IB DP MATHS AI · TOPIC 4 STATISTICS AND PROBABILITY · AHL 4.14–4.19 · ~8 MIN READ

Estimation, Poisson, hypothesis tests and Markov chains

HL statistics for Maths AI: linear combinations of random variables, unbiased estimates, the central limit theorem, confidence intervals for a mean, the Poisson distribution, hypothesis tests with binomial, Poisson and normal distributions, Type I and Type II errors, and Markov chains with transition matrices and steady states.

E(aX + b) & Var(aX + b) | Linear combinations | Unbiased estimators | Central limit theorem | Confidence intervals (z and t) | Poisson distribution | Binomial hypothesis tests | Critical regions

RS by Riyaz Sir
mentorriyaz.in

No tag = **SL and HL** HL ONLY HL students only FORMULA BOOKLET given in the exam
M1 A1 how marks are awarded highlight = comes up again and again Test yourself at the end

IN 30 SECONDS

- **Combining variables:** E(aX + bY) = aE(X) + bE(Y); for independent X and Y, Var(aX + bY) = a^2Var(X) + b^2Var(Y). Variances **add** even when you subtract.
- **Central limit theorem:** for large n (n ≥ 30 is the usual guide), the sample mean $\bar{X}$ is approximately N(μ, $\frac{\sigma^2}{n}$), whatever the shape of the population.
- **Confidence interval** for μ: $\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$ when σ is known, or $\bar{x} \pm t\frac{s_{n-1}}{\sqrt{n}}$ when it isn't.
- **Poisson Po(m):** counts of random events in a fixed interval, with mean = variance = m.
- **Errors:** a Type I error rejects a true H_0 (its probability is the significance level); a Type II error fails to reject a false H_0. A **Markov chain** moves between states with a transition matrix T; the steady state s satisfies Ts = s.

BY THE END YOU CAN

- Use E(aX + b), Var(aX + b) and linear combinations of independent random variables. HL
- Find unbiased estimates of the mean and variance. HL
- Use the central limit theorem and find confidence intervals for a mean. HL
- Use the Poisson distribution, including sums of independent Poisson variables. HL
- Carry out hypothesis tests with binomial, Poisson and normal distributions, find critical regions, and Type I and II error probabilities. HL
- Use transition matrices for Markov chains and find steady states. HL

SYLLABUS NOTE

From first exams in 2029, the Poisson distribution and the hypothesis tests for a normal mean, a Poisson mean and the correlation coefficient leave AI HL. If you sit exams up to November 2028, you still need all of this note.

1 Combining random variables HL

$$E(aX + b) = aE(X) + b,\ Var(aX + b) = a^2Var(X)$$ FORMULA BOOKLET

adding b doesn't change the spread

$$E(a_1X_1 \pm a_2X_2) = a_1E(X_1) \pm a_2E(X_2),\ Var(a_1X_1 \pm a_2X_2) = a_1^2Var(X_1) + a_2^2Var(X_2)$$ FORMULA BOOKLET

X_1, X_2 independent

WORKED EXAMPLE

$X \sim N(50, 3^2)$ and $Y \sim N(30, 4^2)$ are independent. Find the distribution of $X - Y$ and $P(X - Y > 25)$.

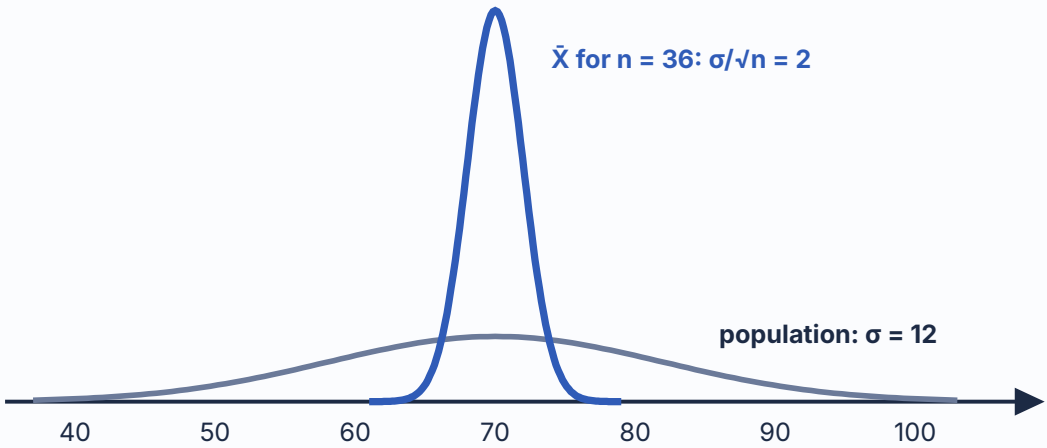
1. $E(X - Y) = 20$, $Var(X - Y) = 9 + 16 = 25$ M1 A1
2. $X - Y \sim N(20, 5^2)$ A1
3. **$P(X - Y > 25) = 0.159$** A1

COMMON MISTAKE

Writing $Var(X - Y) = 9 - 16$. Variances never subtract: two sources of variation always add uncertainty.

- **Unbiased estimates:** $\bar{x}$ estimates μ; $s_{n-1}^2 = \frac{n}{n-1}s_n^2$ estimates σ^2. For 10, 12, 9, 11, 13: $\bar{x} = 11$ and $s_{n-1}^2 = \frac{10}{4} = 2.5$.

2 The central limit theorem and confidence intervals HL



Sample means cluster much more tightly around μ, and are approximately normal for large n.

The distribution of the sample mean is centred at μ and much narrower than the population.

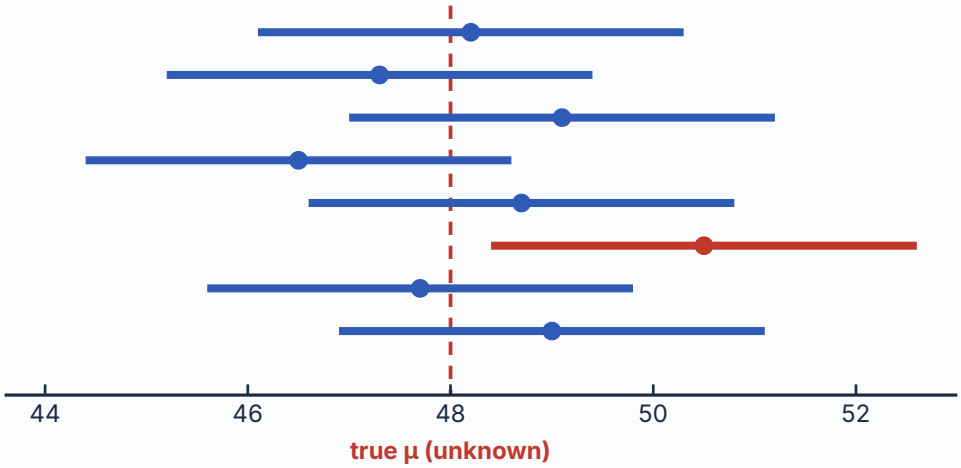
WORKED EXAMPLE

A population has mean 70 and SD 12. A sample of 36 is taken. Find $P(\bar{X} > 73)$.

1. $\bar{X} \sim N(70, \frac{12^2}{36}) = N(70, 2^2)$ (CLT) M1 A1
2. **$P(\bar{X} > 73) = 0.0668$** A1

$$\bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}, \ \bar{x} \pm t \times \frac{s_{n-1}}{\sqrt{n}}$$ FORMULA BOOKLET

confidence intervals for μ



8 samples, 8 intervals: 7 contain μ (blue) and 1 misses it (red).

A 95% interval method captures the true mean in about 95% of samples. Any single interval either contains μ or it doesn't.

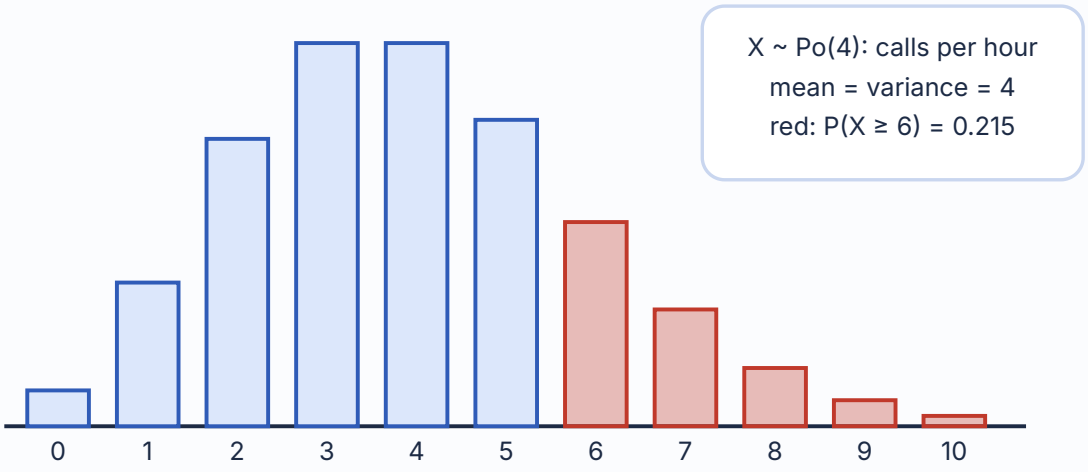
WORKED EXAMPLE

A sample of 25 has $\bar{x} = 48.2$ and $s_{n-1} = 5$. Find a 95% confidence interval for μ.

1. σ unknown, so use t with 24 df: $t = 2.064$ M1
2. $48.2 \pm 2.064 \times \frac{5}{5}$ M1
3. **$46.1 < \mu < 50.3$** A1

- **Interpreting:** "we are 95% confident μ is between 46.1 and 50.3". A wider interval comes from a higher confidence level or a smaller sample.

3 The Poisson distribution HL



Po(4): the number of calls in an hour. The mean is 4 and so is the variance.

$$P(X = x) = \frac{m^x e^{-m}}{x!}, \quad E(X) = \mathrm{Var}(X) = m$$ FORMULA BOOKLET

Poisson with mean m

- **Conditions:** events happen singly, independently, at a constant average rate. If X ~ Po(m) and Y ~ Po(n) are independent, X + Y ~ Po(m + n).

WORKED EXAMPLE

A helpline gets 4 calls an hour on average. Find P(exactly 2 calls in an hour), P(at least 6 calls in an hour), and P(no calls in 30 minutes).

1. X ~ Po(4): P(X = 2) = **0.147** A1
2. P(X ≥ 6) = 1 − P(X ≤ 5) = **0.215** A1
3. **in 30 minutes Y ~ Po(2): P(Y = 0) = e^{-2} = 0.135** M1 A1

EXAM TIP

A quick check for Poisson data: the sample mean and variance should be roughly equal.

Riyaz Sir

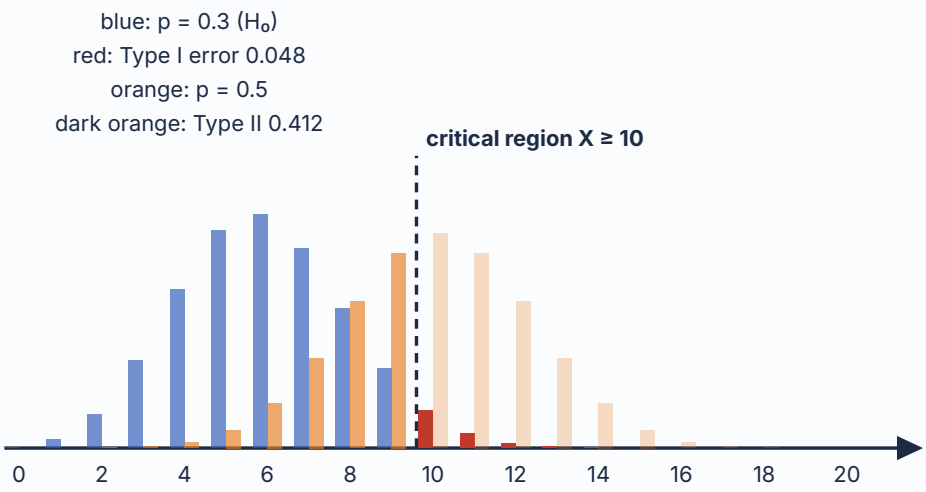
4 Hypothesis tests HL

WORKED EXAMPLE

A seed company says 30% of its seeds fail. A gardener suspects more fail. In a sample of 20, 10 fail. Test at the 5% level.

1. $X \sim B(20, p)$; $H_0: p = 0.3$, $H_1: p > 0.3$ A1
2. $P(X \geq 10 \mid p = 0.3) = 1 - P(X \leq 9) =$ **0.0480** M1 A1
3. **0.0480 < 0.05, so reject H_0: there is evidence that more than 30% fail** R1

- **Critical region:** the set of results that lead to rejecting H_0. Here $P(X \geq 10) = 0.048 \leq 0.05$ but $P(X \geq 9) = 0.113 > 0.05$, so the critical region is **X ≥ 10**.



Type I: reject a true H_0. Type II: fail to reject a false H_0.

The Type I error is the red tail under H_0; the Type II error is the orange part of the true distribution outside the critical region.

WORKED EXAMPLE

For this test, find the probability of a Type I error, and of a Type II error if really p = 0.5.

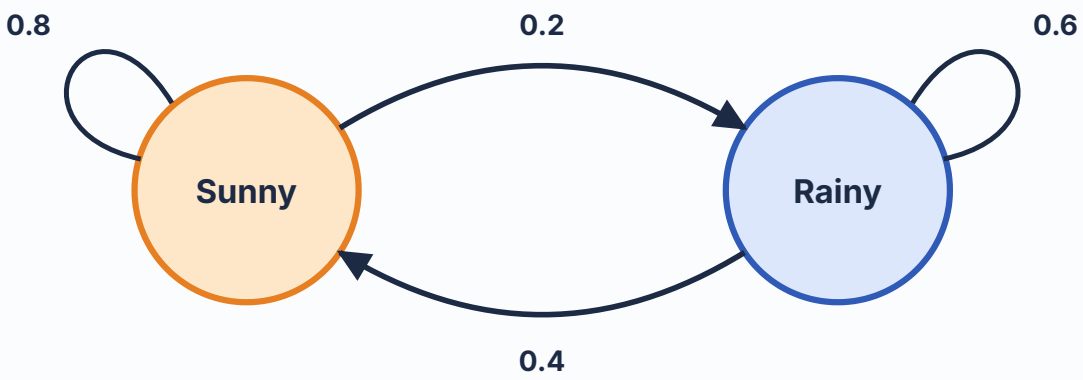
1. Type I: $P(X \geq 10 \mid p = 0.3) =$ **0.0480** A1
2. **Type II: $P(X \leq 9 \mid p = 0.5) = 0.412$** M1 A1

- **Other tests:** a **normal mean** test uses $\bar{X} \sim N(\mu_0, \sigma^2/n)$ (z-test) or a t-test when σ is unknown; a **Poisson mean** test works like the binomial one; a **correlation** test checks $H_0: \rho = 0$ using r and the GDC's linear regression t-test. All three are removed for 2029.

COMMON MISTAKE

Using a discrete distribution as if it hits exactly 5%. With binomial and Poisson tests, the actual significance level is the probability of the critical region (here 4.80%), usually below 5%.

5 Markov chains HL



T = (0.8 0.4; 0.2 0.6): column = from, row = to. Steady state: ⅔ sunny, ⅓ rainy.

A two-state Markov chain. Each column of T adds to 1.

$$s_n = T^n s_0$$

state after n steps; the steady state s satisfies Ts = s

WORKED EXAMPLE

Sunny days are followed by sunny days with probability 0.8, rainy days by sunny days with probability 0.4. Today is sunny. Find the probability that it is sunny in 2 days' time, and the long-term proportion of sunny days.

1. T = (0.8 0.4; 0.2 0.6), s_0 = (1; 0) A1
2. $T^2 s_0$ = (0.72; 0.28), so **P(sunny) = 0.72** M1 A1
3. **steady state: $0.8s + 0.4(1 - s) = s \Rightarrow s = \frac{2}{3}$, so ⅔ of days are sunny in the long run** M1 A1

- **Eigenvector link:** the steady state is the eigenvector of T with eigenvalue 1, scaled so its entries add to 1.

LINKING QUESTION

Linear combinations generalise the binomial and normal ideas of Probability, binomial and normal distributions; the SL χ^2 and t-tests (Chi-squared and t-tests) use the same logic as these HL tests; Markov chains use eigenvectors from Matrices and eigenvalues and random walks from Graph theory.

IDEA FOR YOUR IA / EXPLORATION

Estimate a real population mean (the battery life of phones in your class, the waiting time at a canteen) with a confidence interval, or model customer loyalty between brands as a Markov chain and find its steady state.

Test yourself — cover the answers first

QUICK CHECK

1. HL Var(X) = 4. Find Var(3X − 2).
2. HL X and Y are independent with variances 5 and 7. Find Var(X − Y).
3. HL What does the CLT say about $\bar{X}$ for large n?
4. HL X ~ Po(3). Find Var(X).
5. HL What is a Type I error?

6. HL What must each column of a transition matrix add up to?

EXAM-STYLE

7. HL Cars pass a point at an average of 2 per minute and lorries at 0.5 per minute, independently. Find the probability that exactly 3 vehicles pass in one minute.
8. HL Find a 90% confidence interval for μ from a sample of 50 with $\bar{x} = 120$ when $\sigma = 15$ is known.

Show answers

1. 36
2. 12
3. approximately $N(\mu, \sigma^2/n)$
4. 3
5. rejecting H_0 when it is true
6. 1
7. total $V \sim Po(2.5)$

   $P(V = 3) = \frac{2.5^3 e^{-2.5}}{3!}$

   = 0.214
8. $z = 1.645$

   $120 \pm 1.645 \times \frac{15}{\sqrt{50}} = 120 \pm 3.49$

   $116.5 < \mu < 123.5$

Stuck on a question?

Send it to me on WhatsApp. I read every message.

Share with a friend | Ask Riyaz Sir

Want regular help? Send this to your parent →