PHOTON ACADEMY · IB MATH AI HL CHEATSHEET

# Estimation & Confidence Intervals

AHL 4.14–4.15

ibtuition.sg · free

## FORMULAS YOU MUST KNOW

$\hat{\mu} = \bar{x} = \dfrac{\sum x}{n}$ — the sample mean estimates the population mean $\mu$.

Estimate of the mean:

$\widehat{\sigma^2} = s_{n-1}^2 = \dfrac{\sum (x - \bar{x})^2}{n - 1}$ — divide by $n - 1$, not $n$.

Estimate of the variance:

$E(\bar{x}) = \mu$ and $E(s_{n-1}^2) = \sigma^2$ — right on average, not too big or too small.

Unbiased means:

$\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$ — exact for a normal population, approximate for any population when $n$ is large.

Sampling distribution:

$\mathrm{SE} = \dfrac{\sigma}{\sqrt{n}}$ (known $\sigma$) or $\dfrac{s}{\sqrt{n}}$ (unknown $\sigma$) — the standard deviation of $\bar{x}$.

Standard error:

$\bar{x} \pm E$ — always centred on the sample mean.

The interval:

$\bar{x} \pm z^* \dfrac{\sigma}{\sqrt{n}}$.

$\sigma$ known ($z$-interval):

$\bar{x} \pm t^* \dfrac{s}{\sqrt{n}}$, with $t^*$ read at $\nu = n - 1$ degrees of freedom.

$\sigma$ unknown ($t$-interval):

Enter the raw list → choose TInterval (or ZInterval) → set the C-level, e.g. $0.95$.

Data mode:

Type in $\bar{x}$, $s_x$ (or $\sigma$) and $n$ → same routine, same C-level.

Stats mode:

The interval $(a,\ b)$, ready to quote, together with $\bar{x}$ and $n$.

Output:

$E = t^* \dfrac{s}{\sqrt{n}}$ (or $z^* \dfrac{\sigma}{\sqrt{n}}$) — critical value × standard error.

Margin of error:

## KEY METHODS

- This is the exact opposite of describing a fixed data set (Univariate Statistics), where IB wants $\sigma_x$. Rule of thumb: describing data → $\sigma_x$; estimating a population from a sample → $s_x$.
- Divide by $\sqrt{n}$, never by $n$. Quadruple the sample and the standard error only halves — precision improves slowly, and that is exactly why big samples are expensive.
- Central Limit Theorem: for $n \geq 30$ the sample mean is approximately normal whatever the population's

## EXAM TRAPS TO AVOID

- Dividing by $n$ (the calculator's $\sigma_x$) systematically underestimates the population variance. For estimation IB uses the $n - 1$ divisor — the $s_x$ key — so the estimator is unbiased.
- Pick the interval by what you are given: a stated population $\sigma$ → $z$; only the sample $s$ → $t$. In AI HL exams $\sigma$ is almost always unknown, so the $t$-interval is the default choice.
- $E$ multiplies the standard error $s/\sqrt{n}$ — not $s$ on its own. Dropping the $\sqrt{n}$ inflates every limit by a factor of $\sqrt{n}$.

shape — the reason the interval formulas work so widely in real AI applications.

- $t^* > z^*$ for the same confidence (the $t$-distribution has heavier tails), and $t^* \to z^*$ as $n$ grows — for large samples the two intervals nearly coincide.
- For a $t$-interval, enter the sample standard deviation $s_x$ (the $n-1$ value) in the s.d. field. For a $z$-interval enter the given population $\sigma$. Wrong field, wrong interval.
- If a question wants the margin of error but you built the interval on the GDC, recover it instantly as $E = \dfrac{b-a}{2}$.
- $t^*$ depends on both the confidence level and $\nu = n-1$; read it from the GDC's inverse-$t$ or straight from the interval — never guess it.
- Width is the full span, so keep the factor of $2$: $w = 2E$. The sample mean sits at the midpoint, never at an end.
- Raising the confidence level widens the interval but does not move its centre $\bar{x}$ — more confidence buys less precision, not a different estimate.
- Always round $n$ up to the next whole number — rounding down leaves the interval wider than the target. And $n$ must be a positive integer.
- It is not "there is a $95\%$ probability that $\mu$ lies in this interval". $\mu$ is a fixed number; the interval is what is random. Once computed, this particular interval either contains $\mu$ or it does not.