Chapter 7 From sampling to estimation

Detailed notes pp. 1–4 and 12

Lecture slides PDF pp. 2–42

A confidence interval estimates a fixed population parameter using a random sample. Start by naming the parameter and the population it describes.

$$\boxed{\text{estimate } \pm \text{ critical value} \times \text{SE used}}$$

The product is the **margin of error**; total interval width is twice that margin. For normal-based two-sided confidence $1-\alpha$, use the upper-tail cutoff $z_{\alpha/2}$.

Interpret confidence as a procedure's coverage

Under the model, repeated samples would produce intervals containing the fixed true parameter about $100(1-\alpha)\%$ of the time. A particular observed interval either contains it or does not. The confidence level does not describe the fraction of individuals inside the interval or a posterior probability for the parameter.

Finite populations appear first in the supplied slides

Detailed notes pp. 12

For a simple random sample of size n without replacement from N values, let the finite-population variance σ^2 use divisor N. Then

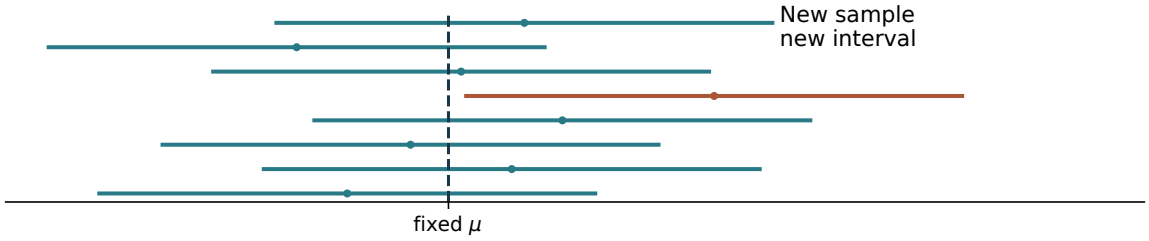
$$\mathrm{E}(\bar{X}) = \mu, \quad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}\frac{N-n}{N-1}, \quad \mathrm{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}.$$

The correction tends to 1 when the sampling fraction is small and to 0 for a census. It is a variance/SE result, not a claim of exact normal shape. With sample variance s^2 defined using divisor $n-1$, the corresponding estimated mean variance is $(1-n/N)s^2/n$.

What changes interval width?

- Higher confidence increases the critical value and widens the interval.
- Larger sample size usually reduces SE and narrows the interval.
- More individual variability increases SE and widens the interval.

A narrow interval does not fix a biased sample or an incorrect model.



The parameter stays fixed; the intervals vary from sample to sample.

Keep in mind. Follow the slide order using the section index. Finite-population sampling is on detailed notes p. 12 even though it begins the lecture deck.

Chapter 7 Intervals for means

Detailed notes pp. 6–7 and 15
Lecture slides PDF pp. 24–53

One population mean

Detailed notes pp. 2: known-SD interval

For independent sampling and known population SD,

$$CI_{1-\alpha}(\mu): \quad \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Coverage is exact under normal sampling with known σ; otherwise a suitable CLT gives an approximation. If σ is unknown, Chapter 7 uses the *large-sample approximate* interval

$$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}.$$

Replacing σ by s does not produce an exactly standard-normal statistic at $n = 30$. For independent normal observations with unknown σ, Chapter 9 uses the exact t_{n-1} reference.

Two independent population means

Fix an order, such as treatment minus control, and keep it throughout:

$$\widehat{\Delta} = \bar{x}_1 - \bar{x}_2, \quad \widehat{\mathrm{SE}}(\widehat{\Delta}) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

$$CI_{1-\alpha}(\mu_1 - \mu_2): \quad (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

This is a large-sample approximation for independent groups. If population SDs are known, use them in place of s_i; normal populations then give an exact normal reference. The formula is not appropriate for paired measurements treated as independent.

Why variances add

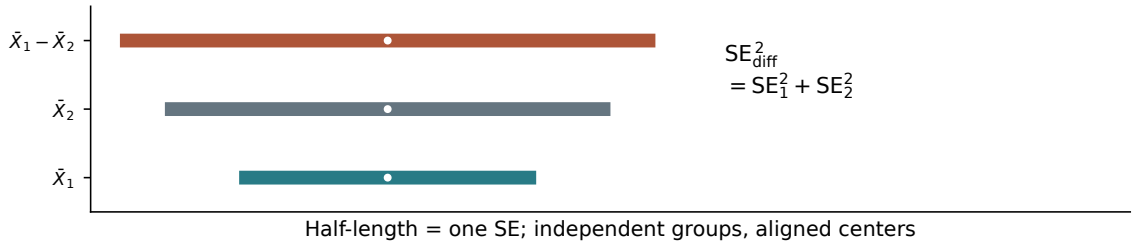
For independent sample means, $\mathrm{Var}(\bar{X}_1 - \bar{X}_2) = \mathrm{Var}(\bar{X}_1) + \mathrm{Var}(\bar{X}_2)$. The minus sign is squared in the variance calculation. Take a square root only after adding variance contributions.

Sample variance and interpretation

Detailed notes pp. 15

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}, \qquad \mathrm{E}(S^2) = \sigma^2 \text{ under iid sampling.}$$

One fitted mean leaves $n - 1$ degrees of freedom (derivation: detailed p. 15). A difference interval containing zero does not establish equality.



For independent estimates, add their variances before taking a square root.

Keep in mind. SD describes individuals; SE describes estimates. State units.

Chapter 7 Proportions and their differences

Detailed notes pp. 8–10
Lecture slides PDF pp. 54–71

A proportion concerns a binary outcome. With x successes in n trials, $\widehat{p} = x/n$ and $\widehat{q} = 1 - \widehat{p}$.

Known model probability versus estimated proportion

If p is specified, independent Bernoulli observations give

$$\mathrm{E}(\widehat{p}) = p, \qquad \mathrm{SE}(\widehat{p}) = \sqrt{\frac{p(1-p)}{n}}.$$

Use this model SE for a sampling-probability question. To estimate an unknown p, substitute observed $\widehat{p}$:

$$CI_{1-\alpha}(p): \quad \widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$$

The course guideline checks observed successes $n\widehat{p} > 5$ and failures $n(1-\widehat{p}) > 5$ for this approximate normal interval. Independence and a small sampling fraction also matter. A very small or near-boundary proportion can make this interval unreliable; endpoints outside $[0, 1]$ warn of a poor normal approximation.

Two independent proportions

$$CI_{1-\alpha}(p_1 - p_2): \quad (\widehat{p}_1 - \widehat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}.$$

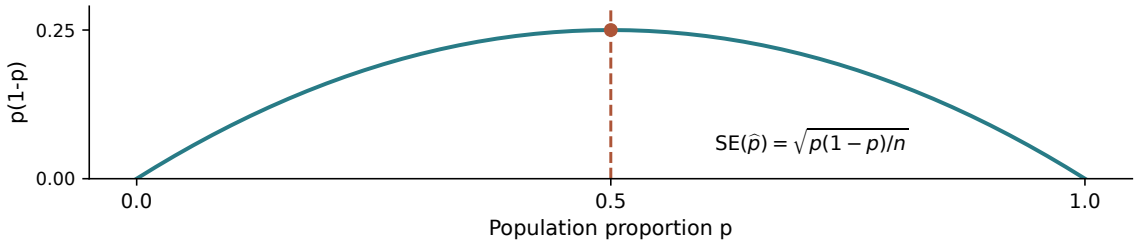
Check success and failure counts separately in *each* group. The SE uses each group's own observed proportion. Do not pool the proportions here: the interval is estimating an unknown difference without imposing equality. Chapter 8 pools specifically for a test of $H_0 : p_1 = p_2$.

Percentages and percentage points

A proportion difference .08 is an 8-percentage-point difference. It is not automatically an 8% relative increase. State both populations and the subtraction order before interpreting the sign.

A link back to Chapter 6

A probability such as $P(\widehat{p} \geq c)$ with stipulated p concerns a random sample result. A confidence interval after observing x estimates a fixed unknown p. These use different information and different SE substitutions.



At fixed sample size, proportion variability is largest at $p = .5$.

Keep in mind. The existing water-sample Problem 3 asks you to distinguish these two tasks. Try the question before opening its released solution.

Chapter 7 Sample size and final checks

Detailed notes pp. 11–14
Lecture slides PDF pp. 72–83

Choose a precision target before collecting data

For a desired two-sided normal-interval margin at most ε,

$$n \geq \left(\frac{z_{\alpha/2}\sigma}{\varepsilon}\right)^2 \quad \text{for one mean,}$$

using a justified planning value for σ if unknown. For a proportion,

$$n \geq \frac{z_{\alpha/2}^2 p(1-p)}{\varepsilon^2}.$$

Use a justified planning estimate or, with no prior information, $p = .5$ because $p(1-p) \leq 1/4$. This choice gives the largest required n. Round the final lower bound *up* to an integer. This gives total sample size unless additional observations are requested.

A target of four percentage points is $\varepsilon = .04$, not 4. A target interval *width* of w means margin $\varepsilon = w/2$.

Planning shortcuts need judgment

The slide's range/4 estimate for SD is a rough planning heuristic motivated by a typical normal spread. A sample minimum-to-maximum range depends on sample size; it is not a fixed four-SD interval. Prefer substantive prior information when available.

Recover information from an interval

For symmetric endpoints $[L, U]$,

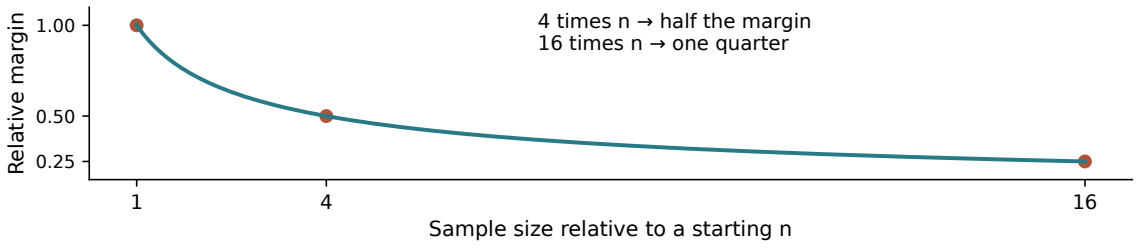
$$\text{estimate} = \frac{L+U}{2}, \qquad M = \frac{U-L}{2}, \qquad \text{SE used} = \frac{M}{z_{\alpha/2}}.$$

Rounded endpoints give approximate recovered values. A margin alone cannot determine both n and the unknown variability.

Before turning to solutions

1. Name the parameter, units, design and task (probability, estimation or planning).
2. Check independence, approximation conditions and the sampling fraction.
3. Show the estimate, SE, critical value and margin separately.
4. Interpret confidence and the subtraction order carefully.

Existing practice: Practice 1, p. 5; Problems 2–3, pp. 16–17. Released solutions are separately labeled on the course page. Keep them closed for the first attempt.



Hold confidence and variability fixed: halving the margin needs four times n.