TEN-YEAR SERIES · WORKED SOLUTIONS

2020 A-Level H2 Mathematics Paper 2

10 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

Differentiation · quadratic curves

Since the curve is quadratic and has a minimum point at $(1, -2)$, we write it in vertex form,

$$y = a(x-1)^2 - 2,$$

where $a > 0$ is a constant to be found.

Differentiating,

$$\frac{dy}{dx} = 2a(x-1).$$

Using the given gradient at $x = 2$:

$$\begin{aligned} 2a(2-1) &= 5 \\ 2a &= 5 \\ a &= \frac{5}{2}. \end{aligned}$$

Substituting back,

$$\begin{aligned} y &= \frac{5}{2}(x-1)^2 - 2 \\ &= \frac{5}{2}x^2 - 5x + \frac{1}{2}. \end{aligned}$$

To check your answer: at $x = 1$, $y = \frac{5}{2}(0)^2 - 2 = -2$, the stated minimum; and $\frac{dy}{dx} = 5(x-1)$ gives $5(1) = 5$ at $x = 2$. ✓

$\therefore$ the equation of the curve is $y = \frac{5}{2}x^2 - 5x + \frac{1}{2}$ (equivalently $y = \frac{5}{2}(x-1)^2 - 2$).

Question 2 12 marks

Sequences and series · recurrence relations · sums

Part (a)

APPROACH

Parts (i) and (ii) all concern the same recurrence $u_{n+1} = 2u_n - 5$. Rather than generate terms one at a time, we look for the fixed point of the recurrence — the value u for which $u = 2u - 5$ — and write $u_n - u$ as a geometric sequence about it. This gives a closed form we can reuse for every part.

Since the fixed point satisfies $u = 2u - 5$, we get $u = 5$. Subtracting this from the recurrence,

$$\begin{aligned} u_{n+1} - 5 &= 2u_n - 10 \\ &= 2\left(u_n - 5\right). \end{aligned}$$

So $u_n - 5$ is geometric with common ratio 2 and first term $u_1 - 5 = p - 5$, giving

$$u_n = 5 + (p-5)\cdot 2^{n-1}.$$

Part (i)

Part (A) 1 mark

With $p = 7$,

$$\begin{aligned} u_n &= 5 + 2\cdot 2^{n-1} \\ &= 5 + 2^n. \end{aligned}$$

Explanation: since $2^n \to \infty$ as $n \to \infty$, the terms increase strictly and without bound; the sequence diverges to infinity.

Part (B) 1 mark

With $p = 5$,

$$u_n = 5 + 0\cdot 2^{n-1} = 5.$$

Explanation: every term equals 5; the sequence is constant.

Part (ii) 2 marks

Using the formula with $n = 5$:

$$\begin{aligned} u_5 &= 5 + (p-5)\cdot 2^4 \\ &= 5 + 16(p-5). \end{aligned}$$

Using the given value $u_5 = 101$:

$$\begin{aligned} 5 + 16(p-5) &= 101 \\ 16(p-5) &= 96 \\ p - 5 &= 6 \\ p &= 11. \end{aligned}$$

$\therefore p = 11.$

Part (b)

> **APPROACH**
>
> We need v_3 and v_4 in terms of a and b before we can use the condition $v_4 = 2v_3$.

Using the recurrence with $n = 1$:

$$v_3 = v_1 + 2v_2 - 7 = a + 2b - 7.$$

Using the recurrence with $n = 2$:

$$\begin{aligned} v_4 &= v_2 + 2v_3 - 7 \\ &= b + 2(a + 2b - 7) - 7 \\ &= 2a + 5b - 21. \end{aligned}$$

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

Part (i) 3 marks

Substituting into $v_4 = 2v_3$:

$$2a + 5b - 21 = 2(a + 2b - 7)$$
$$2a + 5b - 21 = 2a + 4b - 14$$
$$b = 7.$$

$\therefore b = 7.$

Part (ii) 1 mark

With $b = 7$,

$$v_3 = a + 2(7) - 7 = a + 7,$$
$$v_4 = 2v_3 = 2a + 14.$$

Using the recurrence with $n = 3$:

$$v_5 = v_3 + 2v_4 - 7$$
$$= (a + 7) + 2(2a + 14) - 7$$
$$= 5a + 28.$$

$\therefore v_5 = 5a + 28.$

Part (c)

Part (i) 2 marks

Since $S_n = n^3 - 11n^2 + 4n$ is the sum of the first n terms, the nth term is $T_n = S_n - S_{n-1}$ for $n > 1$.

$$T_n = \left(n^3 - 11n^2 + 4n\right) - \left[(n-1)^3 - 11(n-1)^2 + 4(n-1)\right]$$
$$= 3n^2 - 25n + 16.$$

Note: checking $n = 1$, $T_1 = 3 - 25 + 16 = -6 = S_1$, so this formula also gives the first term correctly and holds for all $n \geq 1$.

$\therefore T_n = 3n^2 - 25n + 16.$

Part (ii) 2 marks

APPROACH

"The sum of the first m terms equals the sum of the first three terms" means $S_m = S_3$, so $n = m$ and $n = 3$ are both roots of $S_n = S_3$; we factor out $(n - 3)$ to find the others.

$$S_3 = 3^3 - 11(3)^2 + 4(3) = -60.$$

Setting $S_m = -60$:

$$m^3 - 11m^2 + 4m = -60$$
$$m^3 - 11m^2 + 4m + 60 = 0.$$

Since $m = 3$ is a root, we factor it out:

$$(m-3)\left(m^2 - 8m - 20\right) = 0$$
$$(m-3)(m-10)(m+2) = 0$$
$$m = 3,\ 10,\ -2.$$

Since $m > 3$, $m = 3$ and $m = -2$ are rejected ($m = -2$ also rejected as m is a positive integer).

To check your answer: $S_{10} = 1000 - 1100 + 40 = -60 = S_3$. ✓

$\therefore m = 10.$

Question 3 12 marks

Parametric equations · normals · area

Part (i) 5 marks

APPROACH

We first find the parameter t at $(14, 11)$, then differentiate parametrically to get the tangent gradient, and take the negative reciprocal for the normal.

Since $x = 14$ and $y = 11$,

$$3t^2 + 2 = 14 \quad \text{and} \quad 6t - 1 = 11,$$

both giving $t = 2$ (and $t = 2 \geq \frac{1}{6}$, so this is a valid point on C).

Differentiating each equation with respect to t,

$$\frac{dx}{dt} = 6t, \qquad \frac{dy}{dt} = 6.$$

Using the chain rule,

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{6}{6t} = \frac{1}{t}.$$

At $t = 2$, the gradient of the tangent to C is $\frac{1}{2}$, so the gradient of the normal N (perpendicular to the tangent) is -2.

Since N passes through $(14, 11)$ with gradient -2,

$$y - 11 = -2(x - 14)$$
$$y - 11 = -2x + 28$$
$$2x + y = 39.$$

To check your answer: the tangent and normal gradients, $\frac{1}{2}$ and -2, multiply to -1; and $2(14) + 11 = 39$. ✓

$\therefore$ the equation of N is $2x + y = 39$.

Part (ii) 4 marks

APPROACH

The region bounded by C, N and the x-axis splits naturally at $x = 14$: for x between the start of C and 14 the upper boundary is the curve, and for x between 14 and where N meets the x-axis the upper boundary is the line. We find both end-points first.

Since $t \geq \frac{1}{6}$, the curve C starts where $t = \frac{1}{6}$:

$$x = 3\left(\frac{1}{6}\right)^2 + 2 = \frac{25}{12}, \qquad y = 6\left(\frac{1}{6}\right) - 1 = 0,$$

i.e. C starts on the x-axis, at $\left(\frac{25}{12},\ 0\right)$.

Note: both x and y increase with t, so C rises monotonically from $\left(\frac{25}{12}, 0\right)$ to $(14, 11)$.

The line N meets the x-axis where $y = 0$ in $2x + y = 39$:

$$2x = 39$$
$$x = \frac{39}{2}.$$

The enclosed region runs along the x-axis from $x = \frac{25}{12}$ to $x = \frac{39}{2}$, bounded above by C up to $x = 14$ and by N from $x = 14$ to $x = \frac{39}{2}$. Using $\int y \frac{dx}{dt}\, dt$ for the area under C,

$$\begin{aligned}\text{Area under } C &= \int_{1/6}^{2} (6t - 1)(6t)\, dt \\ &= \int_{1/6}^{2} \left(36t^2 - 6t\right) dt \\ &= \left[12t^3 - 3t^2\right]_{1/6}^{2} \\ &= 84 - \left(-\frac{1}{36}\right) \\ &= \frac{3025}{36}.\end{aligned}$$

The area under N, from $x = 14$ to $x = \frac{39}{2}$, is a right-angled triangle of base $\frac{39}{2} - 14 = \frac{11}{2}$ and height 11:

$$\text{Area under } N = \frac{1}{2} \cdot \frac{11}{2} \cdot 11 = \frac{121}{4}.$$

Adding the two areas,

$$\begin{aligned}\text{Total area} &= \frac{3025}{36} + \frac{121}{4} \\ &= \frac{3025}{36} + \frac{1089}{36} \\ &= \frac{2057}{18}.\end{aligned}$$

∴ the area enclosed by C, N and the x-axis is $\frac{2057}{18} \approx 114 \text{ units}^2$ (3 s.f.).

Part (iii)(a) 1 mark

Since a two-way stretch with scale factor 2 in the x-direction and scale factor 3 in the y-direction scales every area by $2 \times 3 = 6$,

$$\text{Area enclosed by } D,\ M \text{ and the } x\text{-axis} = 6 \times \frac{2057}{18} = \frac{2057}{3}.$$

∴ the area is $\frac{2057}{3} \approx 686 \text{ units}^2$ (3 s.f.).

Part (iii)(b) 2 marks

APPROACH

We first eliminate t to get the cartesian equation of C, then apply the stretch $x \mapsto 2x,\ y \mapsto 3y$ directly to that equation.

From $x = 3t^2 + 2$, $t^2 = \frac{x-2}{3}$. From $y = 6t - 1$, $t = \frac{y+1}{6}$, so $t^2 = \frac{(y+1)^2}{36}$. Equating,

$$\frac{(y+1)^2}{36} = \frac{x-2}{3}$$
$$(y+1)^2 = 12(x-2).$$

Under the stretch, a point (x, y) on C maps to $(X, Y) = (2x, 3y)$ on D, so $x = \frac{X}{2}$ and $y = \frac{Y}{3}$. Substituting,

$$\left(\frac{Y}{3} + 1\right)^2 = 12\left(\frac{X}{2} - 2\right)$$
$$\frac{(Y+3)^2}{9} = 6(X-4)$$
$$(Y+3)^2 = 54(X-4).$$

Note: since $y \geq 0$ on C (as $t \geq \frac{1}{6}$), the image satisfies $Y \geq 0$, so D is the upper branch of this curve.

∴ the cartesian equation of D is $(y+3)^2 = 54(x-4)$, with $y \geq 0$.

Question 4 11 marks

Differentiation · maximum volume

Part (i) 2 marks

APPROACH

We need two relations: how a and h are tied together by the 30 cm cardboard (Fig. 1), and how h, H and a are tied together inside the folded pyramid (Fig. 2).

From Fig. 1, the net's overall width is 30 cm, made up of the central square of side a together with two triangular flaps of height h on either side, so

$$a + 2h = 30$$
$$h = \frac{30 - a}{2}.$$

In the folded pyramid, h is the slant height from the apex to the midpoint of a base edge, H is the vertical height, and the horizontal distance from the centre of the square base to the midpoint of an edge is $\frac{a}{2}$. These three lengths form a right-angled triangle, so by Pythagoras' theorem,

$$h^2 = H^2 + \left(\frac{a}{2}\right)^2$$
$$H^2 = h^2 - \frac{a^2}{4}.$$

Substituting $h = \frac{30-a}{2}$,

$$\begin{aligned} H^2 &= \left(\frac{30 - a}{2}\right)^2 - \frac{a^2}{4} \\ &= \frac{(30 - a)^2 - a^2}{4} \\ &= \frac{900 - 60a}{4} \\ &= 225 - 15a. \quad \text{(shown)} \end{aligned}$$

Part (ii) 5 marks

> **APPROACH**
>
> With $H = \sqrt{225 - 15a}$ from (i), the volume is a function of a alone; we differentiate and solve $\frac{dV}{da} = 0$.

Since the volume of the pyramid is $\frac{1}{3} \times$ base area $\times$ height,

$$V = \frac{1}{3}a^2\sqrt{225 - 15a}, \qquad 0 < a < 15$$

(the upper bound comes from requiring $H^2 > 0$).

Differentiating using the product rule,

$$\begin{aligned} \frac{dV}{da} &= \frac{1}{3}\left[2a\sqrt{225 - 15a} + a^2 \cdot \frac{-15}{2\sqrt{225 - 15a}}\right] \\ &= \frac{a}{3\sqrt{225 - 15a}}\left[2(225 - 15a) - \frac{15a}{2}\right] \\ &= \frac{a}{3\sqrt{225 - 15a}}\left(450 - \frac{75a}{2}\right). \end{aligned}$$

Setting $\frac{dV}{da} = 0$: since $0 < a < 15$, $a \neq 0$ and $\sqrt{225 - 15a} \neq 0$, so

$$450 - \frac{75a}{2} = 0$$
$$a = 12.$$

At $a = 12$,

$$H^2 = 225 - 15(12) = 45, \qquad H = 3\sqrt{5},$$
$$V = \frac{1}{3}(12)^2\left(3\sqrt{5}\right) = 144\sqrt{5}.$$

∴ the maximum possible volume of the pyramid is $144\sqrt{5} \approx 322\ cm^3$ (3 s.f.).

Part (iii)

Part (a) 3 marks

APPROACH

The total surface area of the four triangular faces is $4 \times \frac{1}{2}ah = 2ah$; substitute $h = \frac{30-a}{2}$ from (i) to write it in terms of a alone, then maximise.

$$S = 2ah = 2a \cdot \frac{30 - a}{2}$$
$$= 30a - a^2.$$

Differentiating,

$$\frac{dS}{da} = 30 - 2a.$$

Setting $\frac{dS}{da} = 0$,

$$30 - 2a = 0$$
$$a = 15.$$

∴ the total surface area of the four triangular faces is a maximum when $a = 15$.

Note: at $a = 15$, $H^2 = 225 - 15(15) = 0$, so this is exactly the boundary value at which the pyramid degenerates — see (b).

Part (b) 1 mark

Since $H = 0$ when $a = 15$, the solid has zero height: each triangular flap folds all the way down onto the base square instead of standing up. At $a = 15$, $h = \frac{30-15}{2} = 7.5 = \frac{a}{2}$, which is exactly the distance from the centre of the square base to the midpoint of an edge — so each folded-down triangle reaches exactly to the centre, and all four meet there.

∴ the net folds completely flat: no pyramid is formed at all, only a flat square shape (of diagonal 30 cm, whose corners are the midpoints of the sides of the original 30 cm cardboard square).

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

Question 5 10 marks

Discrete random variables · expectation and variance

Part (i) 1 mark

The bag has one green disc (worth 0), r red discs (worth 5 each) and $2r$ blue discs (worth 2 each). Tina's score is the product of the two discs' point values, so the possible pairs and their scores are: green with anything gives 0; red with red gives $5 \times 5 = 25$; red with blue gives $5 \times 2 = 10$; blue with blue gives $2 \times 2 = 4$.

$\therefore$ Tina's possible scores are $0,\ 4,\ 10,\ 25$.

Part (ii) 7 marks

APPROACH

There are $3r + 1$ discs in total, so we find the probability of each score by counting unordered pairs of discs, using combinations. Writing every probability over the common denominator $3(3r + 1)$ makes the sum $E(X)$ come out as a single fraction that matches the given expression.

Let X be Tina's score. The total number of ways to take 2 discs from the $3r + 1$ discs is $\binom{3r+1}{2} = \frac{(3r+1)(3r)}{2}$.

Counting the ways to get each score:

$$P(X = 25) = \frac{\binom{r}{2}}{\binom{3r+1}{2}} = \frac{r(r-1)}{(3r+1)(3r)} = \frac{r-1}{3(3r+1)}$$

$$P(X = 10) = \frac{r \cdot 2r}{\binom{3r+1}{2}} = \frac{2r^2}{\frac{1}{2}(3r+1)(3r)} = \frac{4r}{3(3r+1)}$$

$$P(X = 4) = \frac{\binom{2r}{2}}{\binom{3r+1}{2}} = \frac{2r(2r-1)}{(3r+1)(3r)} = \frac{2(2r-1)}{3(3r+1)}$$

$$P(X = 0) = \frac{1 \cdot 3r}{\binom{3r+1}{2}} = \frac{3r}{\frac{1}{2}(3r+1)(3r)} = \frac{2}{3r+1}$$

To check your answer: the four probabilities should sum to 1. Writing $P(X = 0) = \frac{6}{3(3r+1)}$, the numerators are $(r - 1) + 4r + 2(2r - 1) + 6 = 9r + 3 = 3(3r + 1)$, so the sum is $\frac{3(3r+1)}{3(3r+1)} = 1$. ✓

Then

$$\begin{aligned} &E(X) \\ &= 0 \cdot P(X = 0) + 4 \cdot P(X = 4) + 10 \cdot P(X = 10) + 25 \cdot P(X = 25) \\ &= \frac{4 \cdot 2(2r-1) + 10 \cdot 4r + 25(r-1)}{3(3r+1)} \\ &= \frac{(16r - 8) + 40r + (25r - 25)}{3(3r+1)} \end{aligned}$$

$$= \frac{81r - 33}{3(3r+1)}$$
$$= \frac{27r - 11}{3r+1} \quad \text{(shown)}$$

For the variance we also need $E\left(X^2\right)$:

$$E\left(X^2\right)$$
$$= 16 \cdot P(X=4) + 100 \cdot P(X=10) + 625 \cdot P(X=25)$$
$$= \frac{16 \cdot 2(2r-1) + 100 \cdot 4r + 625(r-1)}{3(3r+1)}$$
$$= \frac{(64r - 32) + 400r + (625r - 625)}{3(3r+1)}$$
$$= \frac{1089r - 657}{3(3r+1)}$$
$$= \frac{3(121r - 73)}{3r+1}$$

Then $Var(X) = E\left(X^2\right) - [E(X)]^2$:

$$Var(X) = \frac{3(121r-73)}{3r+1} - \left(\frac{27r-11}{3r+1}\right)^2$$
$$= \frac{3(121r-73)(3r+1) - (27r-11)^2}{(3r+1)^2}$$
$$= \frac{\left(1089r^2 - 294r - 219\right) - \left(729r^2 - 594r + 121\right)}{(3r+1)^2}$$
$$= \frac{360r^2 + 300r - 340}{(3r+1)^2}$$
$$\therefore\ Var(X) = \frac{20\left(18r^2 + 15r - 17\right)}{(3r+1)^2}$$

Part (iii) 2 marks

APPROACH

setting the variance expression from (ii) equal to 38 gives a quadratic in r; since $r > 1$, only one root can be accepted.

$$\frac{20\left(18r^2 + 15r - 17\right)}{(3r+1)^2} = 38$$
$$360r^2 + 300r - 340 = 38(3r+1)^2$$
$$360r^2 + 300r - 340 = 342r^2 + 228r + 38$$
$$18r^2 + 72r - 378 = 0$$
$$r^2 + 4r - 21 = 0$$
$$(r+7)(r-3) = 0$$
$$r = -7 \ \text{(rej. as } r > 1), \quad r = 3$$

To check your answer: with $r = 3$ there are $1 + 3 + 6 = 10$ discs; enumerating all $\binom{10}{2} = 45$ equally likely pairs of discs directly gives $E(X) = 7$ (matching $\frac{27(3)-11}{3(3)+1} = \frac{70}{10} = 7$) and $Var(X) = 38$. ✓

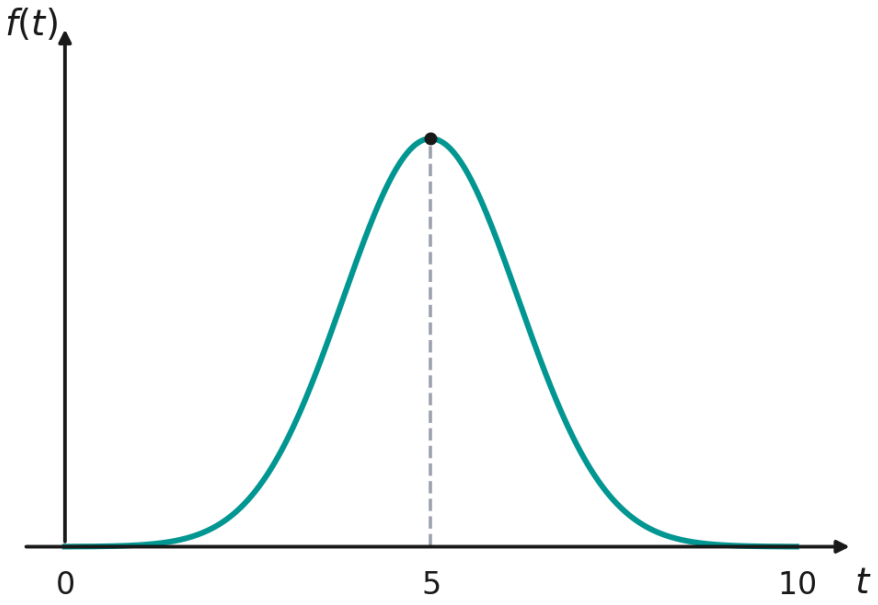
$\therefore r = 3.$

Question 6 10 marks

Normal distribution · sums of normal variables

Part (i) 2 marks

Since $T \sim N\left(5,\ 1.2^2\right)$, the density is a symmetric bell curve centred at $t = 5$, with the period from 8 am to 8.10 am corresponding to $t \in [0, 10]$.



Part (ii) 1 mark

James leaves later than 8.06 am means $T > 6$.

$$P(T > 6) = 0.202 \text{ (by GC)}$$

$\therefore P(T > 6) = 0.202$, correct to 3 s.f.

Part (iii) 2 marks

APPROACH

James leaves at T minutes past 8 am and then walks for W minutes, so his arrival time is $T + W$ minutes past 8 am. He is late if this exceeds 30 (i.e. past 8.30 am). Since T and W are independent normal random variables, $T + W$ is also normal, with mean and variance found by adding.

Let $T + W$ be James's arrival time, in minutes past 8 am, on a day he walks. Since $T \sim N\left(5,\ 1.2^2\right)$ and $W \sim N\left(21,\ 3^2\right)$ are independent,

$$\begin{aligned} T + W &\sim N\left(5 + 21,\ 1.2^2 + 3^2\right) \\ &= N(26,\ 10.44) \end{aligned}$$

James is late if $T + W > 30$, so

$$P(T + W > 30) = 0.108 \text{ (by GC)}$$

$\therefore$ the probability that James is late for work on a day he walks is 0.108, correct to 3 s.f.

Part (iv) 5 marks

APPROACH

we are asked for $P\left(\text{fine} \mid \text{late}\right)$, a reversed conditional probability, so we use Bayes' theorem. This needs $P\left(\text{late}\right)$ overall, which splits into the walking case (already found in (iii)) and a driving case that we find the same way.

Let F be the event that the weather is fine on a given day, and L be the event that James is late for work that day. We are given $P(F) = 0.7$, so $P(F') = 0.3$.

On a day the weather is not fine, James drives, arriving $T + D$ minutes past 8 am, where $T \sim N\left(5,\ 1.2^2\right)$ and $D \sim N\left(19,\ 6^2\right)$ are independent, so

$$\begin{aligned} T + D &\sim N\left(5 + 19,\ 1.2^2 + 6^2\right) \\ &= N(24,\ 37.44) \end{aligned}$$

giving

$$P(L \mid F') = P(T + D > 30) = 0.163 \text{ (by GC)}$$

From part **(iii)**, $P(L \mid F) = 0.10786\ldots$ (unrounded). By the law of total probability,

$$\begin{aligned} &P(L) \\ &= P(L \mid F)P(F) + P(L \mid F')P(F') \\ &= (0.10786\ldots)(0.7) + (0.16340\ldots)(0.3) \\ &= 0.12452\ldots \end{aligned}$$

By Bayes' theorem,

$$\begin{aligned} &P(F \mid L) \\ &= \frac{P(L \mid F)P(F)}{P(L)} \\ &= \frac{(0.10786\ldots)(0.7)}{0.12452\ldots} \\ &= 0.606\ldots \end{aligned}$$

$\therefore P(F \mid L) = 0.606$, correct to 3 s.f.

Question 7 8 marks

Correlation and regression · transformed models

Part (i) 2 marks



$$r = -0.848 \text{ (by GC)}$$

$\therefore$ the product moment correlation coefficient between d and t is -0.848, correct to 3 s.f.

Part (ii) 5 marks

APPROACH

the model $d = -a + bu$ is linear in u, so we regress d on $u = \frac{1}{t}$ (not on t) to find a and b.

Using $u = \frac{1}{t}$, the data become

u	$\frac{1}{32}$	$\frac{1}{41}$	$\frac{1}{50}$	$\frac{1}{59}$	$\frac{1}{68}$
d	172	57	27	19	14

By GC, the regression line of d on u is

$$d = -145 + 9460u \text{ (by GC)}$$

so $-a = -145$ and $b = 9460$, giving

$$a = 145, \qquad b = 9460$$

both correct to 3 s.f.

$$r = 0.941 \text{ (by GC)}$$

Physics. Math. Chemistry.

$\therefore a = 145$, $b = 9460$, and the product moment correlation coefficient between d and u is 0.941, all correct to 3 s.f.

Part (iii) 1 mark

At $t = 86$, using Lim's model,

$$\begin{aligned} d &= -145 + \frac{9460}{86} \\ &= -35.0 \ (3 \text{ s.f.}) \end{aligned}$$

This is negative, which is meaningless for a number of days, and very far from the actual value 32.

Explanation: $t = 86$ lies well outside the range of the given data ($32 \leq t \leq 68$), so using the model here is an extrapolation; a negative predicted number of days shows the model does not fit this additional data, even though $|r| = 0.941$ is close to 1 within the original data range.

$\therefore$ Lim's model does not fit the additional data.

Part (iv)

APPROACH

we substitute t in terms of T into the equation from (ii). Rearranging $C = \frac{5}{9}(F - 32)$ for F (with $F = t$ and $C = T$ here) gives t in terms of T, which we substitute into $d = -a + bu = -a + \frac{b}{t}$.

Since $T = \frac{5}{9}(t - 32)$,

$$t = \frac{9}{5}T + 32$$

Substituting into $d = -145 + \frac{9460}{t}$ from part **(ii)**,

$$d = -145 + \frac{9460}{\frac{9}{5}T + 32}$$

$\therefore d = -145 + \frac{9460}{\frac{9}{5}T + 32}$.

thelearningcave.com.sg
Physics. Math. Chemistry.

Question 8 7 marks

Probability · selections · discrete random variables

Part (i) 2 marks

APPROACH

A direct comparison of $P(R=1)$ and $P(R=2)$ as decimals hides the structure. Since both probabilities share the same denominator $\binom{28}{12}$, it is quicker to form the ratio $P(R=2) \div P(R=1)$ and show it exceeds 1.

Let R be the number of rectangles chosen. Since 12 shapes are chosen from 11 circles and 17 rectangles, R has a hypergeometric distribution, and

$$P(R=1) = \frac{\binom{17}{1}\binom{11}{11}}{\binom{28}{12}}, \quad P(R=2) = \frac{\binom{17}{2}\binom{11}{10}}{\binom{28}{12}}.$$

The two probabilities have the same denominator, so we compare numerators. Simplifying:

$$\begin{aligned} &\frac{P(R=2)}{P(R=1)} \\ &= \frac{\binom{17}{2}\binom{11}{10}}{\binom{17}{1}\binom{11}{11}} \\ &= \frac{\binom{17}{2}}{\binom{17}{1}} \times \frac{\binom{11}{10}}{\binom{11}{11}} \\ &= \frac{136}{17} \times \frac{11}{1} \\ &= 8 \times 11 \\ &= 88. \end{aligned}$$

Since $\frac{P(R=2)}{P(R=1)} = 88 > 1$ and $P(R=1) > 0$, we have $P(R=2) > P(R=1)$, i.e. $P(R=1) < P(R=2)$. (shown)

To check your answer: By GC, $P(R=1) = \frac{17}{30421755} \approx 5.59 \times 10^{-7}$ and $P(R=2) = \frac{1496}{30421755} \approx 4.92 \times 10^{-5}$, and indeed $88 \times 5.59 \times 10^{-7} \approx 4.92 \times 10^{-5}$. ✓

Part (ii) 5 marks

APPROACH

With r extra rectangles, the total number of shapes and the size of the choice both change, but $\binom{28+r}{12}$ again appears in every probability and cancels out of the given ratio, leaving an equation in r alone.

With the number of rectangles increased by r, there are $(17+r)$ rectangles and 11 circles, so

$$P(R=3) = \frac{\binom{17+r}{3}\binom{11}{9}}{\binom{28+r}{12}}, \quad P(R=4) = \frac{\binom{17+r}{4}\binom{11}{8}}{\binom{28+r}{12}}.$$

Using the given condition $P(R=4) = 15\,P(R=3)$, the common denominator $\binom{28+r}{12}$ cancels:

$$\binom{17+r}{4}\binom{11}{8} = 15\binom{17+r}{3}\binom{11}{9}$$

$$\Rightarrow \frac{\binom{17+r}{4}}{\binom{17+r}{3}} = 15 \times \frac{\binom{11}{9}}{\binom{11}{8}}.$$

Now $\dfrac{\binom{17+r}{4}}{\binom{17+r}{3}} = \dfrac{14+r}{4}$, and $\binom{11}{9} = 55$, $\binom{11}{8} = 165$, so $\dfrac{\binom{11}{9}}{\binom{11}{8}} = \dfrac{1}{3}$. Substituting:

$$\frac{14+r}{4} = 15 \times \frac{1}{3}$$

$$= 5.$$

$$14 + r = 20$$

$$r = 6.$$

$\therefore r = 6.$

Question 9 12 marks

Binomial distribution · probabilities in context

Part (i) 1 mark

A random sample in this context means that each of the 10 pens tested is selected from the box of 100 pens in such a way that every pen has an equal chance of being chosen, and the pens are chosen independently of one another (so that whether one pen is faulty does not affect the chance that another is chosen).

Part (ii) 1 mark

Let X be the number of faulty pens in a random sample of 10 pens from a box. Since each pen is independently faulty with probability 0.06,

$$X \sim B(10,\ 0.06).$$

The box is accepted if 2 or fewer faulty pens are found, so

$$P\,(\text{box accepted})$$

$$= P(X \leq 2)$$

$$= 0.981 \ \ (\text{by GC}).$$

$\therefore P\,(\text{box accepted}) = 0.981.$

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

Part (iii) 4 marks

> **APPROACH**
>
> "More than 5% of these boxes are rejected" is a statement about the number of boxes out of 75, so we need a second binomial variable, Y, counting rejected boxes among the 75, with success probability $P\,(\text{reject}) = 1 - P\,(\text{accept})$ found in part (ii).

From part (ii), the probability that a box is rejected is

$$P\,(\text{reject}) = 1 - 0.9811621... = 0.0188378... \;\; (\text{by GC, unrounded value kept}).$$

Let Y be the number of rejected boxes among the 75 boxes tested. Since the boxes are tested independently,

$$Y \sim B(75,\ 0.0188378...).$$

5% of 75 boxes is 3.75 boxes, so "more than 5% of these boxes are rejected" means $Y > 3.75$, i.e. $Y \geq 4$. Converting the tail to cumulative form,

$$\begin{aligned} & P(Y \geq 4) \\ \Rightarrow\ & 1 - P(Y \leq 3) \\ = \ & 1 - 0.9465 \\ = \ & 0.0535 \;\; (\text{by GC}). \end{aligned}$$

$\therefore P\,(\text{more than } 5\% \text{ of boxes rejected}) = 0.0535.$

Part (iv) 5 marks

> **APPROACH**
>
> The alternative procedure has three outcomes after the first sample of 5 (accept outright, reject outright, or take a second sample), so $P\,(\text{accept})$ is built up as a sum over these cases rather than as a single binomial tail.

Let W be the number of faulty pens in the first sample of 5 pens, so $W \sim B(5,\ 0.06)$. The box is accepted immediately if $W = 0$. If $W = 1$ or $W = 2$, a second sample of 5 pens is taken; let V be the number of faulty pens in this second sample, so $V \sim B(5,\ 0.06)$ independently of W, and the box is accepted if $W + V \leq 2$.

- If $W = 1$, the box is accepted when $V \leq 1$.
- If $W = 2$, the box is accepted when $V \leq 0$.

By GC,

$$P(W = 0) = 0.733904, \;\; P(W = 1) = 0.234225, \;\; P(W = 2) = 0.0299010,$$
$$P(V \leq 1) = 0.968129, \;\; P(V \leq 0) = 0.733904.$$

Therefore

$$\begin{aligned} & P\,(\text{box accepted}) \\ = \ & P(W = 0) + P(W = 1)P(V \leq 1) + P(W = 2)P(V \leq 0) \end{aligned}$$

$$= 0.733904 + (0.234225)(0.968129) + (0.0299010)(0.733904)$$
$$= 0.983 \ \text{(by GC)}.$$

$\therefore P(\text{box accepted under alternative procedure}) = 0.983.$

Part (v) 1 mark

Explanation: Under the alternative procedure a second sample of 5 pens is only needed when the first sample gives 1 or 2 faulty pens; otherwise testing stops after just 5 pens. The expected number of pens tested per box is therefore

$$5 + 5 \times P(W = 1 \text{ or } W = 2) = 5 + 5(0.264126) \approx 6.32,$$

which is fewer than the 10 pens always tested under the original procedure, while the probability of accepting a good box (0.983) is about the same as before (0.981). The factory manager might prefer the alternative procedure because it tests fewer pens on average, reducing the cost, time and product wasted in testing, without materially changing the chance that a box is correctly accepted.

Question 10 13 marks

Hypothesis testing · test of a mean

Part (i) 4 marks

Let $X\%$ be the percentage of carbon, by weight, in a round steel bar, so $X \sim N\left(1.5,\ 0.09^2\right)$. The production manager tests

$$H_0 : \mu = 1.5 \quad \text{against} \quad H_1 : \mu \neq 1.5$$

at the 5% level of significance, where μ is the population mean percentage of carbon in round steel bars. Under H_0, since X is normally distributed with known variance 0.09^2, the sample mean of $n = 15$ bars is exactly normal:

$$\overline{X} \sim N\left(1.5,\ \frac{0.09^2}{15}\right).$$

Standardising, under H_0 the test statistic

$$Z = \frac{\overline{X} - 1.5}{0.09/\sqrt{15}} \sim N(0,\ 1).$$

Since this is a two-tailed test at the 5% level, H_0 is rejected if $|Z| > 1.9600$ (by GC, the upper 2.5% point of $N(0, 1)$), i.e.

$$\left|\overline{X} - 1.5\right| > 1.9600 \times \frac{0.09}{\sqrt{15}}$$
$$= 0.0455 \ \text{(by GC, unrounded } 0.045546\ldots).$$

So the critical region is $\overline{X} < 1.4544\ldots$ or $\overline{X} > 1.5455\ldots$, i.e.

$$\overline{x} < 1.45 \quad \text{or} \quad \overline{x} > 1.55 \ \text{(to 3 s.f.)}.$$

∴ the critical region is $\overline{X} < 1.45$ or $\overline{X} > 1.55$.

Part (ii) 2 marks

Explanation: In the earlier test, the percentage of carbon in round steel bars is given to be normally distributed, so the sample mean $\overline{X}$ is exactly normally distributed whatever the sample size, and a small sample of 15 bars is enough for the test to be valid. For the new flat bars, the distribution of the percentage of carbon is not stated to be normal, so the production manager needs a large sample (here $n = 40$) so that, by the Central Limit Theorem, the sample mean $\overline{X}$ is approximately normally distributed even though the underlying distribution of carbon percentage in a single bar is unknown.

Part (iii) 2 marks

Since $n = 40$, $\Sigma x = 10.16$ and $\Sigma x^2 = 2.586342$, the unbiased estimate of the population mean is

$$\overline{x} = \frac{\Sigma x}{n} = \frac{10.16}{40} = 0.254.$$

The unbiased estimate of the population variance is

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right) \\ &= \frac{1}{39}\left(2.586342 - \frac{10.16^2}{40}\right) \\ &= 0.000146 \ \ (3 \text{ s.f., unrounded } 0.0001462\ldots). \end{aligned}$$

∴ the unbiased estimates are $\overline{x} = 0.254$ and $s^2 = 0.000146$.

Part (iv) 5 marks

Let μ be the population mean percentage of carbon, by weight, in flat bars. The production manager tests

$$H_0 : \mu = 0.25 \quad \text{against} \quad H_1 : \mu > 0.25$$

at the $2\frac{1}{2}\%$ level of significance.

APPROACH

The underlying distribution of carbon percentage in a flat bar is not given as normal, but $n = 40$ is large, so we use the Central Limit Theorem (part (ii)) with the unbiased variance estimate from part (iii) in place of the population variance, rather than an exact normal test.

Since $n = 40$ is large, by the Central Limit Theorem, under H_0,

$$\overline{X} \sim N\left(0.25, \frac{s^2}{n}\right) \text{ approximately, where } s^2 = 0.0001462\ldots \text{ from part (iii)}.$$

Using the given sample data,

$$z$$

$$
\begin{aligned}
&= \frac{\overline{x} - 0.25}{\sqrt{s^2/n}} \\
&= \frac{0.254 - 0.25}{\sqrt{0.0001462.../40}} \\
&= 2.09 \ \ (\text{by GC, unrounded } 2.0922...).
\end{aligned}
$$

By GC, the p-value is

$$P(Z \geq 2.0922...) = 0.0182.$$

Since $0.0182 < 0.025$, i.e. the test statistic $z = 2.09$ lies beyond the critical value $z = 1.960$ (the upper $2\frac{1}{2}\%$ point of $N(0,1)$), we reject H_0 at the $2\frac{1}{2}\%$ level of significance.

To check your answer: the sample mean 0.254 is only 0.004 above 0.25, but the standard error $\sqrt{s^2/n} \approx 0.00191$ is small, so even this small difference is several standard errors out — consistent with rejecting H_0. ✓

There is sufficient evidence, at the $2\frac{1}{2}\%$ level of significance, that the mean percentage of carbon in the flat bars is more than 0.25%, supporting the production manager's suspicion that too much carbon has been added.

thelearningcave.com.sg
Physics. Math. Chemistry.