1008.96}{x} \quad (2 \text{ d.p.})$	A few did not follow the instruction of giving the answers to 2 d.p.

(d) [2]	The light intensity at a distance 120 cm away from the light source is $y \approx 0.86938 + \frac{1008.9551}{120} = 9.28 \text{ lux (3 s.f.)}$ The value $x = 120$ lies outside the given range of values of x, which is $[10, 100]$ and hence extrapolation is used to obtain the estimate. The estimate would be unreliable as the linear model may not be applicable outside the given range of data.	A more accurate answer is obtained when using more accurate coefficients found in part (c). Quite a few did not state the given range of x which is $[10, 100]$ in their answer.
(e) [1]	The value of $c = 0.87$ lux represents the theoretical light intensity picked up by the sensor when the sensor is at a very large distance from the light source.	c is the value obtained when $x \rightarrow \infty$. Do note the need to give the answer in the context of the question (light intensity, distance, light source)
(f) [1]	We expect $y \rightarrow 0$ as $x \rightarrow \infty$. Thus, the constant c might represent ambient or background light in the environment or systematic error due to the faulty sensor.	Parts (e) and (f) are distinct. Part (e) asks what c represents <i>in the regression model</i> in the context of the question, while part (f) asks for a <i>physical/real-world</i> interpretation, and students should thus suggest something physically meaningful.

- 9 A quality control inspector at an electronics factory tests components to see if any are defective. The inspector randomly selects 30 components for testing in a day. Historical data shows that 0.8% of the components are defective.

- (a)** State, in context, two assumptions needed for the number of defective components found each day to be well-modelled by a binomial distribution. [2]

Assume now that the number of defective components found each day has a binomial distribution.

- (b) Find the probability that at most 1 component is found to be defective. [1]
- (c) Find the probability that the first defective component is the fourth component tested and a total of 3 defective components are found in a day. [2]
- (d) Given that the inspector finds at least 1 defective component in a day, find the probability that she finds less than 3 defective components. [2]
- (e) Find the probability that, in a randomly chosen 5-day working week, at most 1 component is found to be defective on more than 3 days. [2]

The owner of the factory decides to expand his business by producing memory chips. On average, 5% of the memory chips are defective. Another quality control inspector was tasked to randomly select 100 memory chips for testing in a day. You may assume that the number of defective memory chips found each day has a binomial distribution.

- (f) The inspector records down the number of defective memory chips she found each day. Estimate the probability that the average number of defective memory chips found on 50 randomly chosen days is between 4.5 and 5.5. [3]

Solution		Comments
(a) [2]	The probability of a component being defective remains constant at 0.008 for all components. The event of a component being defective is independent of the event of another component being defective.	Generally, most were able to state the assumptions in the context of the question. A small minority still used the word “probability” instead of the word “event” for the assumption on independence. There was also a handful that stated the event as “selecting/choosing a defective component” instead of “the component being defective”.
(b) [1]	Let X be the number of defective components out of 30 components. $X \sim B(30, 0.008)$ $P(X \leq 1) \approx 0.97600 = 0.976$ (3 s.f.)	This part was well done but some were careless and wrote p as 0.08 instead of 0.008 which affected the rest of question too.

(c) [2]	P(first defective component is the fourth component tested and a total of 3 defective components are found in a day) $= \left[(0.992)^3 (0.008) \right] {}^{26}C_2 (0.008)^2 (0.992)^{24}$ ≈ 0.00013396 $= 0.000134 \text{ (3 s.f.)}$	Many were unable to do this part correctly as they consider the same distribution used in part (b) where the number of components is 30 but in this part, we first find the P(1st defective component is the 4th component tested) followed by the P(finding 2 defective components in the remaining 26 components).
(d) [2]	$P(X < 3 X \geq 1)$ $= \frac{P((X \leq 2) \cap (X \geq 1))}{P(X \geq 1)}$ $= \frac{P(1 \leq X \leq 2)}{P(X \geq 1)}$ $= \frac{P(X=1) + P(X=2)}{1 - P(X=0)} \text{ or } \frac{P(X \leq 2) - P(X=0)}{1 - P(X=0)}$ ≈ 0.99174 $= 0.992 \text{ (3 s.f.)}$	Some were careless in stating the probability and wrote it as $P(X \geq 1 X < 3)$ instead. Note that $P((X \leq 2) \cap (X \geq 1))$ $= P(X \leq 2) \times P(X \geq 1)$ is not correct as the two events are not independent
(e) [2]	Let Y be number of days with at most 1 defective component, out of 5 days. Then $Y \sim B(5, 0.97600)$ $P(Y > 3) = 1 - P(Y \leq 3)$ $\approx 0.99451 = 0.995 \text{ (3 s.f.)}$ OR $P(Y > 3) = P(Y = 4) + P(Y = 5)$ $\approx 0.99451 = 0.995 \text{ (3 s.f.)}$ OR Required probability $= {}^5C_4 (0.97600)^4 (1 - 0.97600) + (0.97600)^5$ $\approx 0.99451 = 0.995 \text{ (3 s.f.)}$	A common mistake: $P(Y > 3) = 1 - P(Y \leq 2)$ Do exercise greater care when working out the complement!

(f) [3]	Let W be the number of defective memory chips out of 100 memory chips. Then $W \sim B(100, 0.05)$ $E(W) = 100 \times 0.05 = 5$ $\text{Var}(W) = 10 \times 0.05 \times 0.95 = 4.75$ Since $n = 50$ is large, by Central Limit Theorem, $\bar{W} \sim N\left(5, \frac{4.75}{50}\right)$ approximately. $P(4.5 \leq \bar{W} \leq 5.5) \approx 0.89524 = 0.895$ (3 s.f.)	Students should learn to identify key words in the question - “<i>Estimate</i>” the probability that the <i>average number</i>” which are clues for us to consider Central Limit Theorem. Also note that the distribution of $\bar{W}$ follows normal distribution approximately, so it is important to state that in the solution.
----------------------------------	---	--

- 10 (a)** The waiting time for a bus during peak hours in a city is X minutes and it is known that the standard deviation of X is 1.2 minutes. The Metropolitan Transport Board (MTB) which runs a large fleet of buses claims that the mean waiting time for buses during peak hours in the city is 6.5 minutes.

With various changes to the infrastructure in the city, MTB wants to review whether the mean waiting time has changed since the claim was made. MTB records a random sample of 20 waiting times and found that the mean waiting time was $\bar{x}$ minutes.

- (i)** Given that $\bar{x} = 6.9$, test at the 5% level of significance, whether the mean waiting time differs from 6.5 minutes.

State an assumption for the test to be valid. [4]

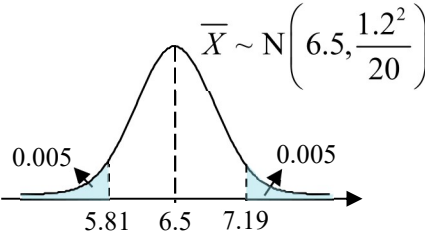
- (ii)** Explain what you understand by the phrase ‘at the 5% level of significance’ in the context of this question. [1]

- (iii)** Find the range of values of $\bar{x}$ for which there is sufficient evidence at 1% significance level that the mean waiting time differs from 6.5 minutes. [2]

- (b)** MTB operates the train system in the city as well and has received public feedback that the mean waiting time for trains has been getting worse. It is known that the waiting times for trains is normally distributed. MTB wishes to test, at the 5% level of significance, if the mean waiting time for trains is more than 3.7 minutes. A random sample of n observations, where $n \geq 30$, was taken and a mean of 3.95 minutes and standard deviation of 0.9 minutes were obtained.

Given that the null hypothesis is not rejected, obtain an inequality involving n , and hence find the set of values that n can take. [5]

Solution		Comments
(a) (i) [4]	Let X denote the waiting time for a bus in minutes and μ denote the population mean waiting time for buses in minutes. Assume X follows a normal distribution. $H_0 : \mu = 6.5$ vs $H_1 : \mu \neq 6.5$ Perform a 2-tail test at 5 % significance level. Under H_0, $\bar{X} \sim N\left(6.5, \frac{1.2^2}{20}\right)$. Using a z-test, $p\text{-value} = 2P(\bar{X} \geq 6.9) = 0.136$ (3 s.f) Since $p\text{-value} = 0.136 > 0.05$, we do not reject H_0 and conclude that there is insufficient evidence at 5% level of significance, that the mean waiting time differs from 6.5 minutes.	If $n < 30$ and the distribution of the population is unknown, the first and foremost assumption is that the population has a normal distribution, otherwise the test can't be conducted. The p-value of a 2-tail test is twice the p-value of a 1-tail test. Include “insufficient evidence at 5% significance level” so the conclusion isn't too assertive.
(ii) [1]	‘at the 5% level of significance’ means there is a probability of 0.05 of wrongly concluding that the mean waiting time differs from 6.5 minutes when it is in fact 6.5 minutes.	Phrase it as “the probability of concluding H_1 is true when H_0 is true” in the context of the question. We assume H_0 is true for the test, so must include the part in bold.

(iii) [2]	$H_0 : \mu = 6.5$ vs $H_1 : \mu \neq 6.5$ Under H_0, $\bar{X} \sim N\left(6.5, \frac{1.2^2}{20}\right)$. At 1% significance level, reject H_0 if $p\text{-value} \leq 0.01$ $P(\bar{X} \leq \bar{x}) \leq 0.005$ or $P(\bar{X} \geq \bar{x}) \leq 0.005$  Using GC, $0 \leq \bar{x} \leq 5.81$ (3 s.f.) or $\bar{x} \geq 7.19$ (3 s.f.) GC screenshots: 	The range of values of $\bar{x}$ to reject H_0 is simply the critical region.
(b) [5]	Let Y denote waiting time for a train in minutes and μ_y denote the population mean waiting time for trains in minutes. Unbiased estimate of the population variance is $s_y^2 = \frac{n}{n-1}(0.9^2)$ Null hypothesis, $H_0: \mu_y = 3.7$ Alternative hypothesis, $H_1: \mu_y > 3.7$ We perform a 1-tail test at 5% level of significance. Under H_0, $\bar{Y} \sim N\left(3.7, \frac{\frac{n}{n-1}0.9^2}{n}\right)$ approximately i.e $\bar{Y} \sim N\left(3.7, \frac{0.9^2}{n-1}\right)$ approximately	Note that s^2 is not the sample variance Remember to state the distribution. Don't apply CLT since population is normally distributed.

	Since H_0 is not rejected, $p\text{-value} > 0.05$. $P(\bar{Y} \geq 3.95) > 0.05$ $P\left(Z \geq \frac{3.95 - 3.7}{\frac{0.9}{\sqrt{n-1}}}\right) > 0.05$ $P\left(Z \geq \frac{0.25\sqrt{n-1}}{0.9}\right) > 0.05$ From GC, $P(Z \geq 1.6449) = 0.05$ From the diagram, $\frac{0.25\sqrt{n-1}}{0.9} < 1.6449$ $\sqrt{n-1} < 1.6449 \left(\frac{0.9}{0.25}\right)$ $n < 36.1 \text{ (3 s.f.)}$ $\therefore$ The set of values of n is $\{n \in \mathbb{Z} : 30 \leq n \leq 36\}$ OR $\{30, 31, 32, 33, 34, 35, 36\}$	A sketch helps in getting the correct inequality.
--	--	--

