IB Math AI HL — Topic 4

Random Variables, Linear Combinations & Unbiased Estimates

Discrete Distributions · E(X) & Var(X) · Linear Combinations · Unbiased Estimates s_{n-1}^2

Difficulty	EASY	MEDIUM	HARD
Questions	10	10	10
Total Marks	30	40	50

Grand Total: **120 marks**

Scan me for help

Name:______________________________ Date:_______________

Score:__________/120

Instructions: Answer all questions. Show all working. Give answers correct to 3 significant figures unless stated. A GDC is permitted throughout.

Key Formulae — Random Variables & Unbiased Estimates

Expected Value (discrete):

$$\text{E}(X) = \sum x\, P(X = x)$$

Variance (discrete):

$$\text{Var}(X) = \text{E}(X^2) - [\text{E}(X)]^2$$

Linear scaling:

$$\text{E}(aX + b) = a\,\text{E}(X) + b$$

$$\text{Var}(aX + b) = a^2\,\text{Var}(X)$$

Sum/difference of *independent* RVs:

$$\text{E}(X \pm Y) = \text{E}(X) \pm \text{E}(Y)$$

$$\text{Var}(aX \pm bY) = a^2\,\text{Var}(X) + b^2\,\text{Var}(Y)$$

Unbiased estimate of population mean:

$$\bar{x} = \frac{\sum x_i}{n}$$

Unbiased estimate of population variance:

$$s_{n-1}^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

GDC Tips

- **1-Var Stats**: Enter data in a list, then `STAT` > `CALC` > `1-Var Stats`. $\bar{x}$ gives the sample mean; S_x gives s_{n-1} (the unbiased SD); σ_x gives s_n (population SD). **Use S_x for inference questions.**
- **Probability distributions**: For a custom discrete distribution, enter values in L1 and probabilities in L2, then compute E(X) = `sum(L1*L2)` and E(X^2) = `sum(L1^2*L2)`.
- **Linear combinations**: The GDC cannot compute Var$(aX + bY)$ directly — always apply the formula by hand.

Section A — Easy EASY (Q1–Q10, 3 marks each = 30 marks)

1. EASY Calculator **[3 marks]**

The table below gives the probability distribution of a discrete random variable X.

x	1	2	3	4
$P(X = x)$	0.15	0.35	0.40	p

(a) Write down the value of p. [1]

(b) Find E(X). [1]

(c) Find Var(X). [1]

2. EASY Calculator **[3 marks]**

A random variable X has E$(X) = 5$ and Var$(X) = 4$.
Let $Y = 3X - 2$.

(a) Find E(Y). [1]

(b) Find Var(Y). [1]

(c) Write down the standard deviation of Y. [1]

3. **EASY** **Calculator** **[3 marks]**

A random variable X has E$(X) = 8$ and Var$(X) = 9$.
A second random variable Y, independent of X, has E$(Y) = 3$ and Var$(Y) = 4$.
Find:

(a) E$(X + Y)$ [1]

(b) Var$(X + Y)$ [1]

(c) Var$(X - Y)$ [1]

4. **EASY** **Calculator** **[3 marks]**

The table below gives the probability distribution of a discrete random variable X.

x	0	2	5
$P(X = x)$	0.50	0.30	0.20

Find E(X) and E(X^2), and hence find Var(X).

5. **EASY** **Calculator** **[3 marks]**

A sample of six measurements is taken:

$$12, \quad 15, \quad 14, \quad 18, \quad 11, \quad 16$$

(a) Write down the sample mean $\bar{x}$. [1]

(b) Calculate the unbiased estimate of the population variance, s^2_{n-1}. Give your answer correct to 3 significant figures. [2]

6. **EASY** **Calculator** **[3 marks]**

A random variable X has E$(X) = 7$ and Var$(X) = 3$.
Find:

(a) E$(2X)$ [1]

(b) E(X^2) [1]

(c) Var$(5 - X)$ [1]

7. EASY Calculator **[3 marks]**

The table below gives the probability distribution of X, the daily profit (in hundreds of dollars) earned by a market stall.

Profit x ($100s)	-1	0	2	4
$P(X = x)$	0.10	0.25	0.45	0.20

(a) Write down $P(X = 0)$. [1]

(b) Find the expected daily profit. [2]

8. EASY Calculator **[3 marks]**

Two independent random variables X and Y satisfy

$$\mathsf{E}(X) = 10, \quad \mathsf{Var}(X) = 6, \quad \mathsf{E}(Y) = 4, \quad \mathsf{Var}(Y) = 2.$$

Find:

(a) $\mathsf{E}(X - Y)$ [1]

(b) $\mathsf{Var}(2X)$ [1]

(c) $\mathsf{Var}(X + Y)$ [1]

9. EASY Calculator **[3 marks]**

A sample of five exam scores is: $\{62,\ 70,\ 68,\ 75,\ 65\}$.

(a) Find the sample mean $\bar{x}$. [1]

(b) Hence find s_{n-1}^2, the unbiased estimate of the population variance. Give your answer correct to 3 significant figures. [2]

10. EASY Calculator **[3 marks]**

A discrete random variable X has E$(X) = 4$ and Var$(X) = 5$.
A new variable $W = -2X + 1$ is defined.

(a) Write down E(W). [1]

(b) Write down Var(W). [1]

(c) Write down the standard deviation of W. [1]

Section B — Medium MEDIUM (Q11–Q20, 4 marks each = 40 marks)

11. MEDIUM Calculator **[4 marks]**

The table below gives the incomplete probability distribution of a discrete random variable X.

x	1	2	3	4	5
$P(X = x)$	a	0.20	$2a$	0.15	0.10

(a) Find the value of a. [2]

(b) Find E(X) and Var(X). [2]

12. MEDIUM Calculator **[4 marks]**

A small factory produces components whose lengths X mm are modelled as a discrete random variable with E$(X) = 50$ and Var$(X) = 16$.

Quality control multiplies each measurement by a constant factor of 0.5 and adds 3 to obtain a scaled measurement $W = 0.5X + 3$.

(a) Find E(W) and Var(W). [2]

(b) Two independent components are chosen. Let $T = X_1 + X_2$ be the total length. Find E(T) and Var(T). [2]

13. MEDIUM Calculator **[4 marks]**

A nutritionist records the daily protein intake (in grams) for a sample of seven athletes:

$$155, \quad 162, \quad 148, \quad 170, \quad 159, \quad 163, \quad 151$$

(a) Calculate $\bar{x}$, the sample mean. [1]

(b) Calculate s_{n-1}^2, the unbiased estimate of the population variance. Give your answer correct to 3 significant figures. [2]

(c) State what s_{n-1}^2 estimates and why it is preferred to s_n^2 for this purpose. [1]

14. MEDIUM Calculator **[4 marks]**

X and Y are independent random variables with E$(X) = 6,$ Var$(X) = 4,$ E$(Y) = 9,$ Var$(Y) = 9.$
Let $T = 2X - Y + 5.$

(a) Show that E$(T) = 8.$ [2]

(b) Find Var$(T).$ [2]

15. MEDIUM Calculator **[4 marks]**

The discrete random variable X has the probability distribution shown in the table.

x	0	1	3	6
$P(X = x)$	0.20	0.30	0.40	0.10

(a) Calculate E(X) and Var(X). [3]

(b) Hence find the standard deviation of $3X - 4$. [1]

16. MEDIUM Calculator **[4 marks]**

A game at a school fair uses a spinner with sectors labelled $1, 2, 3$. The probability of landing on sector k is proportional to k. A player pays $1.50 to spin. The prize paid out equals the number on the sector in dollars.
Let X be the number on the sector.

(a) Write down the probability distribution of X. [2]

(b) Find the expected profit per spin for the organiser. [2]

17. MEDIUM Calculator **[4 marks]**

X_1, X_2, X_3 are three independent observations of a random variable X with $\text{E}(X) = \mu$ and $\text{Var}(X) = \sigma^2$. Two estimators of μ are proposed:

$$T_1 = \frac{X_1 + X_2 + X_3}{3}, \qquad T_2 = \frac{X_1 + 2X_2 + X_3}{4}.$$

(a) Show that T_1 is an unbiased estimator of μ. [2]

(b) Determine whether T_2 is unbiased. Justify your answer. [2]

18. MEDIUM Calculator **[4 marks]**

A courier firm delivers two types of parcels each day. The weight of a type-A parcel is modelled by the random variable A with $\text{E}(A) = 3.2$ kg and $\text{Var}(A) = 0.49$ kg^2. The weight of a type-B parcel is modelled by the random variable B with $\text{E}(B) = 1.8$ kg and $\text{Var}(B) = 0.16$ kg^2. The two types are independent.

(a) A van carries one type-A and one type-B parcel. Find the expected total weight and the standard deviation of the total weight. [2]

(b) A second delivery consists of 3 type-A parcels. Find the variance of the total weight of these three parcels. [2]

19. MEDIUM Calculator **[4 marks]**

Eight temperature readings (in °C) recorded at a weather station are:

$$21.4, \quad 22.1, \quad 20.8, \quad 23.0, \quad 21.9, \quad 22.6, \quad 20.5, \quad 22.3$$

(a) Find $\bar{x}$ and s_{n-1}, the unbiased standard deviation. Give s_{n-1} correct to 3 significant figures. [2]

(b) A new probe is calibrated so that each reading is multiplied by 0.1 and then increased by 273.15, converting to kelvin. Write down the mean and variance of the converted readings. [2]

20. MEDIUM Calculator **[4 marks]**

The number of errors X found per page when proofreading a manuscript is modelled by the probability distribution in the table below.

x	0	1	2	3
$P(X = x)$	0.40	0.35	0.20	0.05

(a) Find E(X) and Var(X). [2]

(b) A chapter has 12 independent pages. Let S be the total number of errors in the chapter. Write down E(S) and Var(S). [2]

Section C — Hard HARD (Q21–Q30, 5 marks each = 50 marks)

21. HARD Calculator **[5 marks]**

The discrete random variable X has the following probability distribution, where k is a positive constant.

x	1	2	3	4
$P(X = x)$	k	$2k$	$3k$	$4k$

(a) Find the value of k. [1]

(b) Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$, giving exact values. [3]

(c) Y is an independent random variable with $\mathrm{E}(Y) = \mathrm{E}(X)$ and $\mathrm{Var}(Y) = 2\,\mathrm{Var}(X)$. Find the standard deviation of $X + 2Y$. [1]

22. HARD Calculator **[5 marks]**

A random variable X has E$(X) = \mu$ and Var$(X) = \sigma^2$.
Three independent observations X_1, X_2, X_3 are taken. Consider the estimator

$$\hat{\mu} = \frac{1}{6}X_1 + \frac{1}{3}X_2 + cX_3.$$

(a) Find the value of c that makes $\hat{\mu}$ an unbiased estimator of μ. [2]

(b) Using this value of c, find Var$(\hat{\mu})$ in terms of σ^2. [2]

(c) Determine whether $\hat{\mu}$ or the estimator $\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3)$ has smaller variance. Justify your answer. [1]

23. HARD Calculator **[5 marks]**

A factory produces steel rods whose lengths are modelled by a random variable L cm with E$(L) = 30$ and Var$(L) = 0.25$.

A second machine trims each rod by an independent amount T cm where E$(T) = 2$ and Var$(T) = 0.04$. The trimmed rod has length $R = L - T$.

(a) Find E(R) and Var(R). [2]

(b) Five trimmed rods are packaged together. Let $S = R_1 + R_2 + R_3 + R_4 + R_5$ be the total length. Find E(S) and Var(S). [2]

(c) A crate can hold rods with a combined length of at most 145.5 cm. Find the value v such that E$(S) = v$ and comment on whether the crate is likely to be overfilled. [1]

24. HARD Calculator **[5 marks]**

The discrete random variable X has the probability distribution given below, where m is an unknown constant.

x	0	1	2	3	4
$P(X = x)$	0.10	m	0.30	$2m$	0.10

(a) Find m. [1]

(b) Find E(X) and Var(X). [2]

(c) Y is independent of X with E$(Y) = 3$ and Var$(Y) = 2$. Find the values of a and b such that E$(aX - bY) = 0$ and Var$(aX - bY) = 26$, given that $a, b > 0$. [2]

25. HARD Calculator **[5 marks]**

A marine biologist collects a random sample of n observations of fish lengths (in cm). The sample mean is $\bar{x}_1 = 26.0$ and the unbiased variance estimate is $s^2_{n-1,1} = 9.0$.

A second biologist independently collects 4 observations:

$$22.0, \quad 24.0, \quad 26.0, \quad 28.0$$

(a) For the second sample, find $\bar{x}_2$ and $s^2_{n-1,2}$, the unbiased variance estimate. [2]

(b) The combined sample mean of all $n+4$ observations is $\overline{x}_{\text{combined}} = 25.6$. Find the value of n. [2]

(c) State one reason why $s^2_{n-1,1}$ may not be a reliable estimate of the population variance. [1]

26. HARD Calculator **[5 marks]**

Two independent random variables X and Y have the same distribution with $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

(a) Show that $\mathrm{E}(X - Y) = 0$ and find $\mathrm{Var}(X - Y)$. [2]

(b) A third random variable $D = aX - aY$ where $a > 0$. Find the value of a such that $\mathrm{Var}(D) = 16\sigma^2$. [2]

(c) Using your answer, find $\mathrm{E}(D^2)$. [1]

27. HARD Calculator **[5 marks]**

Ten students sit an admissions test. Their raw scores are:

$$55, \quad 62, \quad 48, \quad 71, \quad 58, \quad 66, \quad 53, \quad 60, \quad 73, \quad 54$$

(a) Find $\bar{x}$ and s_{n-1}^2. Give s_{n-1}^2 correct to 3 significant figures. [2]

(b) The scores are standardised using the transformation $Z = \dfrac{X - \bar{x}}{s_{n-1}}$. Show that the mean of the Z-scores is 0. [1]

(c) Find the unbiased variance estimate of the Z-scores. Justify your answer without computing all ten values. [2]

28. HARD Calculator **[5 marks]**

Two independent random variables X and Y satisfy

$$\mathrm{E}(X) = \mu, \quad \mathrm{Var}(X) = 9, \quad \mathrm{E}(Y) = \mu, \quad \mathrm{Var}(Y) = 4.$$

Let $W = aX + bY$, where a and b are constants with $a + b = 1$.

(a) Show that W is an unbiased estimator of μ for any values of a and b satisfying $a + b = 1$. [2]

(b) Show that $\mathrm{Var}(W) = 9a^2 + 4(1-a)^2$. [1]

(c) Find the value of a that minimises $\mathrm{Var}(W)$, and state the corresponding minimum variance. [2]

29. HARD Calculator **[5 marks]**

The discrete random variable X has probability distribution given by

$$P(X = x) = \frac{x}{c}, \quad x = 1, 2, 3, 4, 5,$$

where c is a constant.

(a) Write down the value of c. [1]

(b) Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$, giving exact values. [3]

(c) Find the value of k such that $\mathrm{Var}(kX - 3) = \mathrm{E}(X)^2$, given $k > 0$. [1]

30. HARD Calculator **[5 marks]**

Two independent random variables X and Y have expectations E$(X) = p$ and E$(Y) = q$, and variances Var$(X) = p(1-p)$ and Var$(Y) = q(1-q)$, where $0 < p < 1$ and $0 < q < 1$.
Let $D = X - Y$.

(a) Write down E(D) and Var(D). [2]

(b) It is known that E$(D) = 0.2$ and Var$(D) = 0.34$. Using the facts that $p + q = 1$ when E$(D) = 0.2$ would require specific values, find p and q given that $p - q = 0.2$ and $p(1-p) + q(1-q) = 0.34$. [2]

(c) State whether p and q you found are consistent with the constraint $0 < p, q < 1$. [1]

Common Mistakes to Avoid

1. **SD vs Variance trap.** $\text{Var}(aX) = a^2\text{Var}(X)$, but $\text{SD}(aX) = |a|\,\text{SD}(X)$. Never add standard deviations directly.
2. **Variance of a difference.** Variances always *add* for independent random variables: $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$ (not subtract).
3. **s_n vs s_{n-1}.** On the GDC, S_x is s_{n-1} (unbiased) and σ_x is s_n (biased). *Always use S_x for inference and unbiased estimation.*
4. **Constants shift the mean, not the variance.** $\text{Var}(X + b) = \text{Var}(X)$ regardless of b.
5. **Sum of n copies.** $\text{Var}(X_1 + \cdots + X_n) = n\,\text{Var}(X)$ (not $n^2\,\text{Var}(X)$, which would be $\text{Var}(nX)$).
6. **Unbiased estimator check.** To verify T is unbiased, compute $\text{E}(T)$ and confirm it equals the parameter. Simply stating "it is unbiased" without showing $\text{E}(T) = \mu$ earns no marks.

Worked Solutions — Section A (Easy)

Q1 — Discrete Probability Distribution EASY

(a) Sum of probabilities $= 1$: **M1**

$0.15 + 0.35 + 0.40 + p = 1 \implies p = 0.10$ **A1**

(b) $\text{E}(X) = 1(0.15) + 2(0.35) + 3(0.40) + 4(0.10)$

$= 0.15 + 0.70 + 1.20 + 0.40 = 2.45$ **A1**

(c) $\text{E}(X^2) = 1(0.15) + 4(0.35) + 9(0.40) + 16(0.10) = 0.15 + 1.40 + 3.60 + 1.60 = 6.75$

$\text{Var}(X) = 6.75 - (2.45)^2 = 6.75 - 6.0025 = 0.7475 \approx 0.748$ **A1**

Note: Award A1 for exact value 0.7475 *or correctly rounded value.*

Q2 — Linear Scaling EASY

(a) $\text{E}(Y) = 3\text{E}(X) - 2 = 3(5) - 2 = 13$ **A1**

(b) $\text{Var}(Y) = 3^2\text{Var}(X) = 9(4) = 36$ **A1**

(c) $\text{SD}(Y) = \sqrt{36} = 6$ **A1**

Q3 — Sum/Difference of Independent RVs EASY

(a) $\text{E}(X + Y) = 8 + 3 = 11$ **A1**

(b) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = 9 + 4 = 13$ **A1**

(c) $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) = 9 + 4 = 13$ **A1**

Note: Variances add for both sum and difference of independent RVs.

Q4 — $\text{E}(X^2)$ and Variance EASY

$\text{E}(X) = 0(0.50) + 2(0.30) + 5(0.20) = 0 + 0.60 + 1.00 = 1.60$ **A1**

$\text{E}(X^2) = 0(0.50) + 4(0.30) + 25(0.20) = 0 + 1.20 + 5.00 = 6.20$ **A1**

$\text{Var}(X) = 6.20 - (1.60)^2 = 6.20 - 2.56 = 3.64$ **A1**

Q5 — Unbiased Variance Estimate EASY

(a) $\bar{x} = \dfrac{12 + 15 + 14 + 18 + 11 + 16}{6} = \dfrac{86}{6} = 14.\overline{3} \approx 14.3$ **A1**

(b) Using GDC 1-Var Stats (S_x value): **(M1)**

$s_{n-1} = 2.5820\ldots, \quad s_{n-1}^2 = (2.5820\ldots)^2 = 6.6\overline{6} \approx 6.67$ **A1**

Note: Award M1 for attempting $\sum(x_i - \bar{x})^2/(n-1)$*; A1 for* 6.67 *(3 s.f.).*

Q6 — Properties of $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ EASY

(a) $\mathrm{E}(2X) = 2 \times 7 = 14$ **A1**

(b) $\mathrm{E}(X^2) = \mathrm{Var}(X) + [\mathrm{E}(X)]^2 = 3 + 49 = 52$ **A1**

(c) $\mathrm{Var}(5 - X) = (-1)^2\mathrm{Var}(X) = \mathrm{Var}(X) = 3$ **A1**

Q7 — Expected Profit EASY

(a) $P(X = 0) = 0.25$ **A1**

(b) $\mathrm{E}(X) = (-1)(0.10) + 0(0.25) + 2(0.45) + 4(0.20)$ **M1**

$= -0.10 + 0 + 0.90 + 0.80 = 1.60$ (hundred dollars per day)

Expected daily profit $= \$160$ **A1**

Q8 — Independent RVs EASY

(a) $\mathrm{E}(X - Y) = 10 - 4 = 6$ **A1**

(b) $\mathrm{Var}(2X) = 4\,\mathrm{Var}(X) = 4(6) = 24$ **A1**

(c) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 6 + 2 = 8$ **A1**

Q9 — Sample Variance EASY

(a) $\bar{x} = \dfrac{62 + 70 + 68 + 75 + 65}{5} = \dfrac{340}{5} = 68$ **A1**

(b) $\sum(x_i - \bar{x})^2 = 36 + 4 + 0 + 49 + 9 = 98$ **(M1)**

$s_{n-1}^2 = \dfrac{98}{4} = 24.5$ **A1**

Q10 — Linear Transformation EASY

$W = -2X + 1$

(a) $\mathrm{E}(W) = -2(4) + 1 = -7$ **A1**

(b) $\mathrm{Var}(W) = (-2)^2\mathrm{Var}(X) = 4(5) = 20$ **A1**

(c) $\mathrm{SD}(W) = \sqrt{20} = 2\sqrt{5} \approx 4.47$ **A1**

Worked Solutions — Section B (Medium)

Q11 — Unknown Parameter in Distribution MEDIUM

(a) Sum of probabilities $= 1$: **M1**

$a + 0.20 + 2a + 0.15 + 0.10 = 1$

$3a + 0.45 = 1 \implies a = \dfrac{0.55}{3} = \dfrac{11}{60} \approx 0.183$ **A1**

(b) With $a = 11/60$: $P(X = 1) = 11/60,\ P(X = 3) = 22/60 = 11/30$

$\mathsf{E}(X) = 1 \cdot \frac{11}{60} + 2(0.20) + 3 \cdot \frac{22}{60} + 4(0.15) + 5(0.10)$ **M1**

$= \frac{11}{60} + 0.40 + \frac{66}{60} + 0.60 + 0.50 = \frac{77}{60} + 1.50 = 1.2833\ldots + 1.50 = 2.783\ldots \approx 2.78$

$\mathsf{E}(X^2) = 1 \cdot \frac{11}{60} + 4(0.20) + 9 \cdot \frac{22}{60} + 16(0.15) + 25(0.10)$

$= \frac{11}{60} + 0.80 + \frac{198}{60} + 2.40 + 2.50 = \frac{209}{60} + 5.70 = 3.4833 + 5.70 = 9.183\ldots$

$\mathsf{Var}(X) = 9.183\ldots - (2.783\ldots)^2 = 9.183\ldots - 7.747\ldots = 1.44$ (3 s.f.) **A1**

Note: Accept $\mathsf{E}(X) = 2.78$ *and* $\mathsf{Var}(X) = 1.44$.

Q12 — Linear Combinations & Sum of Copies MEDIUM

(a) $\mathsf{E}(W) = 0.5(50) + 3 = 25 + 3 = 28$ **A1**

$\mathsf{Var}(W) = (0.5)^2(16) = 0.25(16) = 4$ **A1**

(b) $\mathsf{E}(T) = 2\mathsf{E}(X) = 2(50) = 100$ **A1**

$\mathsf{Var}(T) = 2\,\mathsf{Var}(X) = 2(16) = 32$ **A1**

Note: $\mathsf{Var}(X_1 + X_2) = \mathsf{Var}(X_1) + \mathsf{Var}(X_2)$ *since* X_1, X_2 *independent.*

Q13 — Unbiased Variance from Data MEDIUM

(a) $\bar{x} = \dfrac{155 + 162 + 148 + 170 + 159 + 163 + 151}{7} = \dfrac{1108}{7} = 158.286\ldots \approx 158$ (g) **A1**

(b) Using GDC (S_x): **(M1)**

$s^2_{n-1} = (7.5657\ldots)^2 = 57.2$ (3 s.f.) **A1**

(c) s^2_{n-1} estimates the *population variance*; it is preferred because it is an **unbiased** estimator (its expected value equals the true population variance), whereas s^2_n systematically underestimates it. **R1**

Q14 — Linear Combination, Show That MEDIUM

(a) $\mathsf{E}(T) = 2\mathsf{E}(X) - \mathsf{E}(Y) + 5 = 2(6) - 9 + 5 = 12 - 9 + 5 = 8$ **M1A1**

(Result obtained as required.) **AG**

(b) $\mathsf{Var}(T) = 2^2\mathsf{Var}(X) + (-1)^2\mathsf{Var}(Y) = 4(4) + 1(9)$ **M1**

$= 16 + 9 = 25$ **A1**

Note: The constant $+5$ *does not contribute to variance.*

Q15 — Var(X) then SD$(aX+b)$ MEDIUM

(a) $\text{E}(X) = 0(0.20) + 1(0.30) + 3(0.40) + 6(0.10)$ **M1**

$= 0 + 0.30 + 1.20 + 0.60 = 2.10$

$\text{E}(X^2) = 0(0.20) + 1(0.30) + 9(0.40) + 36(0.10) = 0 + 0.30 + 3.60 + 3.60 = 7.50$

$\text{Var}(X) = 7.50 - (2.10)^2 = 7.50 - 4.41 = 3.09$ **A1A1**

(b) $\text{Var}(3X-4) = 9\,\text{Var}(X) = 9(3.09) = 27.81$

$\text{SD}(3X-4) = \sqrt{27.81} = 5.27\ldots \approx 5.27$ **A1**

Q16 — Probability Distribution from Context MEDIUM

(a) Total "weight" $= 1 + 2 + 3 = 6$. $P(X = k) = k/6$.

x	1	2	3
$P(X = x)$	$1/6$	$2/6$	$3/6$

A1A1

(b) $\text{E}(X) = 1\cdot\frac{1}{6} + 2\cdot\frac{2}{6} + 3\cdot\frac{3}{6} = \frac{1+4+9}{6} = \frac{14}{6} = \frac{7}{3}$ **M1**

Prize paid out $= \text{E}(X) = \frac{7}{3} \approx \$2.33$ per spin.

Organiser collects \$1.50 and pays out $\approx \$2.33$, so expected profit per spin $= 1.50 - \frac{7}{3} = -\frac{5}{6} \approx -\$0.833$ (a loss of \$0.833 per spin). **A1**

Note: Award A1 for $-5/6$ or -0.833.

Q17 — Unbiasedness of Estimators MEDIUM

(a) $\text{E}(T_1) = \dfrac{\text{E}(X_1) + \text{E}(X_2) + \text{E}(X_3)}{3} = \dfrac{\mu + \mu + \mu}{3} = \dfrac{3\mu}{3} = \mu$ **M1**

Since $\text{E}(T_1) = \mu$, T_1 is an unbiased estimator of μ. **AG**

(b) $\text{E}(T_2) = \dfrac{\text{E}(X_1) + 2\text{E}(X_2) + \text{E}(X_3)}{4} = \dfrac{\mu + 2\mu + \mu}{4} = \dfrac{4\mu}{4} = \mu$ **M1**

Since $\text{E}(T_2) = \mu$, T_2 is also an unbiased estimator of μ. **A1**

Note: Award R1 for the correct conclusion with correct justification.

Q18 — Independent Weights MEDIUM

(a) $\text{E}(A+B) = 3.2 + 1.8 = 5.0$ kg **A1**

$\text{Var}(A+B) = 0.49 + 0.16 = 0.65$ kg^2 **M1**

$\text{SD}(A+B) = \sqrt{0.65} = 0.806\ldots \approx 0.806$ kg **A1**

(b) $\text{Var}(A_1 + A_2 + A_3) = 3\,\text{Var}(A) = 3(0.49) = 1.47$ kg^2 **A1**

Note: Only 3 marks available in this question; (a) carries 3 and (b) carries 1.

Q19 — Transformation of Data MEDIUM

(a) Using GDC: $\bar{x} = 21.825$ °C ≈ 21.8 °C **A1**

$s_{n-1} = 0.8681\ldots \approx 0.868$ °C **A1**

(b) For the transformation $K = 0.1X + 273.15$:

$\mathsf{E}(K) = 0.1(21.825) + 273.15 = 2.1825 + 273.15 = 275.333\ldots \approx 275$ K **A1**

$\mathsf{Var}(K) = (0.1)^2 s_{n-1}^2 = 0.01 \times (0.8681\ldots)^2 = 0.01 \times 0.7536\ldots = 7.54 \times 10^{-3}$ **A1**

Note: Accept $\mathsf{Var}(K) = 0.00754$ *(3 s.f.).*

Q20 — Sum of Independent RVs MEDIUM

(a) $\mathsf{E}(X) = 0(0.40) + 1(0.35) + 2(0.20) + 3(0.05)$ **M1**

$= 0 + 0.35 + 0.40 + 0.15 = 0.90$

$\mathsf{E}(X^2) = 0(0.40) + 1(0.35) + 4(0.20) + 9(0.05) = 0 + 0.35 + 0.80 + 0.45 = 1.60$

$\mathsf{Var}(X) = 1.60 - (0.90)^2 = 1.60 - 0.81 = 0.79$ **A1**

(b) $\mathsf{E}(S) = 12\,\mathsf{E}(X) = 12(0.90) = 10.8$ **A1**

$\mathsf{Var}(S) = 12\,\mathsf{Var}(X) = 12(0.79) = 9.48$ **A1**

Worked Solutions — Section C (Hard)

Q21 — Unknown k; Synthesis HARD

(a) $k + 2k + 3k + 4k = 10k = 1 \implies k = 0.1$ **A1**

(b) $\mathrm{E}(X) = 1(0.1) + 2(0.2) + 3(0.3) + 4(0.4) = 0.1 + 0.4 + 0.9 + 1.6 = 3$ **M1A1**
$\mathrm{E}(X^2) = 1(0.1) + 4(0.2) + 9(0.3) + 16(0.4) = 0.1 + 0.8 + 2.7 + 6.4 = 10$
$\mathrm{Var}(X) = 10 - 9 = 1$ **A1**

(c) Given $\mathrm{E}(Y) = 3$ and $\mathrm{Var}(Y) = 2(1) = 2$:
$\mathrm{Var}(X + 2Y) = \mathrm{Var}(X) + 4\,\mathrm{Var}(Y) = 1 + 4(2) = 9$ **M1**
$\mathrm{SD}(X + 2Y) = \sqrt{9} = 3$ **A1**

Q22 — Optimal Unbiased Estimator HARD

(a) $\mathrm{E}(\hat{\mu}) = \frac{1}{6}\mu + \frac{1}{3}\mu + c\mu = \mu\left(\frac{1}{6} + \frac{1}{3} + c\right) = \mu\left(\frac{1}{2} + c\right)$ **M1**
For unbiasedness: $\frac{1}{2} + c = 1 \implies c = \frac{1}{2}$ **A1**

(b) $\mathrm{Var}(\hat{\mu}) = \left(\frac{1}{6}\right)^2\sigma^2 + \left(\frac{1}{3}\right)^2\sigma^2 + \left(\frac{1}{2}\right)^2\sigma^2$ **M1**
$= \left(\frac{1}{36} + \frac{1}{9} + \frac{1}{4}\right)\sigma^2 = \left(\frac{1}{36} + \frac{4}{36} + \frac{9}{36}\right)\sigma^2 = \frac{14}{36}\sigma^2 = \frac{7}{18}\sigma^2$ **A1**

(c) $\mathrm{Var}(\bar{X}) = \dfrac{\sigma^2}{3} = \dfrac{6\sigma^2}{18}$
Since $\frac{6}{18} < \frac{7}{18}$, $\bar{X}$ has smaller variance and is a more efficient estimator. **R1**

Q23 — Trimmed Rod / Sum of 5 HARD

(a) $\mathrm{E}(R) = \mathrm{E}(L) - \mathrm{E}(T) = 30 - 2 = 28$ cm **A1**
$\mathrm{Var}(R) = \mathrm{Var}(L) + \mathrm{Var}(T) = 0.25 + 0.04 = 0.29$ cm^2 **A1**

(b) $\mathrm{E}(S) = 5\,\mathrm{E}(R) = 5(28) = 140$ cm **A1**
$\mathrm{Var}(S) = 5\,\mathrm{Var}(R) = 5(0.29) = 1.45$ cm^2 **A1**

(c) $v = \mathrm{E}(S) = 140$ cm.
Since $\mathrm{E}(S) = 140 < 145.5$, the expected total length is well within the crate limit, so on average the crate is unlikely to be overfilled. **R1**

Q24 — Unknown Parameter; Simultaneous System HARD

(a) $0.10 + m + 0.30 + 2m + 0.10 = 1 \implies 3m = 0.50 \implies m = \frac{1}{6} \approx 0.167$ **A1**

(b) $\mathrm{E}(X) = 0(0.10) + 1\cdot\frac{1}{6} + 2(0.30) + 3\cdot\frac{2}{6} + 4(0.10)$ **M1**
$= 0 + \frac{1}{6} + 0.60 + 1.00 + 0.40 = \frac{1}{6} + 2.00 = 2.1\overline{6}$
$\mathrm{E}(X^2) = 0 + \frac{1}{6} + 4(0.30) + 9\cdot\frac{2}{6} + 16(0.10) = \frac{1}{6} + 1.20 + 3.00 + 1.60 = \frac{1}{6} + 5.80 = 5.9\overline{6}$
$\mathrm{Var}(X) = 5.9\overline{6} - (2.1\overline{6})^2 = 5.9\overline{6} - 4.694\ldots = 1.27$ (3 s.f.) **A1**

(c) $\mathsf{E}(aX - bY) = a(2.1\overline{6}) - 3b = 0 \implies b = \dfrac{13a}{18}$ **M1**

$\mathsf{Var}(aX - bY) = a^2(1.27) + b^2(2) = 26$

Substituting $b = \frac{13a}{18}$:

$1.27a^2 + 2\cdot\left(\frac{13a}{18}\right)^2 = 26 \implies 1.27a^2 + \frac{338a^2}{324} = 26 \implies (1.27 + 1.0432\ldots)a^2 = 26$

$2.313\ldots\ a^2 = 26 \implies a^2 = 11.24\ldots \implies a = 3.35\ (3\text{ s.f.})$

$b = \dfrac{13(3.35)}{18} = 2.42\ (3\text{ s.f.})$ **A1**

Note: Accept $a = 3.35$, $b = 2.42$.

Q25 — Combined Sample Mean HARD

(a) $\bar{x}_2 = \dfrac{22.0 + 24.0 + 26.0 + 28.0}{4} = \dfrac{100.0}{4} = 25.0$ **A1**

$\sum(x_i - \bar{x}_2)^2 = 9 + 1 + 1 + 9 = 20, \quad s^2_{n-1,2} = \dfrac{20}{3} = 6.\overline{6} \approx 6.67$ **A1**

(b) Combined mean formula: **M1**

$\bar{x}_{\text{combined}} = \dfrac{n\bar{x}_1 + 4\bar{x}_2}{n+4}$

$25.6 = \dfrac{26.0n + 4(25.0)}{n+4} = \dfrac{26.0n + 100}{n+4}$

$25.6(n+4) = 26.0n + 100$

$25.6n + 102.4 = 26.0n + 100$

$2.4 = 0.4n \implies n = 6$ **A1**

(c) The sample size $n = 6$ is small, so $s^2_{n-1,1}$ has high variability and may not be a reliable (precise) estimate of the population variance. **R1**

Q26 — Variance of Difference; Finding Constant HARD

(a) $\mathsf{E}(X - Y) = \mathsf{E}(X) - \mathsf{E}(Y) = \mu - \mu = 0$ **M1**

$\mathsf{Var}(X - Y) = \mathsf{Var}(X) + \mathsf{Var}(Y) = \sigma^2 + \sigma^2 = 2\sigma^2$ **A1**

(b) $\mathsf{Var}(D) = \mathsf{Var}(aX - aY) = a^2\mathsf{Var}(X) + a^2\mathsf{Var}(Y) = 2a^2\sigma^2$ **M1**

Set equal to $16\sigma^2$: $2a^2\sigma^2 = 16\sigma^2 \implies a^2 = 8 \implies a = 2\sqrt{2}$ **A1**

(c) $\mathsf{E}(D) = a\mathsf{E}(X) - a\mathsf{E}(Y) = a(\mu - \mu) = 0$

$\mathsf{E}(D^2) = \mathsf{Var}(D) + [\mathsf{E}(D)]^2 = 16\sigma^2 + 0 = 16\sigma^2$ **A1**

Q27 — Standardisation; Properties of Z-scores HARD

(a) Data: $55, 62, 48, 71, 58, 66, 53, 60, 73, 54$

$\bar{x} = \dfrac{600}{10} = 60$ **A1**

GDC (S_x): $s_{n-1} = 8.0829\ldots$, $s_{n-1}^2 = 65.3$ (3 s.f.) **A1**

(b) Mean of Z-scores: $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} \frac{x_i - \bar{x}}{s_{n-1}} = \frac{1}{s_{n-1}} \cdot \frac{\sum(x_i - \bar{x})}{n} = \frac{1}{s_{n-1}} \cdot 0 = 0$ **M1**

(Since $\sum(x_i - \bar{x}) = 0$ by definition of the mean.) **AG**

(c) $Z_i = \frac{X_i - \bar{x}}{s_{n-1}} = \frac{1}{s_{n-1}} X_i - \frac{\bar{x}}{s_{n-1}}$, a linear transformation with scale factor $\frac{1}{s_{n-1}}$. **M1**

$s_{Z,n-1}^2 = \left(\frac{1}{s_{n-1}}\right)^2 s_{n-1}^2 = 1$ **A1**

The unbiased variance estimate of the Z-scores is 1.

Q28 — Optimal Unbiased Estimator via Minimisation HARD

(a) $\mathrm{E}(W) = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y) = a\mu + b\mu = (a+b)\mu = 1 \cdot \mu = \mu$ **M1**

Since $\mathrm{E}(W) = \mu$, W is an unbiased estimator for any $a + b = 1$. **AG**

(b) $b = 1 - a$, so **M1**

$\mathrm{Var}(W) = a^2\,\mathrm{Var}(X) + (1-a)^2\,\mathrm{Var}(Y) = 9a^2 + 4(1-a)^2$ **AG**

(c) $\frac{d}{da}[9a^2 + 4(1-a)^2] = 18a - 8(1-a) = 18a - 8 + 8a = 26a - 8$ **M1**

Set to zero: $26a = 8 \implies a = \frac{4}{13}$, so $b = \frac{9}{13}$. **A1**

Minimum variance $= 9\left(\frac{4}{13}\right)^2 + 4\left(\frac{9}{13}\right)^2 = \frac{144 + 324}{169} = \frac{468}{169} = \frac{36}{13} \approx 2.77$ **A1**

Q29 — Parametric Distribution; External Condition HARD

(a) $\sum_{x=1}^{5} \frac{x}{c} = \frac{1+2+3+4+5}{c} = \frac{15}{c} = 1 \implies c = 15$ **A1**

(b) $\mathrm{E}(X) = \sum_{x=1}^{5} x \cdot \frac{x}{15} = \frac{1}{15}\sum_{x=1}^{5} x^2 = \frac{1^2+2^2+3^2+4^2+5^2}{15} = \frac{55}{15} = \frac{11}{3}$ **M1A1**

$\mathrm{E}(X^2) = \sum_{x=1}^{5} x^2 \cdot \frac{x}{15} = \frac{1}{15}(1 + 8 + 27 + 64 + 125) = \frac{225}{15} = 15$

$\mathrm{Var}(X) = 15 - \left(\frac{11}{3}\right)^2 = 15 - \frac{121}{9} = \frac{135 - 121}{9} = \frac{14}{9}$ **A1**

(c) $\mathrm{Var}(kX - 3) = k^2\mathrm{Var}(X) = k^2 \cdot \frac{14}{9}$

Set equal to $[\mathrm{E}(X)]^2 = \left(\frac{11}{3}\right)^2 = \frac{121}{9}$: **M1**

$k^2 \cdot \frac{14}{9} = \frac{121}{9} \implies k^2 = \frac{121}{14} \implies k = \frac{11}{\sqrt{14}} = \frac{11\sqrt{14}}{14} \approx 2.94$ **A1**

Q30 — Bernoulli-type Parameters; System **HARD**

(a) $\mathsf{E}(D) = \mathsf{E}(X) - \mathsf{E}(Y) = p - q$ **A1**

$\mathsf{Var}(D) = \mathsf{Var}(X) + \mathsf{Var}(Y) = p(1-p) + q(1-q)$ **A1**

(b) From $p - q = 0.2$: $p = q + 0.2$ **M1**

Substituting into $p(1-p) + q(1-q) = 0.34$:

$(q + 0.2)(1 - q - 0.2) + q(1-q) = 0.34$

$(q + 0.2)(0.8 - q) + q - q^2 = 0.34$

$0.8q - q^2 + 0.16 - 0.2q + q - q^2 = 0.34$

$-2q^2 + 1.6q + 0.16 = 0.34$

$2q^2 - 1.6q + 0.18 = 0$

$q^2 - 0.8q + 0.09 = 0$

$$q = \frac{0.8 \pm \sqrt{0.64 - 0.36}}{2} = \frac{0.8 \pm \sqrt{0.28}}{2} = \frac{0.8 \pm 0.5292\ldots}{2}$$

$q = 0.6646\ldots$ or $q = 0.1354\ldots$

Correspondingly $p = 0.8646\ldots$ or $p = 0.3354\ldots$ **A1**

(c) Both solutions satisfy $0 < p, q < 1$, so both are consistent with the constraint. **R1**

Note: Accept both pairs of solutions or state there are two valid solutions.