TEN-YEAR SERIES · WORKED SOLUTIONS

# 2021 A-Level H2 Mathematics Paper 2

11 questions · 100 marks

The question paper is © UCLES/SEAB and is not reproduced here — have your own copy of the paper open as you work through these solutions. The worked solutions and diagrams are original work by The Learning Cave and are for personal study only.

thelearningcave.com.sg
Physics. Math. Chemistry.

# Question 1 5 marks

Complex numbers · conjugate roots

**APPROACH**

Since $a$ and $b$ are real, the non-real roots of the cubic must occur in conjugate pairs, which immediately gives a second root. The third (real) root then follows from Vieta's formula for the sum of the roots, after which $a$ and $b$ follow from the remaining Vieta relations.

As $a$ and $b$ are real, the complex roots of $x^3 + 2x^2 + ax + b = 0$ occur in conjugate pairs. Since $1 + \frac{1}{2}i$ is a root, its conjugate $1 - \frac{1}{2}i$ is also a root.

Let the third root be $r$, which must be real. By Vieta's formula, the sum of the roots equals $-2$ (the negative of the coefficient of $x^2$):

$$\left(1 + \frac{1}{2}i\right) + \left(1 - \frac{1}{2}i\right) + r = -2$$
$$\Rightarrow 2 + r = -2$$
$$\Rightarrow r = -4$$

**To check your answer:** the three roots $1 + \frac{1}{2}i$, $1 - \frac{1}{2}i$, $-4$ sum to $2 + (-4) = -2$, as required ✓

By Vieta's formula, the product of the roots equals $-b$:

$$\left(1 + \frac{1}{2}i\right)\left(1 - \frac{1}{2}i\right)(-4) = -b$$

Since $\left(1 + \frac{1}{2}i\right)\left(1 - \frac{1}{2}i\right) = 1^2 - \left(\frac{1}{2}i\right)^2 = 1 + \frac{1}{4} = \frac{5}{4}$,

$$\frac{5}{4} \times (-4) = -b$$
$$\Rightarrow -5 = -b$$
$$\Rightarrow b = 5$$

Also by Vieta's formula, the sum of the products of the roots taken two at a time equals $a$:

$$a = \left(1 + \frac{1}{2}i\right)\left(1 - \frac{1}{2}i\right) + \left(1 - \frac{1}{2}i\right)(-4) + (-4)\left(1 + \frac{1}{2}i\right)$$
$$= \frac{5}{4} + (-4 + 2i) + (-4 - 2i)$$
$$= \frac{5}{4} - 8$$
$$= -\frac{27}{4}$$

**To check your answer:** expanding $\left(x - 1 - \frac{1}{2}i\right)\left(x - 1 + \frac{1}{2}i\right)(x + 4)$ directly gives $x^3 + 2x^2 - \frac{27}{4}x + 5$, matching $a = -\frac{27}{4}$ and $b = 5$ ✓

$\therefore$ the other roots are $x = 1 - \frac{1}{2}i$ and $x = -4$, with $a = -\frac{27}{4}$ and $b = 5$.

THE LEARNING CAVE thelearningcave.com.sg Physics. Math. Chemistry.

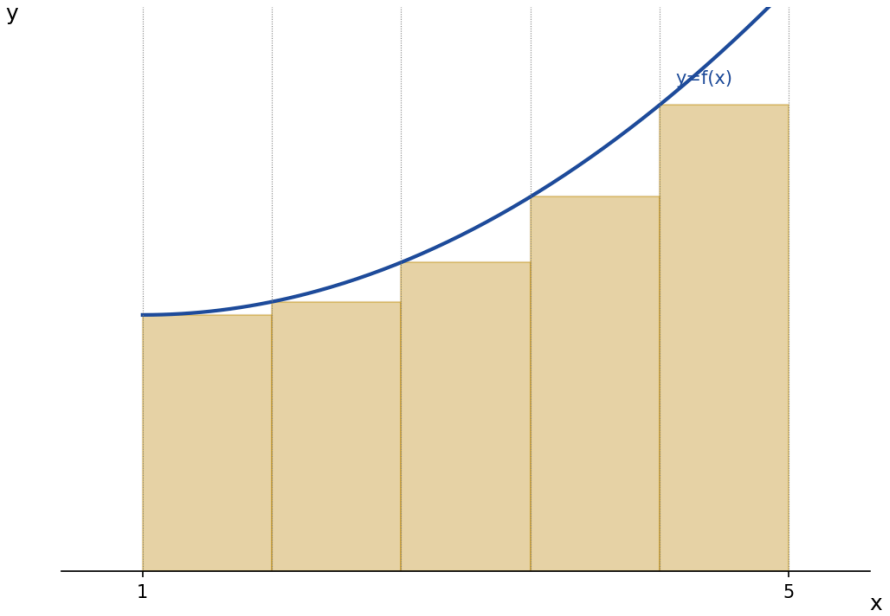
# Question 2 6 marks

Integration · rectangle estimates of area

## Part (a) 2 marks

$$h = \frac{5-1}{5} = 0.8$$

Since $y = f(x)$ is increasing on $[1, 5]$, as shown in the given sketch, using the value at the **left** endpoint of each strip as the rectangle's height gives a rectangle that lies entirely below the curve throughout that strip — the curve only gets higher moving right within the strip. So the total area of these rectangles is less than the area of $A$:




$\therefore h = 0.8$ and $\sum_{n=0}^{4} f(1+nh)\,h < \text{Area of } A.$

## Part (b) 3 marks

Using the value at the **right** endpoint of each strip instead gives a rectangle that lies entirely above the curve throughout that strip, since $f$ is increasing and the right endpoint gives the largest value of $f$ in the strip. So this total is greater than the area of $A$:

$\therefore \sum_{n=1}^{5} f(1+nh)\,h > \text{Area of } A.$

## Part (c)

With $f(x) = \frac{1}{20}x^2 + 1$ and $h = 0.8$, the strip boundaries are $x = 1,\ 1.8,\ 2.6,\ 3.4,\ 4.2,\ 5$.

$$f(1) = \frac{21}{20} = 1.050, \quad f(1.8) = \frac{581}{500} = 1.162, \quad f(2.6) = \frac{669}{500} = 1.338$$

$$f(3.4) = \frac{789}{500} = 1.578, \quad f(4.2) = \frac{941}{500} = 1.882, \quad f(5) = \frac{9}{4} = 2.250$$

Lower bound, using the left-endpoint sum from part (a):

$$\sum_{n=0}^{4} f(1+nh)\,h$$
$$= (1.050 + 1.162 + 1.338 + 1.578 + 1.882) \times 0.8$$
$$= \frac{701}{125} = 5.608$$

Upper bound, using the right-endpoint sum from part (b):

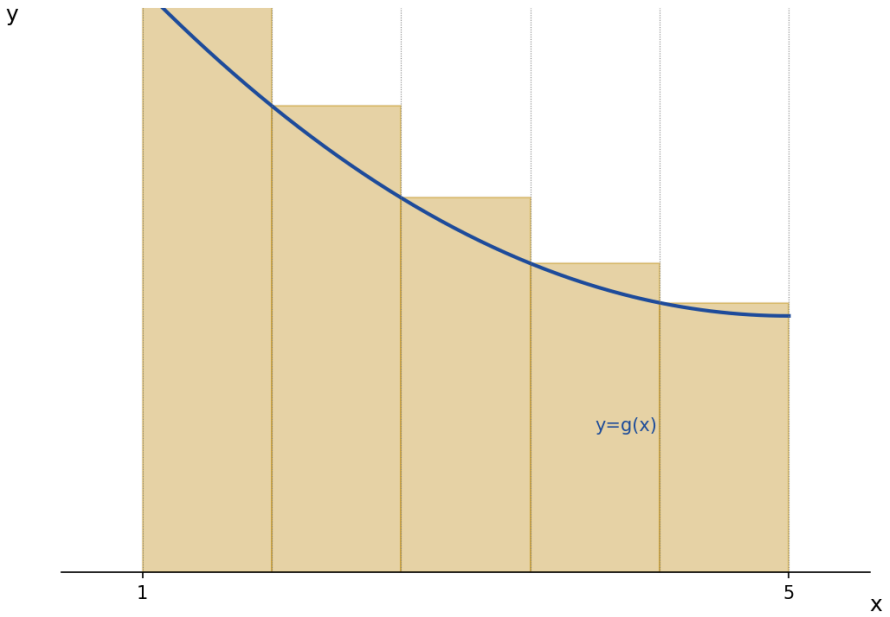
$$\sum_{n=1}^{5} f(1+nh)\,h$$
$$= (1.162 + 1.338 + 1.578 + 1.882 + 2.250) \times 0.8$$
$$= \frac{821}{125} = 6.568$$

**To check your answer:** the exact area, $\int_1^5 \left(\frac{x^2}{20} + 1\right) dx = \frac{91}{15} \approx 6.067$, lies between the lower bound $5.608$ and the upper bound $6.568$, as it must ✓

$\therefore \frac{701}{125} < \text{Area of } A < \frac{821}{125}$, i.e. $5.608 < \text{Area of } A < 6.568$.

## Part (d) 1 mark

For the area under $y = g(x)$ to be **less than** the left-endpoint sum $\sum_{n=0}^{4} g(1+nh)\,h$, that sum must **overestimate** the true area. This happens when $g$ is **decreasing** on $[1, 5]$: the left endpoint of each strip then gives the largest value of $g$ within that strip, so each rectangle lies entirely above the curve.




$\therefore$ any decreasing function $g(x)$ on $[1, 5]$ works, e.g. as sketched above.

# Question 3 9 marks

Functions · composite · inverse functions

## Part (a)(i) 2 marks

We first evaluate $h(2)$.

$$\begin{aligned} h(2) &= \frac{1}{2}(2)^2 + 3 \\ &= 2 + 3 \\ &= 5 \end{aligned}$$

Since $5 \neq 0.2$, this lies in the domain of $g$, so we substitute into $g$.

$$\begin{aligned} gh(2) &= g(5) \\ &= \frac{5+1}{5(5)-1} \\ &= \frac{6}{24} \\ &= \frac{1}{4} \end{aligned}$$

**To check your answer:** $h(2) = 5$ is within the domain of $g$ (since $5 \neq 0.2$), and $g(5) = \frac{6}{24} = 0.25$, a valid output. ✓

$\therefore gh(2) = \frac{1}{4}$.

## Part (a)(ii) 1 mark

$$x = \frac{2}{5}$$

Explanation: setting $g(x) = 1.4 = \frac{7}{5}$ and solving,

$$\begin{aligned} \frac{x+1}{5x-1} &= \frac{7}{5} \\ \Rightarrow 5(x+1) &= 7(5x-1) \\ \Rightarrow 5x + 5 &= 35x - 7 \\ \Rightarrow 30x &= 12 \\ \Rightarrow x &= \frac{2}{5} \end{aligned}$$

## Part (b)(i) 2 marks

$$k = -\frac{b}{2}$$

Explanation: $f$ is undefined when its denominator vanishes, i.e. when $2x + b = 0$, giving $x = -\frac{b}{2}$. Since division by zero is undefined, this value of $x$ must be excluded from the domain of $f$.

THE LEARNING CAVE thelearningcave.com.sg Physics. Math. Chemistry.

## Part (b)(ii) 3 marks

**APPROACH**

$f = f^{-1}$ means applying $f$ twice returns $x$ for every $x$ in the domain, so we find $f^{-1}$ explicitly and require $f^{-1}(x) = f(x)$ identically. This produces an equation that must hold for *every* $x$, not just specific values, so we collect the result by powers of $x$ and set each coefficient to zero.

Let $y = \frac{x+a}{2x+b}$ and solve for $x$ in terms of $y$.

$$\begin{aligned} y(2x+b) &= x+a \\ \Rightarrow 2xy + by &= x+a \\ \Rightarrow 2xy - x &= a - by \\ \Rightarrow x(2y-1) &= a - by \\ \Rightarrow x &= \frac{a-by}{2y-1} \end{aligned}$$

Replacing $y$ by $x$, $f^{-1}(x) = \frac{a-bx}{2x-1}$. Setting $f(x) = f^{-1}(x)$ for all $x$ in the domain,

$$\frac{x+a}{2x+b} = \frac{a-bx}{2x-1}$$

Cross-multiplying and expanding,

$$\begin{aligned} (x+a)(2x-1) &= (a-bx)(2x+b) \\ \Rightarrow 2x^2 - x + 2ax - a &= 2ax + ab - 2bx^2 - b^2x \end{aligned}$$

The $2ax$ terms cancel from both sides. Since this must hold for every $x$, we collect by powers of $x$ and equate each coefficient to zero.

$$\begin{aligned} (2+2b)x^2 + \left(b^2 - 1\right)x - a(1+b) &= 0 \text{ for all } x \\ \Rightarrow 2 + 2b = 0, \quad b^2 - 1 &= 0, \quad a(1+b) = 0 \end{aligned}$$

From the first equation, $b = -1$, which also satisfies $b^2 - 1 = 0$. With $b = -1$, the third equation becomes $a(0) = 0$, true for every value of $a$, so $a$ is not fixed by this condition.

Note: $f$ must still be a genuine function with an inverse, so its numerator cannot be a scalar multiple of its denominator (this would make $f$ constant). With $b = -1$, $x + a = \frac{1}{2}(2x - 1)$ when $a = -\frac{1}{2}$, giving $f(x) = \frac{1}{2}$ identically, which has no inverse, so this value of $a$ is excluded.

**To check your answer:** with $b = -1$ and, say, $a = 5$: $f(x) = \frac{x+5}{2x-1}$; computing $f\left(f(x)\right)$ symbolically gives exactly $x$, confirming $f$ is self-inverse. ✓

$\therefore b = -1$, and $a$ may be any real number except $a = -\frac{1}{2}$.

## Part (b)(iii) 1 mark

$$f^{-1}(-4) = \frac{4-a}{9}$$

Explanation: since $f = f^{-1}$, $f^{-1}(-4) = f(-4)$.

$$f(-4) = \frac{-4+a}{2(-4)+(-1)}$$
$$= \frac{a-4}{-9}$$
$$= \frac{4-a}{9}$$

## Question 4 9 marks

Sequences and series · AP and GP in context

### Part (a) 2 marks

We model Mrs Wong's ten times as an arithmetic progression (AP) with first term $a = 40$, last term $l = 25$ and $n = 10$ terms. The sum of an AP is $S_n = \frac{n}{2}(a+l)$.

$$S_{10} = \frac{10}{2}(40+25)$$
$$= 5(65)$$
$$= 325$$

**To check your answer:** the common difference is $d = \frac{25-40}{10-1} = -\frac{5}{3}$; the ten times $40,\ 38\frac{1}{3},\ 36\frac{2}{3},\ \ldots,\ 25$ sum to $325$. ✓

∴ the total time to swim 10 lengths is $325\ s$.

### Part (b) 5 marks

**APPROACH**

Suzie's 30-length swim splits into three blocks of 10 lengths, each under a different rule — Mrs Wong's AP (part (a)), a constant $25\ s$ per length, and Alfie's GP — so we find the total time as the sum of these three blocks, then divide the total distance by it.

For Alfie's programme, the times form a geometric progression (GP) with first term $a = 25$ and $10$ th term $ar^9 = 40$.

$$ar^9 = 40$$
$$\Rightarrow r^9 = \frac{40}{25} = \frac{8}{5}$$
$$\Rightarrow r = \left(\frac{8}{5}\right)^{1/9}$$

By GC, $r \approx 1.05361$. Using $S_n = \frac{a(r^n-1)}{r-1}$,

$$S_{10} = \frac{25\left(r^{10}-1\right)}{r-1}$$
$$\approx 319.797\ s\ (\text{by GC})$$

Suzie's total time for the 30 lengths is the sum of the three blocks: Mrs Wong's programme, the middle 10 lengths at $25\ s$ each, and Alfie's programme.

$$\begin{aligned}\text{Total time} &= 325 + 10(25) + 319.797 \\ &= 894.797\ s\end{aligned}$$

Her total distance is $30$ lengths of the $35\ m$ pool.

$$\text{Total distance} = 30 \times 35 = 1050\ m$$

$$\begin{aligned}\text{Average speed} &= \frac{1050}{894.797} \\ &\approx 1.17\ \text{m s}^{-1}\ \ (3\ \text{s.f.})\end{aligned}$$

**To check your answer:** $r^9 \approx 1.05361^9 \approx 1.600 = \frac{8}{5}$ ✓, matching the required ratio; and the average speed of $1.17\ \text{m s}^{-1}$ is consistent with covering $1050\ m$ in roughly $15$ minutes. ✓

∴ Suzie's average speed is $1.17\ \text{m s}^{-1}$ (3 s.f.).

### Part (c) 2 marks

We track the cumulative time at the end of each length. At the end of length $10$ (Mrs Wong's programme), the cumulative time is $325\ s$ (part (a)). Lengths $11$ to $20$ each take a constant $25\ s$, so the cumulative time at the end of length $10 + k$ is $325 + 25k$.

$$\begin{aligned}325 + 25k &= 480 \\ \Rightarrow k &= 6.2\end{aligned}$$

Since $k = 6$ gives $t = 475\ s$ (end of length 16) and $k = 7$ gives $t = 500\ s$ (end of length 17), and $475 < 480 < 500$, at $t = 480\ s$ Suzie is partway through length 17.

**To check your answer:** cumulative time at the end of length $16$ is $325 + 25(6) = 475\ s$ and at the end of length $17$ is $325 + 25(7) = 500\ s$; since $475 < 480 < 500$, $t = 480\ s$ does fall within length 17. ✓

Note: consecutive lengths alternate direction along the pool, with the first length swum away from the starting point, so odd-numbered lengths are swum away from it and even-numbered lengths towards it. Length $17$ is odd.

∴ exactly 8 minutes after starting, Suzie is swimming away from her starting point (she is on length 17).

## Question 5 11 marks

Integration · techniques of integration

### Part (a) 2 marks

We use the identity $\tan^2\theta = \sec^2\theta - 1$ to rewrite the integrand.

$$\begin{aligned}&\int \tan^2 5x \, dx \\ &= \int \left(\sec^2 5x - 1\right) \, dx \\ &= \frac{1}{5}\tan 5x - x + C\end{aligned}$$

**To check your answer:** differentiating $\frac{1}{5}\tan 5x - x + C$ gives $\sec^2 5x - 1 = \tan^2 5x$, the original integrand. ✓

$\therefore \int \tan^2 5x \, dx = \frac{1}{5}\tan 5x - x + C.$

## Part (b) 3 marks

**APPROACH**

the product $\sin 2x \sin 3x$ has no elementary antiderivative as it stands, so we first turn it into a sum of cosines using the product-to-sum identity, which we can integrate term by term.

We use $\sin A \sin B = \frac{1}{2}\left[\cos(A-B) - \cos(A+B)\right]$ with $A = 2x$, $B = 3x$.

$$\begin{aligned}\sin 2x \sin 3x &= \frac{1}{2}\left[\cos(2x - 3x) - \cos(2x + 3x)\right] \\ &= \frac{1}{2}\left[\cos(-x) - \cos 5x\right] \\ &= \frac{1}{2}\left(\cos x - \cos 5x\right)\end{aligned}$$

since $\cos$ is even. Integrating term by term,

$$\begin{aligned}&\int_0^b \sin 2x \sin 3x \, dx \\ &= \frac{1}{2}\int_0^b \left(\cos x - \cos 5x\right) \, dx \\ &= \frac{1}{2}\left[\sin x - \frac{1}{5}\sin 5x\right]_0^b \\ &= \frac{1}{2}\sin b - \frac{1}{10}\sin 5b\end{aligned}$$

(the lower limit contributes $0$, since $\sin 0 = 0$).

**To check your answer:** at $b = \frac{\pi}{2}$: $\frac{1}{2}\sin\frac{\pi}{2} - \frac{1}{10}\sin\frac{5\pi}{2} = \frac{1}{2}(1) - \frac{1}{10}(1) = 0.4$, matching direct numerical integration of $\sin 2x \sin 3x$ from $0$ to $\frac{\pi}{2}$. ✓

$\therefore \int_0^b \sin 2x \sin 3x \, dx = \frac{1}{2}\sin b - \frac{1}{10}\sin 5b.$

## Part (c) 3 marks

Since the derivative of $\ln x$ is $\frac{1}{x}$, we let $u = \ln x$, so $du = \frac{1}{x}\,dx$. When $x = a$, $u = \ln a$; when $x = b$, $u = \ln b$.

$$\begin{aligned}
&\int_a^b \frac{1}{x \ln x}\, dx \\
&= \int_{\ln a}^{\ln b} \frac{1}{u}\, du \\
&= \left[\ln |u|\right]_{\ln a}^{\ln b} \\
&= \ln(\ln b) - \ln(\ln a)
\end{aligned}$$

Since $1 < a < b$, both $\ln a > 0$ and $\ln b > 0$, so by the quotient rule for logarithms,

$$\ln(\ln b) - \ln(\ln a) = \ln\left(\frac{\ln b}{\ln a}\right)$$

**To check your answer:** with $a = e$, $b = e^2$: $\ln(\ln b) - \ln(\ln a) = \ln 2 - \ln 1 = \ln 2 \approx 0.693$; direct numerical integration of $\frac{1}{x \ln x}$ from $e$ to $e^2$ also gives $\approx 0.693$. ✓

$\therefore \int_a^b \frac{1}{x \ln x}\, dx = \ln\left(\frac{\ln b}{\ln a}\right)$.

## Part (d) 3 marks

We are given $u = 1 + e^{2x}$, so $\frac{du}{dx} = 2e^{2x}$, i.e. $e^{2x}\, dx = \frac{1}{2}\, du$.

$$\begin{aligned}
&\int \frac{e^{2x}}{\left(1 + e^{2x}\right)^3}\, dx \\
&= \int \frac{1}{u^3} \cdot \frac{1}{2}\, du \\
&= \frac{1}{2} \int u^{-3}\, du \\
&= \frac{1}{2}\left(\frac{u^{-2}}{-2}\right) + C \\
&= -\frac{1}{4u^2} + C
\end{aligned}$$

Substituting back $u = 1 + e^{2x}$,

$$-\frac{1}{4u^2} + C = -\frac{1}{4\left(1 + e^{2x}\right)^2} + C$$

**To check your answer:** differentiating $-\frac{1}{4(1+e^{2x})^2} + C$ by the chain rule gives $-\frac{1}{4} \times (-2)\left(1 + e^{2x}\right)^{-3} \times 2e^{2x} = \frac{e^{2x}}{(1+e^{2x})^3}$, the original integrand. ✓

$\therefore \int \frac{e^{2x}}{(1+e^{2x})^3}\, dx = -\frac{1}{4(1+e^{2x})^2} + C$.

thelearningcave.com.sg
Physics. Math. Chemistry.

# Question 6 7 marks

Discrete random variables · expectation and variance

**APPROACH**

we have two unknowns, $p$ and $q$, but only one equation follows directly from the probabilities summing to 1. A second comes from the given variance, which needs both $E(X)$ and $E\left(X^2\right)$. We express $q$ in terms of $p$ from the first equation, then substitute into $Var(X) = E\left(X^2\right) - [E(X)]^2 = 1.61$ to get a quadratic in $p$ alone.

Let $X$ be the score obtained on the spinner, with distribution

| $x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $P(X = x)$ | 0.2 | 0.3 | $p$ | $p$ | $q$ |

Since the probabilities sum to 1,

$$0.2 + 0.3 + p + p + q = 1$$
$$\Rightarrow 2p + q = 0.5 \quad (1)$$

The mean is

$$E(X) = 1(0.2) + 2(0.3) + 3p + 4p + 5q$$
$$= 0.8 + 7p + 5q$$

Using (1), $q = 0.5 - 2p$, so

$$E(X) = 0.8 + 7p + 5(0.5 - 2p)$$
$$= 3.3 - 3p \quad (2)$$

Similarly,

$$E\left(X^2\right) = 1^2(0.2) + 2^2(0.3) + 3^2p + 4^2p + 5^2q$$
$$= 1.4 + 25p + 25q$$
$$= 1.4 + 25p + 25(0.5 - 2p)$$
$$= 13.9 - 25p \quad (3)$$

Using $Var(X) = E\left(X^2\right) - [E(X)]^2 = 1.61$ and substituting (2) and (3),

$$13.9 - 25p - (3.3 - 3p)^2 = 1.61$$
$$\Rightarrow 13.9 - 25p - \left(10.89 - 19.8p + 9p^2\right) = 1.61$$
$$\Rightarrow 3.01 - 5.2p - 9p^2 = 1.61$$
$$\Rightarrow 9p^2 + 5.2p - 1.4 = 0$$

By the quadratic formula (or GC),

$$p = \frac{-5.2 \pm \sqrt{5.2^2 + 4(9)(1.4)}}{2(9)}$$
$$= \frac{-5.2 \pm \sqrt{77.44}}{18}$$

$$= \frac{-5.2 \pm 8.8}{18}$$

This gives $p = 0.2$ or $p = -\frac{7}{9}$. Since $p$ is a probability, $0 \leq p \leq 1$, so $p = -\frac{7}{9}$ $(\text{rej. } \text{p} < 0)$, and $p = 0.2$.

$$q = 0.5 - 2(0.2) = 0.1$$

**To check your answer:** with $p = 0.2$, $q = 0.1$, all five probabilities $0.2, 0.3, 0.2, 0.2, 0.1$ are non-negative and sum to $1$. Also $E\left(X^2\right) = 13.9 - 25(0.2) = 8.9$, giving $Var(X) = 8.9 - 2.7^2 = 1.61$, matching the given variance. ✓

Substituting into (2),

$$E(X) = 3.3 - 3(0.2) = 2.7$$

$\therefore$ the mean score is $2.7$.

## Question 7 8 marks

Permutations and combinations · repeated letters

The word ABRACADABRA has $11$ letters: $5$ A's, $2$ B's, $2$ R's, $1$ C, and $1$ D.

### Part (a) 2 marks

Since the letters repeat, we divide the number of permutations of $11$ distinct objects by the factorial of each repeated letter's count.

$$\begin{aligned} &\text{no. of arrangements} \\ &= \frac{11!}{5!\,2!\,2!} \\ &= 83160 \end{aligned}$$

**To check your answer:** $11! = 39916800$ and $5! \times 2! \times 2! = 480$; $39916800 \div 480 = 83160$, a positive integer as expected. ✓

$\therefore$ the number of different arrangements is $83160$.

### Part (b) 3 marks

**APPROACH**

each condition glues a group of letters together, so we treat each group as a single unit and arrange the units. The A's need care: "exactly 4 of the 5 A's together" is not just "glue 4 A's into a block" — we must also make sure that block does not end up next to the 5th A, since that would merge into a run of all 5. We count all arrangements with a block of 4 A's, then subtract those where the block sits next to the lone A.

Since the 2 B's must be together, treat them as one unit $[BB]$. Since the 2 R's must be together, treat them as one unit $[RR]$. Since C is next to D, treat them as one unit $[CD]$, which has 2 internal orders ($CD$ or $DC$). Since exactly 4 of the A's are together, treat those 4 as one unit $[\text{AAAA}]$, leaving the 5th A as a separate letter.

This gives 5 units to arrange in a row: $[BB]$, $[RR]$, $[CD]$, $[\text{AAAA}]$, $A$.

$$\begin{aligned}&\text{no. of arrangements of the 5 units}\\&= 5! \times 2\\&= 240\end{aligned}$$

From these 240, we subtract the arrangements in which $[\text{AAAA}]$ and the lone $A$ are adjacent, since that would give a run of 5 A's rather than exactly 4. Gluing $[\text{AAAA}]$ and $A$ together (in either order) leaves 4 units to arrange:

$$\begin{aligned}&\text{no. with all 5 } A'\text{s together}\\&= 4! \times 2 \text{ (order of the glued A-block)} \times 2 \text{ ($CD$ order)}\\&= 96\end{aligned}$$

$$\text{no. of valid arrangements} = 240 - 96 = 144$$

**To check your answer:** a direct computer enumeration of all 83160 arrangements of ABRACADABRA, filtered by "BB adjacent", "RR adjacent", "C next to D" and "the A's form a run of length exactly 4", also gives 144. ✓

∴ the number of such arrangements is 144.

## Part (c) 3 marks

Treating all 5 A's as one unit $[\text{AAAAA}]$, we have 7 units to arrange: $[\text{AAAAA}]$, $B$, $B$, $R$, $R$, $C$, $D$, with the B's and R's each repeated.

$$\begin{aligned}&\text{no. of arrangements with all 5 } A'\text{s together}\\&= \frac{7!}{2!\,2!}\\&= 1260\end{aligned}$$

Since all 83160 arrangements from **(a)** are equally likely,

$$\begin{aligned}P(\text{all 5 A's together}) &= \frac{1260}{83160}\\&= \frac{1}{66}\end{aligned}$$

**To check your answer:** $83160 \div 1260 = 66$ exactly, so the fraction reduces to the unit fraction $\frac{1}{66}$. ✓

∴ $P(\text{all 5 A's together}) = \frac{1}{66}$.

# Question 8 9 marks

Hypothesis testing · test of a mean

## Part (a) 3 marks

The sales manager wants to test only whether the tyres last **longer** than $20\,000$ miles (a specific direction), not merely whether the life span differs from $20\,000$ miles in either direction. This one-directional claim is why a $1$-tail test is appropriate.

Let $\mu$ thousand miles be the population mean life span of the front tyres.

$$H_0 : \mu = 20$$
$$H_1 : \mu > 20$$

where the test is carried out at the $5\%$ significance level.

∴ a $1$-tail test is appropriate, with $H_0 : \mu = 20$ tested against $H_1 : \mu > 20$ at the $5\%$ level of significance.

## Part (b) 2 marks

With $n = 50$, $\Sigma(x - 20) = 9.4$ and $\Sigma(x - 20)^2 = 38.76$, we find the unbiased estimates from the coded data.

$$\begin{aligned}\overline{x} &= 20 + \frac{\Sigma(x - 20)}{n} \\ &= 20 + \frac{9.4}{50} \\ &= 20.188\end{aligned}$$

$$\begin{aligned}s^2 &= \frac{1}{n-1}\left[\Sigma(x-20)^2 - \frac{(\Sigma(x-20))^2}{n}\right] \\ &= \frac{1}{49}\left[38.76 - \frac{9.4^2}{50}\right] \\ &= \frac{36.9928}{49} \\ &= 0.755 \ \ (3 \text{ s.f.})\end{aligned}$$

∴ the unbiased estimate of the population mean is $\overline{x} = 20.188$ thousand miles, and the unbiased estimate of the population variance is $s^2 = 0.755$ (thousand miles)$^2$ (3 s.f.).

## Part (c) 3 marks

Since $n = 50$ is large, by the Central Limit Theorem $\overline{X}$ is approximately normal even though the population distribution of tyre life spans is not stated. Under $H_0$,

$$\overline{X} \sim N\left(20, \frac{s^2}{n}\right) \text{ approximately}$$

Using the sample data, the test statistic is

$$z = \frac{\overline{x} - 20}{\sqrt{s^2/n}}$$
$$= \frac{0.188}{\sqrt{0.75496/50}}$$
$$= 1.53 \text{ (3 s.f., by GC)}$$

For a 1-tail test at the $5\%$ level, the critical value is $z = 1.6449$, i.e. the critical region is $z > 1.6449$.

**To check your answer:** since $z \approx 1.53 < 1.6449$, the test statistic does not fall in the critical region, consistent with $\overline{x} = 20.188$ being only modestly above $20$ relative to the spread of the data. ✓

Since $z \approx 1.53 < 1.6449$, we do not reject $H_0$.

∴ there is insufficient evidence, at the $5\%$ level of significance, to conclude that the mean life span of front tyres is more than $20\,000$ miles.

### Part (d) 1 mark

The test in **(c)** would no longer be valid.

Explanation: with only $15$ tyres, the sample is too small for the Central Limit Theorem to guarantee that $\overline{X}$ is approximately normal. Since the population distribution of tyre life spans is not given to be normal, $\overline{X}$ can no longer be assumed approximately normal, so the $z$-test in **(c)** breaks down.

∴ the test would be inappropriate, since the Central Limit Theorem no longer applies for a sample of only $15$ tyres.

## Question 9 7 marks

Correlation and regression · transformed models

### Part (a) 1 mark

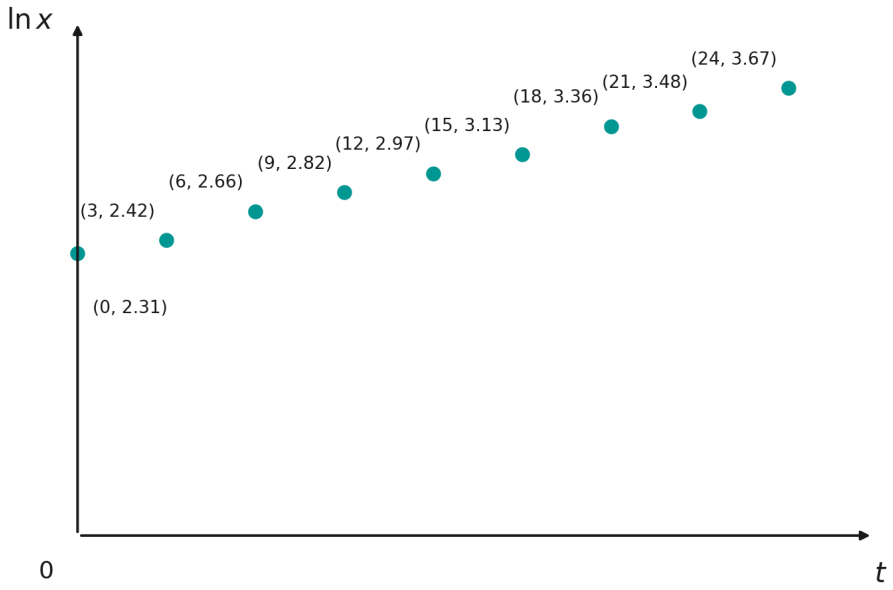
No — a linear model is not likely to fit well.

Explanation: although $r = 0.9803$ is close to $1$, indicating a strong *linear* association, the scatter diagram shows the points bending upward as $t$ increases, a curved pattern rather than a straight line. A high value of $r$ measures only the strength of a linear relationship; it does not rule out a non-linear model fitting even better, so the curvature visible in the diagram means a straight line is not a good model here.

∴ a linear model is not likely to be a good model for $x$ against $t$.

### Part (b)(i) 1 mark

We compute $\ln x$ for each value of $x$ in the table and plot it against $t$.




∴ the points now lie close to a straight line with positive gradient, with none of the curvature seen in the original $x$ against $t$ diagram.

## Part (b)(ii) 3 marks

By GC, performing linear regression of $\ln x$ on $t$,

$$\ln x = c + dt$$

$$c \approx 2.29, \quad d \approx 0.0573 \text{ (3 s.f., by GC)}$$

and the product moment correlation coefficient for this model is

$$r \approx 0.999 \text{ (3 s.f., by GC)}.$$

**To check your answer:** at $t = 24$, $\ln x \approx 2.293 + 0.0573(24) = 3.68$, so $x \approx e^{3.68} \approx 39.6$, close to the given $x = 39.3$. ✓

∴ $c \approx 2.29$, $d \approx 0.0573$, and $r \approx 0.999$ (3 s.f.).

## Part (c) 2 marks

The model $\ln x = c + dt$ has $r \approx 0.999$, closer to 1 than the linear model's $r = 0.9803$, and its scatter diagram in **(b)(i)** shows the points lying close to a straight line with no visible curvature, unlike the original diagram. Since $\ln x = c + dt$ is equivalent to $x = e^c e^{dt} = Ae^{dt}$ (an exponential growth model, where $A = e^c$), Toby's second model fits the data much better than his first, straight-line model for $x$ against $t$.

∴ the exponential model $x = Ae^{dt}$, obtained from $\ln x = c + dt$, is a much better fit than the straight-line model $x$ against $t$.

thelearningcave.com.sg
Physics. Math. Chemistry.

# Question 10 15 marks

Normal distribution · sums of normal variables

## Part (a) 3 marks

Let $X$ kg be the mass of a seat, where $X \sim N\left(\mu,\ \sigma^2\right)$. We are given $P(X < 2.1) = 0.80$ and $P(X < 1.95) = 0.15$.

By GC (inverse normal),

$$\frac{2.1-\mu}{\sigma} = 0.841621\ldots \qquad \text{and} \qquad \frac{1.95-\mu}{\sigma} = -1.036433\ldots$$

$$2.1 = \mu + 0.841621\sigma \qquad \ldots(1)$$

$$1.95 = \mu - 1.036433\sigma \qquad \ldots(2)$$

Subtracting (2) from (1),

$$\begin{aligned} 2.1 - 1.95 &= 0.841621\sigma - (-1.036433\sigma) \\ &= 1.878055\sigma \\ \sigma &= \frac{0.15}{1.878055} \\ &= 0.0799 \ (3\text{ s.f.}) \quad (\text{shown}) \end{aligned}$$

Substituting back into (1),

$$\begin{aligned} \mu &= 2.1 - 0.841621(0.079870) \\ &= 2.03 \ (3\text{ s.f.}) \end{aligned}$$

**To check your answer:** substituting into (2), $\mu - 1.036433\sigma \approx 2.032780 - 1.036433(0.079870) \approx 1.950$, matching the given value. ✓

∴ the mean mass of the seats is $\mu \approx 2.03$ kg, and the standard deviation is $\sigma \approx 0.0799$ kg (both 3 s.f.).

## Part (b) 2 marks

Let $L$ kg be the mass of a leg, $L \sim N\left(1.2,\ 0.02^2\right)$. Standardising,

$$\begin{aligned} &P(L > 1.21) \\ &= P\left(Z > \frac{1.21-1.2}{0.02}\right) \\ &= P(Z > 0.5) \\ &= 0.308538 \ (\text{by GC}) \end{aligned}$$

In a batch of $500$ legs, the expected number with mass more than $1.21$ kg is

$$500 \times 0.308538 = 154.3 \ (4\text{ s.f.})$$

∴ the expected number of legs with mass more than $1.21$ kg is $154$ (nearest whole number).

## Part (c) 3 marks

Let $T = X + L_1 + L_2 + L_3$ be the total mass of a randomly chosen seat and 3 randomly chosen legs, where $X \sim N\left(2.032780,\ 0.079870^2\right)$ (using the unrounded values from **(a)**) and each $L_i \sim N\left(1.2,\ 0.02^2\right)$ independently. Since a sum of independent normal random variables is normal,

$$E(T) = \mu + 3(1.2)$$
$$= 2.032780 + 3.6$$
$$= 5.632780$$
$$Var(T) = \sigma^2 + 3(0.02)^2$$
$$= 0.006379 + 0.0012$$
$$= 0.007579$$
$$T \sim N(5.632780,\ 0.007579)$$
$$P(5.6 < T < 5.7) = 0.427 \text{ (3 s.f., by GC)}$$

**To check your answer:** $sd(T) \approx 0.0871$ kg, and the interval $[5.6,\ 5.7]$ spans roughly $\pm 0.4$ standard deviations either side of the mean $5.633$, consistent with a probability in the moderate range around $0.4$–$0.5$. ✓

$\therefore P(5.6 < T < 5.7) \approx 0.427$ (3 s.f.).

## Part (d) 3 marks

**APPROACH**

drilling reduces every seat's mass by $9\%$, so the drilled mass is the linear transformation $0.91X$, not $X$ minus a fixed amount — scaling a normal random variable by a constant scales its mean by that constant and its variance by the square of that constant.

Since $0.91X \sim N\left(0.91\mu,\ (0.91\sigma)^2\right)$,

$$E(0.91X) = 0.91(2.032780) = 1.849830$$
$$Var(0.91X) = (0.91)^2(0.079870)^2 = 0.006483$$

Let $T' = 0.91X + L_1 + L_2 + L_3$ be the total mass of a drilled seat and 3 legs.

$$E(T') = 1.849830 + 3.6$$
$$= 5.449830$$
$$Var(T') = 0.006483 + 0.0012$$
$$= 0.007683$$
$$T' \sim N(5.449830,\ 0.007683)$$
$$P(T' < 5.6) = 0.969 \text{ (3 s.f., by GC)}$$

**To check your answer:** $sd(T') \approx 0.0805$ kg, so $5.6$ kg is about $1.86$ standard deviations above the new mean of $5.450$ kg, consistent with a high probability close to (but less than) $1$. ✓

thelearningcave.com.sg
Physics Made Chemistry.

$\therefore P(T' < 5.6) \approx 0.969$ (3 s.f.).

### Part (e) 4 marks

**APPROACH**

the fitting conditions are stated in terms of the *difference* between a leg's diameter and its hole's diameter, so we work with $D = M - H$ rather than $M$ and $H$ separately. "No sanding" and "no padding" each translate into a one-sided bound on $D$, and together they give a range for $D$; since a leg is fitted to only one hole, we can find the probability for one leg-hole pair and then use independence across the 3 pairs.

Let $H$ mm be the diameter of a hole, $H \sim N\left(31,\ 0.4^2\right)$, and let $M$ mm be the diameter of a leg, $M \sim N\left(30.7,\ 0.3^2\right)$, with $H$ and $M$ independent. Let $D = M - H$.

$$E(D) = 30.7 - 31 = -0.3$$

$$Var(D) = 0.3^2 + 0.4^2 = 0.09 + 0.16 = 0.25$$

$$D \sim N(-0.3,\ 0.25)$$

No sanding is needed when the leg is not wider than the hole, i.e. $M \leq H$, i.e. $D \leq 0$. No padding is needed when the hole is not more than $0.8$ mm wider than the leg, i.e. $H - M \leq 0.8$, i.e. $D \geq -0.8$. So one leg-hole pair fits without sanding or padding exactly when $-0.8 \leq D \leq 0$.

$$P(-0.8 \leq D \leq 0) = 0.567 \text{ (3 s.f., by GC)}$$

Since the 3 legs are paired with the 3 holes at random, and the diameters of different legs and different holes are all independent of one another, whether each of the 3 leg-hole pairs fits is independent of the other pairs.

$$P\left(\text{all 3 fit}\right) = (0.567092)^3$$

$$= 0.182 \ \text{(3 s.f.)}$$

**To check your answer:** each individual fit has a moderate probability of about $0.567$; cubing this for 3 independent fits gives a noticeably smaller probability, as expected. ✓

$\therefore P\left(\text{all 3 legs fit without sanding or padding}\right) \approx 0.182$ (3 s.f.).

## Question 11 14 marks

Binomial distribution · probabilities in context

### Part (a)(i) 2 marks

Select a simple random sample of the 100 light fittings.

Explanation: number all the light fittings produced that day, then use a random number generator (or random number tables) to choose 100 of the numbers, and test the fittings with those numbers. This gives every light fitting an equal chance of selection, avoiding bias — for instance,

avoiding always testing fittings from the same point in the production run, which might not be representative of the whole day's output.

∴ select a simple random sample (e.g. using random numbers), to avoid bias.

### Part (a)(ii) 1 mark

Using the frequency table, we find the total number of faulty fittings recorded over the $150$ days.

$$\begin{aligned}\Sigma fx &= 0(4) + 1(19) + 2(38) + 3(41) + 4(22) + 5(16) + 6(6) + 7(4) + 8(0) \\ &= 450\end{aligned}$$

The sample mean number of faulty fittings per day is

$$\overline{x} = \frac{450}{150} = 3$$

Since each day's sample contains $100$ fittings, we estimate $p$ by

$$\hat{p} = \frac{\overline{x}}{100} = \frac{3}{100} = 0.03$$

**To check your answer:** the number of days $4 + 19 + 38 + 41 + 22 + 16 + 6 + 4 + 0 = 150$, matching the given total ✓

∴ $\hat{p} = 0.03$.

### Part (a)(iii) 2 marks

Let $Y$ be the number of faulty light fittings found on a day, so $Y \sim B(100,\ 0.03)$, using the estimate of $p$ from part (ii). By GC,

$$\begin{aligned}P(Y = 3) &= P(Y \leq 3) - P(Y \leq 2) \\ &= 0.6472 - 0.4198 \\ &= 0.2275\end{aligned}$$

Over $150$ working days, the expected number of days on which exactly $3$ faulty fittings are found is

$$150 \times 0.2275 \approx 34.1$$

**To check your answer:** $P(Y = 3) \approx 0.2275$ is close to (but slightly less than) the probability at the mean, consistent with $Y$ being roughly symmetric near its mean of $100(0.03) = 3$ ✓

∴ the expected number of days is $34.1$ (3 s.f.).

### Part (b)(i) 2 marks

Two assumptions are needed for $X$ to be well modelled by a binomial distribution:

1. Whether one heating element is faulty is independent of whether any other heating element tested that day is faulty.

2. The probability that a heating element is faulty is constant for every heating element tested on a given day.

$\therefore$ independence between elements, and a constant probability of a fault, across the elements tested each day.

### Part (b)(ii) 2 marks

Assume now $X \sim B(80,\ 0.02)$. By GC,

$$\begin{aligned} P(1 \leq X \leq 4) &= P(X \leq 4) - P(X \leq 0) \\ &= 0.9776 - 0.1986 \\ &= 0.7790 \end{aligned}$$

**To check your answer:** $P(X \leq 0) \approx 0.1986$ is the probability of no faults, plausible for $n = 80$, $p = 0.02$ (mean $=1.6$); the resulting range probability $0.779$ is sensibly less than $1$ ✓

$\therefore P(1 \leq X \leq 4) = 0.779$ (3 s.f.).

### Part (b)(iii) 3 marks

**APPROACH**

"more than 3 elements found faulty" is a daily event; we first find its probability $q$ for a single day, then treat the number of days (out of 5) on which it happens as a second binomial variable.

We first find $q = P(X > 3)$, the probability that more than $3$ elements are faulty on a given day. By GC,

$$\begin{aligned} P(X \leq 3) &= 0.9231 \\ q = P(X > 3) &= 1 - P(X \leq 3) \\ &= 1 - 0.9231 \\ &= 0.0769 \end{aligned}$$

Let $W$ be the number of days, out of $5$, on which more than $3$ elements are found faulty, so $W \sim B(5,\ 0.0769)$. By GC,

$$\begin{aligned} P(W \leq 1) &= 0.9495 \\ P(W \geq 2) &= 1 - P(W \leq 1) \\ &= 1 - 0.9495 \\ &= 0.0505 \end{aligned}$$

**To check your answer:** since $q \approx 0.0769$ is small, "more than 3 faulty" is a rare daily event, so it occurring on $2$ or more of $5$ days should be a small probability, consistent with $0.0505$ ✓

$\therefore P(W \geq 2) = 0.0505$ (3 s.f.).

## Part (b)(iv) 2 marks

Let $T$ be the total number of faulty elements found over a $5$-day working week. Since $T$ is the sum of $5$ independent $B(80,\ 0.02)$ variables, $T \sim B(400,\ 0.02)$. By GC,

$$P(T \leq 8) = 0.593$$

**To check your answer:** $T$ has mean $400(0.02) = 8$, so $P(T \leq 8)$ being just under $0.6$ (slightly more than half) is consistent with $8$ being at the mean of a discrete distribution ✓

$\therefore P(T \leq 8) = 0.593$ (3 s.f.).

THE LEARNING CAVE
thelearningcave.com.sg
Physics. Math. Chemistry.