

LESSON 10 9709 TO 9231

Sampling and estimation

Connect individual measurements, random sample means and estimates of unknown population values.

By the end of this lesson

- Distinguish a population parameter from a sample statistic.
- Find the mean, variance and standard error of a sample mean.
- Choose an exact normal model or a justified CLT approximation.
- Estimate a population mean and variance from raw or summarised data.

Before you begin: Calculate means and descriptive variances; Use expectation and variance for independent sums; Standardise a normal random variable and find normal probabilities.

1. What is random, and what is fixed?

A population is the complete group or probability model we want to describe. A sample contains the observations available to us. A parameter describes the population: its mean μ and variance σ^2 are fixed quantities, even when we do not know their values.

Notation	Meaning
X or X_i	One random observation before its value is known.
$\bar{X}$	The random mean of a sample of size n .
x_i and $\bar{x}$	The observed values and their calculated mean.
μ, σ^2	Population mean and population variance.
s^2	An estimate of population variance from the observed sample.

A statistic is calculated from a sample. Before sampling it is random because another sample could give a different result. After observing the data, $\bar{x}$ is a particular number; it is used to estimate μ . Keep this distinction when writing probabilities: $P(\bar{X} > 54)$ describes possible future sample means.

2. Sampling assumptions matter

The formulas here assume independent observations from the same population, with finite mean and variance. Random selection supports a defensible link between the sample and population. A conve-

nience sample can systematically miss part of a population; taking more observations does not automatically remove that bias.

Independence means that knowing one observation does not change the distribution of another. Sampling with replacement from a finite population provides independent draws. Sampling without replacement is not exactly independent; the simple variance formula below then needs justification or an appropriate correction. In these notes, independence is given or assumed explicitly.

METHOD TIP A random sample can still look unusual

Random sampling does not make every sample perfectly representative. Sampling variability is precisely why we study the distribution of $\bar{X}$.

3. The distribution of sample means

KEY FORMULA Sample mean and its moments

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}, \quad E(\bar{X}) = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

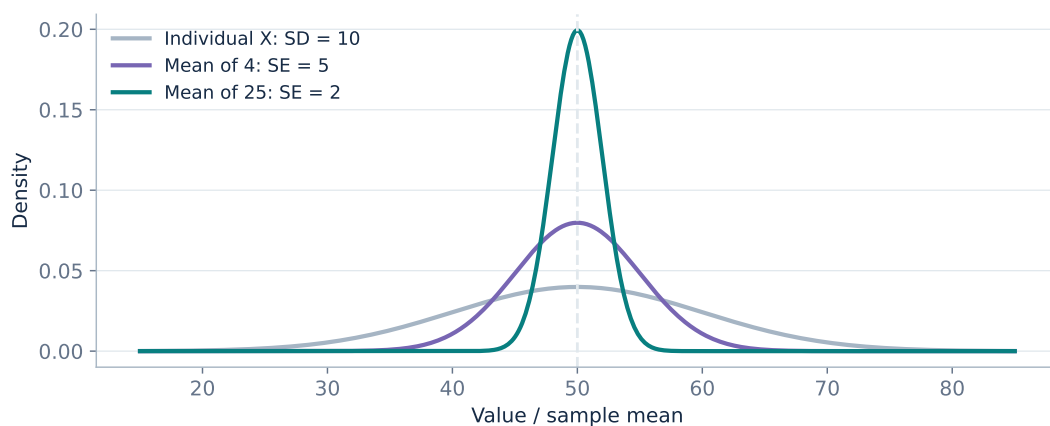
These moment formulas hold for independent observations from the same population with finite variance; normality is not required.

The expectation follows by adding n means and dividing by n . For variance, independence allows us to add the n variances, while dividing the sum by n divides its variance by n^2 . Thus $n\sigma^2/n^2 = \sigma^2/n$. Sample means have the same centre as individual observations but less variation.

KEY FORMULA Standard error of a mean

$$\text{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Standard error means the standard deviation of a statistic. It measures the spread of sample means, rather than the spread of individual observations.



For a normal population with mean 50 and standard deviation 10, sample means have standard deviation 5 when $n = 4$, and 2 when $n = 25$. All three curves are centred at 50.

4. Exact normal sampling distributions

When the population is normal, any independent sample mean is exactly normal. There is no requirement that the sample size exceed a particular threshold. In the notation $N(\mu, \sigma^2)$, the second parameter is the variance, not the standard deviation.

KEY FORMULA Normal population

$$X_i \sim N(\mu, \sigma^2) \implies \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

The observations must be independent and have this same normal distribution.

WORKED EXAMPLE A sample mean above a threshold

Individual measurements follow $N(50, 100)$. For a random independent sample of 25, find the distribution of $\bar{X}$ and $P(\bar{X} > 54)$.

1. Divide the population variance by the sample size.

$$\bar{X} \sim N(50, 100/25) = N(50, 4), \quad \text{SE} = 2$$

2. Standardise using the standard error.

$$P(\bar{X} > 54) = P\left(Z > \frac{54 - 50}{2}\right) = P(Z > 2) \approx 0.0228$$

Answer. $\bar{X} \sim N(50, 4)$ and the probability is approximately 0.0228. Using the individual standard deviation 10 would answer a different question.

5. A normal approximation for other populations

The central limit theorem explains why a sample mean can be approximately normal even when individual observations are not. For independent observations from the same population with finite variance, the approximation improves as the sample size increases. Strong skewness or heavy tails may require much larger samples than a nearly symmetric population.

KEY FORMULA Suitable large-sample approximation

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Use this only when a normal approximation is suitable. There is no universal rule that $n = 30$ guarantees accuracy for every population.

WORKED EXAMPLE A skewed waiting-time population

Waiting times have mean 24 minutes and standard deviation 10 minutes. A normal approximation is suitable for the mean of 100 independent random observations. Estimate $P(\bar{X} < 22)$.

1. The variance formula still applies although the population is skewed.

$$\text{Var}(\bar{X}) = \frac{100}{100} = 1, \quad \text{SE} = 1$$

2. Use the stated normal approximation.

$$P(\bar{X} < 22) \approx P\left(Z < \frac{22 - 24}{1}\right) = P(Z < -2) \approx 0.0228$$

Answer. The estimated probability is 0.0228. The sampling model is approximate, so retain an approximation sign.

6. Estimating unknown population values

An estimator is unbiased when its expected value equals the parameter it estimates. Since $E(\bar{X}) = \mu$, the sample mean is an unbiased estimator of the population mean. This is a statement about repeated sampling; an individual estimate may still be too high or too low.

KEY FORMULA Unbiased estimate of population variance

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

Use $n > 1$ and an independent sample from a population with finite variance. The variance describing the listed data themselves instead uses denominator n .

The deviations from the sample mean sum to zero. Once $n - 1$ deviations are known, the last is determined. The denominator $n - 1$ also corrects the downward bias caused by measuring deviations from the estimated mean. The square root s estimates the population standard deviation; unbiasedness of s^2 does not mean s is itself unbiased.

WORKED EXAMPLE Estimating from raw data

An independent random sample gives 4, 6, 8, 10. Estimate the population mean and variance, then calculate the estimated standard error of the mean.

1. Find the observed mean and squared deviations.

$$\bar{x} = 7, \quad \sum (x_i - \bar{x})^2 = 9 + 1 + 1 + 9 = 20$$

2. Use the sample-variance denominator.

$$s^2 = \frac{20}{3}, \quad s = \sqrt{\frac{20}{3}}$$

3. Divide the estimated standard deviation by the square root of n .

$$\widehat{SE} = \frac{s}{\sqrt{4}} = \sqrt{\frac{5}{3}} \approx 1.291$$

Answer. The mean estimate is 7, the variance estimate is $20/3$, and the estimated standard error is approximately 1.291.

7. Choose the appropriate standard error

If the population standard deviation is known, use $\sigma/\sqrt{n}$. If it is unknown, calculate s from the sample and use the estimated standard error $s/\sqrt{n}$. Estimating the standard error does not, by itself, justify an exact normal confidence interval or test.

COMMON MISTAKE An estimate does not become a known parameter

Replacing σ by s introduces additional uncertainty. A suitable large sample can support approximate normal inference. For a small sample from a normal population with unknown variance, the t methods in Further Statistics are needed.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

A normal population has mean 30 and variance 16. For 16 independent random observations, find the distribution of $\bar{X}$ and $P(29.5 < \bar{X} < 31)$.

1. Divide the variance by n and take its square root.

$$\bar{X} \sim N(30, 16/16) = N(30, 1), \quad SE = 1$$

2. Standardise both endpoints.

$$P(-0.5 < Z < 1) = \Phi(1) - \Phi(-0.5) \approx 0.8413 - 0.3085$$

Answer. $\bar{X} \sim N(30, 1)$ and the probability is approximately 0.5328.

PRACTICE Question 2

For the independent random sample 3, 5, 5, 7, estimate the population mean, variance and standard error of the sample mean.

1. Calculate the mean and sum of squared deviations.

$$\bar{x} = 5, \quad \sum (x_i - \bar{x})^2 = 4 + 0 + 0 + 4 = 8$$

2. Use n minus one for variance and the square root of n for standard error.

$$s^2 = \frac{8}{3}, \quad \widehat{SE} = \frac{\sqrt{8/3}}{2} = \sqrt{\frac{2}{3}}$$

Answer. Mean estimate = 5, variance estimate = 8/3, estimated standard error ≈ 0.8165 .

PRACTICE Question 3

An independent random sample has $n = 10$, $\sum x_i = 52$ and $\sum x_i^2 = 300$. Find $\bar{x}$, s^2 and the estimated standard error.

1. Use the summarised-data formula.

$$\bar{x} = \frac{52}{10} = 5.2, \quad s^2 = \frac{300 - 52^2/10}{9} = \frac{148}{45}$$

2. Take the square root after dividing the estimated variance by n.

$$\widehat{SE} = \sqrt{\frac{148/45}{10}} \approx 0.5735$$

Answer. $\bar{x} = 5.2$, $s^2 = 148/45 \approx 3.289$, and estimated standard error ≈ 0.5735 .

PRACTICE Question 4

(a) How must sample size change to halve the standard error when σ stays fixed? (b) Does this change guarantee that a sample mean from every population is normally distributed?

1. Set the new standard error equal to half the old one.

$$\frac{\sigma}{\sqrt{n_{\text{new}}}} = \frac{1}{2} \frac{\sigma}{\sqrt{n_{\text{old}}}} \Rightarrow n_{\text{new}} = 4n_{\text{old}}$$

2. Normality is exact for independent samples from a normal population. For another population, a suitable CLT approximation needs justification.

Answer. (a) Multiply sample size by 4. (b) No: a larger sample does not provide a universal normality guarantee.

Ready to move on?

- ☐ I can tell an individual observation from a sample mean.
- ☐ I can distinguish the random mean from its observed value.
- ☐ I can calculate variance and standard error on the correct scale.
- ☐ I can state when a normal sampling model is exact or approximate.
- ☐ I can calculate unbiased variance estimates from raw and summarised data.

Next: Hypothesis testing foundations. Use a sampling distribution to assess evidence about a population mean.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

[Cambridge 9709 syllabus, section 6.4: Sampling and estimation](#) · [Cambridge 9231 prior-learning guidance](#)