

Sampling and estimation

A-Level Maths 9709 · Topic 6.4 revision notes

Everything from the 2 lessons of this topic on a few pages: the key idea of each lesson, the diagram to remember, the points to learn and the traps examiners set. Read it, cover it, test yourself.

1. Populations, samples and the distribution of the sample mean

2. Unbiased estimates and confidence intervals

thelucidstem.com

1 Populations, samples and the distribution of the sample mean

Syllabus 6.4

A random sample gives every member of the population an equal chance of selection, so it avoids bias. The sample mean $\bar{X}$ of n independent observations has $E(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$. If X is normal, $\bar{X}$ is exactly normal; if n is large, the Central Limit Theorem says $\bar{X}$ is approximately normal anyway.

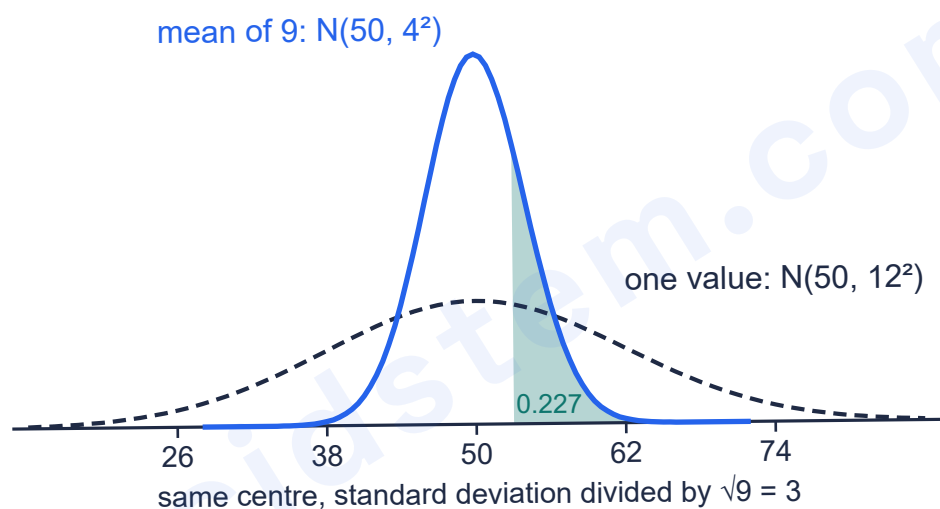


Fig. 1: one value from $N(50, 12^2)$ (dashed) and the mean of a random sample of 9 (blue), drawn to scale. Both are centred on 50, but the sample mean has standard deviation $12/3 = 4$; the shaded area is $P(X > 53) = 0.2266$

Worked example: Choosing with random numbers

A school has 600 students. Describe how to use random numbers to choose a random sample of 30 students.

1. Obtain a list of all 600 students and number them 001 to 600.
2. Generate random three-digit numbers (calculator or table).
3. Ignore numbers 000 and above 600, and ignore repeats.
4. Continue until 30 different numbers are obtained; the students with those numbers form the sample.

Answer: Number the list, select by random numbers, discard out-of-range values and repeats.

Centre μ , spread $\sigma / \sqrt{n}$,
shape normal.

- ✓ **Random sampling:** number the population, select with random numbers, ignore repeats and out-of-range values.
- ✓ **Criticise in context:** who is left out, and which way would that bias the result?
- ✓ **Sample mean:** $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, sd $\frac{\sigma}{\sqrt{n}}$.
- ✓ **Shape:** exact if X is normal; by the CLT if n is large; say which.
- ✓ **Finding n :** set up an inequality in $\sqrt{n}$ and round UP.

Watch out!

"It is too small" or "it is not random" alone rarely scores. Say WHO is excluded and HOW that would push the answer up or down.

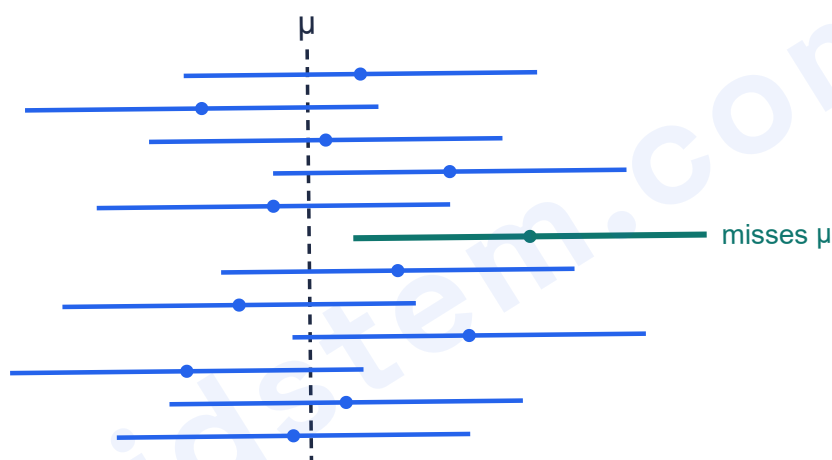
$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$, so the standard deviation is $\frac{\sigma}{\sqrt{n}}$, not $\frac{\sigma}{n}$. With $\sigma = 12$ and $n = 9$ that is 4, not 1.33.

Use the CLT only when the population is NOT normal and n is large, and say so. If the population is normal, $\bar{X}$ is exactly normal and the CLT is not needed; if the population is not normal and n is small, neither applies.

2 Unbiased estimates and confidence intervals

Syllabus 6.4

Unbiased estimates: $\hat{\mu} = \bar{x}$ and $s^2 = \frac{1}{n-1} \left(\sum x^2 - \frac{(\sum x)^2}{n} \right)$. A confidence interval for μ is $\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$ (use s for σ when n is large); for a proportion it is $\hat{p} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. For 95%, $z = 1.96$.



each line: $\bar{x} \pm 1.96\sigma/\sqrt{n}$ from a different random sample

Fig. 1: illustration of twelve 95% confidence intervals from different random samples. Each interval has the same width; their centres vary. In the long run 95% of such intervals contain μ and 5% (about 1 in 20) miss it

Worked example: From raw data

The masses, in grams, of 5 randomly chosen eggs are 12.4, 13.1, 11.8, 12.9, 13.5. Find unbiased estimates of the population mean and variance.

1. $\sum x = 63.7$ and $\sum x^2 = 813.27$.
2. $\bar{x} = \frac{63.7}{5} = 12.74$.
3. $s^2 = \frac{1}{4} \left(813.27 - \frac{63.7^2}{5} \right) = \frac{1}{4} (813.27 - 811.538) = \frac{1.732}{4}$.

Answer: $\bar{x} = 12.74$ g, $s^2 = 0.433$ g²

Estimate, standard error, Z, interval, interpret.

- ✓ **Estimates:** $\bar{x} = \frac{\sum x}{n}$, $s^2 = \frac{1}{n-1} \left(\sum x^2 - \frac{(\sum x)^2}{n} \right)$.
- ✓ **Coded data:** work with $x - a$, then add a back to the mean; the variance is unchanged.
- ✓ **Mean:** $\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$; use s when n is large and σ is unknown.
- ✓ **Proportion:** $\hat{p} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, large n only.
- ✓ **Interpret:** a stated value inside the interval is plausible; outside, it is doubtful.

Watch out!

Read carefully whether the question wants the unbiased estimate (divide by $n - 1$) and do not take the square root unless the standard deviation is asked for. Dividing by n is the usual lost mark.

Interpret correctly: a 95% interval does NOT mean "there is a 95% probability that μ lies in this interval" (μ is fixed). It means that 95% of intervals constructed this way would contain μ (Fig. 1).

Use $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, not $\frac{\hat{p}(1-\hat{p})}{n}$ and not $\sqrt{np(1-p)}$ (that is the standard deviation of the COUNT, not of the proportion).