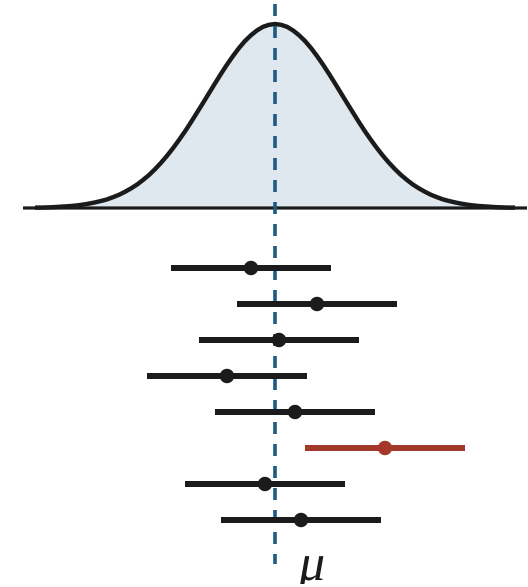


Statistics

What can one sample tell us about a population, and how sure can we be?

1 Sampling → 2 Intervals → 3 Tests



Lecture 1

Probability theory

What can happen, and how does information change it?

Lecture 2

Data, variables & moments

How do measurements vary?

Lecture 3

Statistics

What can one sample tell us about a population?

Lecture 4

Time series

What changes when observations depend on their past?

Lecture 5

Causal inference

Which comparison identifies an effect?

OPENING QUESTION

Smoking falls from 30% to 25% in both tables. Is the evidence the same?

100 people in a cessation programme, asked before and after; two constructed teaching tables with the same margins.

Table A: rows before, columns after

	smokes	not
smokes	20	10
not	5	65

Table B: rows before, columns after

	smokes	not
smokes	25	5
not	0	70

A

A is
stronger

B

B is
stronger

C

The same

YOUR ANSWER

Vote now. Part 2 answers it.

WHY START HERE

Before-and-after data on the same people are everywhere in programme evaluation. Whether a change is real depends on how the data were collected, not only on the headline percentages.

LECTURE 3 · THE ROUTE

From one sample to a range for the truth and a verdict on a claim

PART 1

Samples and sampling distributions

What can one sample tell us about a population, and how much would another sample differ?

population iid estimator
standard error CLT

Notes 3.1–3.3

→

PART 2

Confidence intervals

How precisely can one sample pin down a mean, a proportion or a difference, and what does “95% confidence” promise?

confidence interval coverage
critical value paired data

Notes 3.4–3.5

→

PART 3

Hypothesis tests

When is a sample result strong enough evidence against a claim, and what do α , the p-value and “fail to reject” actually say?

null hypothesis Type I error power
p-value duality

Notes 3.6–3.8

BY THE END OF THIS LECTURE YOU CAN

- ✓ Distinguish a parameter, an estimator and an estimate, and say what the selection mechanism lets a sample tell us.
- ✓ Derive the standard error $\sigma/\sqrt{n}$, and use the LLN and CLT to justify an approximately normal estimate.
- ✓ Construct and interpret 95% intervals for a mean, a proportion or a difference, matching the SE to the design.
- ✓ Carry out a two-sided z test and interpret α , power and the p-value correctly.

PART 1 OF 3

Samples and sampling distributions

What can one sample tell us about a population, and how much would another sample differ?

KEY TERMS

population

iid

estimator

standard error

CLT

[Notes 3.1–3.3 ↗](#)

WHY IT MATTERS

Economists mostly work with samples: income surveys, polls and experiments. A sample average is informative only if the way units were selected targets the right population, and only as precise as its standard error.

IN THIS PART

- 3.1 Who is sampled, and how
- 3.2 Parameter, estimator and estimate
- 3.2 Bias and the standard error
- ▶ Animation: where the square root of n comes from
- 3.3 The law of large numbers and the central limit theorem

3.1 | TARGET, DESIGN AND IID

Say who is sampled and how before using any formula

DEFINITION

Population and sample

The **population** is what we want to describe; the sample is what a stated selection mechanism delivers from a frame, the list units are drawn from.

- Who is missing from the frame?
- Who is selected, and who responds?
- Which steps are random?

COMMON MISTAKE

More respondents do not remove nonresponse bias: a very large sample from the wrong group answers the wrong question very precisely.

DEFINITION

iid

Independent and identically distributed: the draws are jointly independent and each has the same distribution.

WORKED EXAMPLE *ten examination scripts, one graded 90*

Draw two without replacement. Before any draw, $P(X_2 = 90) = 1/10$, the same as X_1 . But

$$P(X_2 = 90 \mid X_1 = 90) = 0 \neq \frac{1}{10}.$$

Identical, yet dependent. With replacement, the draws are iid.

3.2 | PARAMETER, ESTIMATOR, ESTIMATE

A parameter is fixed, an estimator is a random rule, an estimate is one number

Object	Random?	Example
Parameter θ	fixed, unknown	population mean μ
Estimator $\hat{\theta}$	random	$\bar{X}$
Estimate	a number	$\bar{x} = 2.5$

DEFINITION

Sampling distribution

The distribution of an estimator across repeated samples, not the spread of the data within one sample.

KEY FORMULA

DEFINITION

(1)

Sample mean and variance

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

IN WORDS divide by $n - 1$: the n deviations obey one constraint.

WORKED EXAMPLE six observations: 1, 2, 1, 3, 4, 4

$\bar{x} = 15/6 = 2.5$; the squared deviations sum to 9.5, so

$$s^2 = 9.5/5 = 1.9, \quad s \approx 1.378.$$

3.2 | BIAS AND STANDARD ERROR

Judge an estimator by its centre and its spread: bias and standard error

DEFINITION

Bias

$\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$, averaged over repeated samples. $\bar{X}$ is unbiased: $E[\bar{X}] = \mu$ needs only common means.

KEY FORMULA RESULT

(2)

Standard error of the mean

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}, \quad \text{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

IN WORDS estimated by $s/\sqrt{n}$.

NEEDS independent draws with a common variance; dependence adds covariance terms.

WORKED EXAMPLE *one score against the mean of 20*

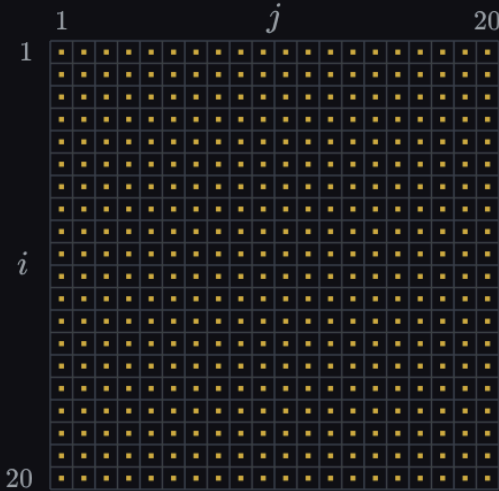
Scores uniform on 0, ..., 100: $\sigma^2 = 850$, so one score has SD $\sqrt{850} \approx 29.15$. The mean of 20 independent scores has

$$\text{SE} = \sqrt{850/20} = \frac{29.15}{\sqrt{20}} \approx 6.52.$$

COMMON MISTAKE

The SD describes individuals; the **standard error** describes the estimator across samples. For 1, 2, 1, 3, 4, 4: $s \approx 1.378$ but $\widehat{\text{SE}} \approx 0.563$.

How much does a mean of 20 scores vary?


$$\text{cell}(i, j) = \text{Cov}(X_i, X_j)$$

spread evenly: 42.5 per cell

$$\begin{aligned}\text{Var}(X_1 + \cdots + X_{20}) &= \text{all 400 cells} \\ &= 20 \times 850 = 17000\end{aligned}$$

$$\text{Var}(\bar{X}) = 17000/20^2 = 42.5$$

the average cell: 850/20



X_i uniform on $0, \dots, 100$: $\sigma^2 = 850$

$$\text{SE}(\bar{X}) = \sqrt{42.5} = 29.15/\sqrt{20} \approx 6.52$$

3.3 | LARGE SAMPLES

Averages concentrate, and their errors become approximately normal

RESULT *the law of large numbers*

For iid data with finite mean and variance, $\bar{X}_n$ converges in probability to μ : $P(|\bar{X}_n - \mu| > \varepsilon) \rightarrow 0$.

KEY FORMULA RESULT

(3)

Central limit theorem

$$\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} N(0, 1)$$

IN WORDS for large n , $\bar{X}_n$ is approximately $N(\mu, \sigma^2/n)$, and s may replace σ .

NEEDS iid data with $0 < \sigma^2 < \infty$.

WORKED EXAMPLE *a U-shaped population*

A $\text{Beta}(\frac{1}{2}, \frac{1}{2})$ variable has SD 0.3536. Means of 50 have SE $\sqrt{(1/8)/50} = 0.05$ and a bell-shaped histogram; at $n = 200$ the SE halves to 0.025. The data stay U-shaped.

COMMON MISTAKE

- The CLT is about means, not the data.
- There is no universal “ $n = 30$ ”.
- No sample size cures a biased selection.

TAKEAWAY An estimate is one draw from a sampling distribution: centred on its target if unbiased, with spread $\sigma/\sqrt{n}$.

PART 2 OF 3

Confidence intervals

How precisely can one sample pin down a mean, a proportion or a difference, and what does “95% confidence” promise?

KEY TERMS

confidence interval

coverage

critical value

paired data

[Notes 3.4–3.5 ↗](#)

SO FAR an estimate is one draw from a sampling distribution; its standard error $\sigma/\sqrt{n}$ measures the spread, and the CLT makes it approximately normal.

WHY IT MATTERS

Decisions need a range, not a point. A 39% approval rate means little without its 2.47-point margin, and a fall in smoking after a programme is convincing or not depending on how many people actually changed.

IN THIS PART

- 3.4 A confidence interval and its coverage
- ▶ Animation: what the 95% in a confidence interval describes
- 3.4 One recipe, three designs
- 3.5 A proportion is a mean of 0/1 data
- 3.5 The opening question answered

3.4 | BUILDING AN INTERVAL

A confidence interval turns a statement about the estimate into a range for the target

KEY FORMULA RESULT

(4)

Confidence interval for a mean

$$\bar{X} \pm c \text{SE}$$

IN WORDS $P(\bar{X} - c \text{SE} \leq \mu \leq \bar{X} + c \text{SE}) = 1 - \alpha$: the random interval covers the fixed μ with probability $1 - \alpha$.

NEEDS exact for normal data with known σ ; approximate for large samples with $s/\sqrt{n}$.

DEFINITION

Coverage

The share of repeated samples whose interval contains the parameter. It belongs to the procedure, not to one computed interval.

Critical values

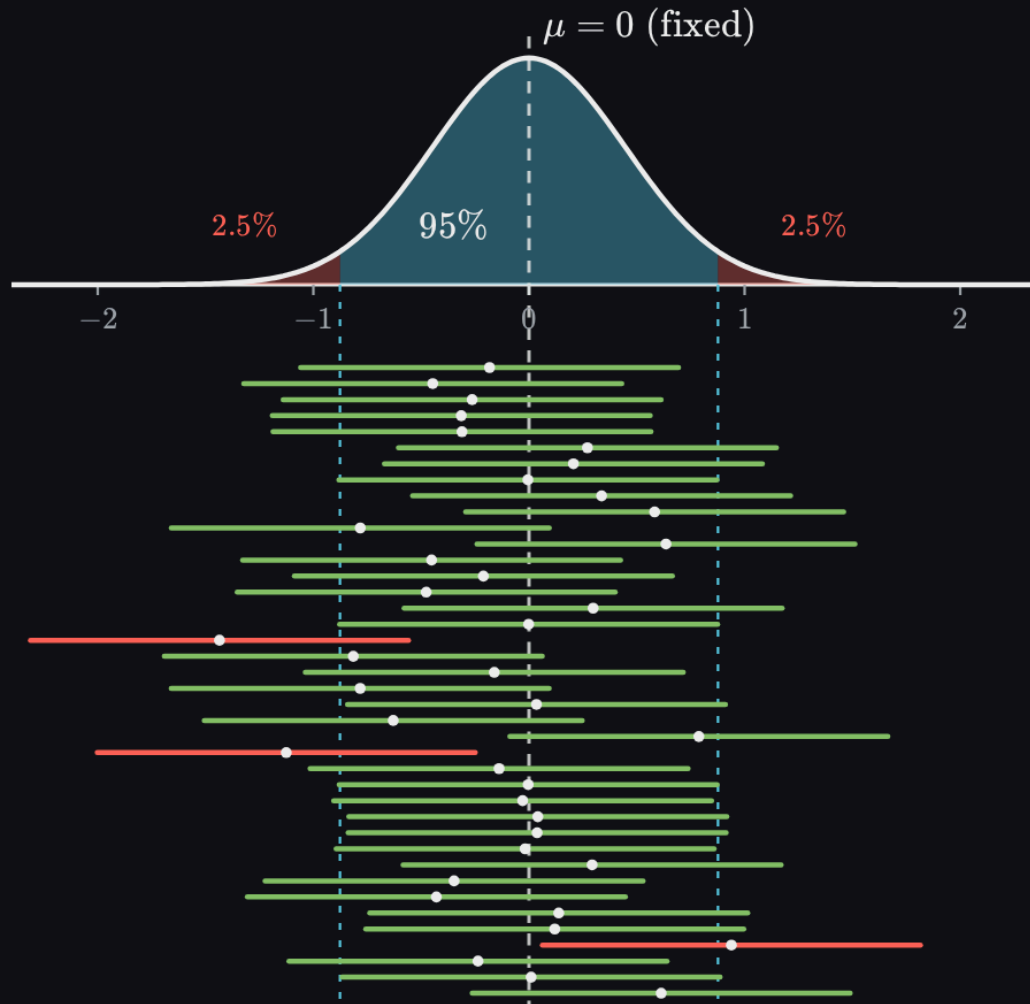
Confidence	90%	95%	99%
c	1.645	1.960	2.576

RESULT *width*

The width is $2c \text{SE}$: more confidence widens it; four times the sample size halves it.

The algebra moves the randomness to the endpoints; μ stays fixed.

What does 95% confidence describe?



$n = 5$ from $N(0, 1)$: $\bar{X} \sim N(0, 1/5)$

$1.96/\sqrt{5} \approx 0.877$: $P(|\bar{X} - \mu| \leq 0.877) = 0.95$

each sample: the interval $\bar{x} \pm 0.877$

$1.00 \pm 0.877 = [0.123, 1.877]$ misses 0

a miss $\iff \bar{x}$ lands outside the 95% band

this run: 37 of 40 contain 0

95% describes the procedure over many samples;
one computed interval either contains μ or not.

3.4 | MEANS IN THREE DESIGNS

One recipe, three designs: the design decides the standard error

Design	Example	Estimate	SE	95% interval
One mean	$n = 100, \bar{x} = 64, s^2 = 8$	64	0.283	[63.446, 64.554]
Independent groups	anchoring prompt 10 vs 65 (43 and 46 students)	-13.6	$\sqrt{7.5} \approx 2.739$	[-18.968, -8.232]
Paired	71 students, exam 1 minus exam 2	-1.8	$\sqrt{124.45/71} \approx 1.324$	[-4.395, 0.795]

KEY FORMULA RESULT

(5)

Standard errors by design

$$\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} \quad \text{or} \quad \frac{s_D}{\sqrt{n}}$$

IN WORDS independent groups: variances add. Paired: analyse the n differences, $s_D^2 = s_X^2 + s_Y^2 - 2s_{XY}$.

COMMON MISTAKE

Treating paired scores as independent:

$\sqrt{117/71 + 151/71}$ ignores the positive covariance and overstates the SE. And [-4.395, 0.795] including 0 does not prove the means equal.

3.5 | PROPORTIONS

A proportion is a mean of 0/1 data, with its own standard error

KEY FORMULA RESULT

(6)

Interval for a proportion (Wald)

$$\hat{p} \pm c \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

IN WORDS $\text{Var}(\hat{p}) = p(1 - p)/n$, estimated by plugging in $\hat{p}$.

NEEDS a large iid sample, away from 0 and 1.

WORKED EXAMPLE *an approval survey*

$n = 1502$, $\hat{p} = 0.39$, treated as an iid unweighted sample:

$$\text{SE} = \sqrt{\frac{0.39(0.61)}{1502}} \approx 0.0126$$

$0.39 \pm 1.96(0.0126)$: from 36.53% to 41.47%, a margin of about 2.47 percentage points, not 2.47%.

COMMON MISTAKE

At 0 successes in 10 the Wald interval is $[0, 0]$. Clipping does not repair it; the Wilson interval, about $[0, 0.2775]$, does better.

3.5 | PAIRED PROPORTIONS

Same fall from 30% to 25%, different uncertainty: the people who changed decide

Each person's change D_i is -1 (quit), 0 (no change) or $+1$ (start). The SE is $s_D/\sqrt{n}$, and only people who changed contribute to s_D .

Table A: 10 quit, 5 start: rows before, columns after

	smokes	not
smokes	20	10
not	5	65

Table B: 5 quit, 0 start: rows before, columns after

	smokes	not
smokes	25	5
not	0	70

WORKED EXAMPLE *the two tables*

Table A: $s_D^2 = 14.75/99$, $SE \approx 0.0386$:

$$-0.05 \pm 1.96(0.0386) \approx [-0.126, 0.026]$$

Table B: $s_D^2 = 4.75/99$, $SE \approx 0.0219$:

$$-0.05 \pm 1.96(0.0219) \approx [-0.093, -0.007]$$

B's interval rests on only five changes, so an exact test is safer; and a before–after fall is not, by itself, the programme's effect.

TAKEAWAY The interval is estimate $\pm$ critical value $\times$ SE, with the SE matched to the design: one sample, independent groups or pairs.

PART 3 OF 3

Hypothesis tests

When is a sample result strong enough evidence against a claim, and what do α , the p-value and “fail to reject” actually say?

KEY TERMS

null hypothesis

Type I error

power

p-value

duality

[Notes 3.6–3.8 ↗](#)

SO FAR a 95% interval is the estimate ± 1.96 standard errors, with the standard error matched to the design.

WHY IT MATTERS

Economists test claims and act on them: an evenly split electorate, no anchoring effect, no change after a programme. Reading a p-value as the probability that a claim is true, or “not significant” as “no effect”, produces false findings.

IN THIS PART

3.6 What a test controls

3.6 One test from start to finish

▶ Animation: what a p-value is an area of

▶ Animation: what a stricter level costs in power

3.7 Tests for means and proportions

3.8 A confidence interval is the set of nulls not rejected

3.6 | HYPOTHESES AND ERROR RATES

A test limits how often it rejects a true claim, not the chance that the claim is true

State $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ about a parameter, with the direction and level fixed in advance.

	H_0 true	H_0 false
Reject	Type I error	correct: power
Do not reject	correct	Type II error

KEY FORMULA DEFINITION

(7)

Level and power

$$P(\text{reject } H_0 \mid H_0 \text{ true}) \leq \alpha$$

IN WORDS power $1 - \beta(\theta)$ is the probability of rejecting at a given alternative.

CHECK

A 5% test rejects. Is there now a 5% chance that H_0 is true?

ANSWER

No: α bounds $P(\text{reject} \mid H_0)$, not $P(H_0 \mid \text{reject})$.
Reversing the condition needs Bayes' rule (Lecture 1).

MANY TESTS

Run 20 tests of true nulls at 5% and $1 - 0.95^{20} \approx 0.64$ is the chance of at least one false finding. Bonferroni uses α/m .

3.6 | THE STATISTIC AND THE P-VALUE

Count the standard errors from the claim; the p-value is the tail beyond them

KEY FORMULA

DEFINITION

(8)

Test statistic

$$Z = \frac{\hat{\theta} - \theta_0}{\widehat{SE}_0(\hat{\theta})}$$

IN WORDS the distance from the claim in null standard errors; reject at 5% if $|Z| > 1.96$.

KEY FORMULA

DEFINITION

(9)

Two-sided p-value

$$p_{\text{val}} = 2\{1 - \Phi(|z_{\text{obs}}|)\}$$

IN WORDS the probability, if H_0 were true, of a statistic at least as extreme as the one observed.

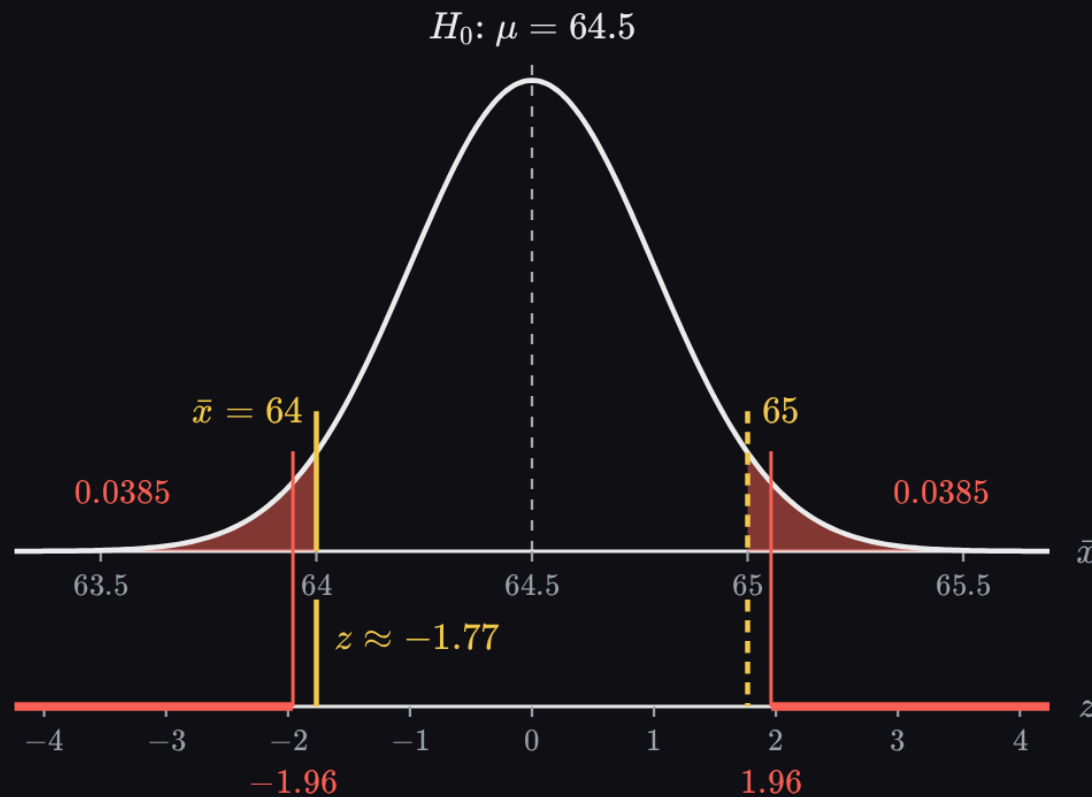
WORKED EXAMPLE $H_0 : \mu = 64.5$

1. $\bar{x} = 64$, $SE \sqrt{8/100} \approx 0.283$.
2. $z = (64 - 64.5)/0.283 \approx -1.77$.
3. $p_{\text{val}} = 2\{1 - \Phi(1.77)\} \approx 0.077$.
4. $0.077 > 0.05$: fail to reject. Report 64, $[63.45, 64.55]$ and $p_{\text{val}} \approx 0.077$.

COMMON MISTAKE

“We failed to reject, so $\mu = 64.5$.” The data are just as compatible with 64.3.

What is a p-value an area of?



$$H_0: \mu = 64.5; n = 100, s^2 = 8$$

$$SE = \sqrt{8/100} \approx 0.2828; \bar{x} = 64$$

$$z = (64 - 64.5)/0.2828 \approx -1.77$$

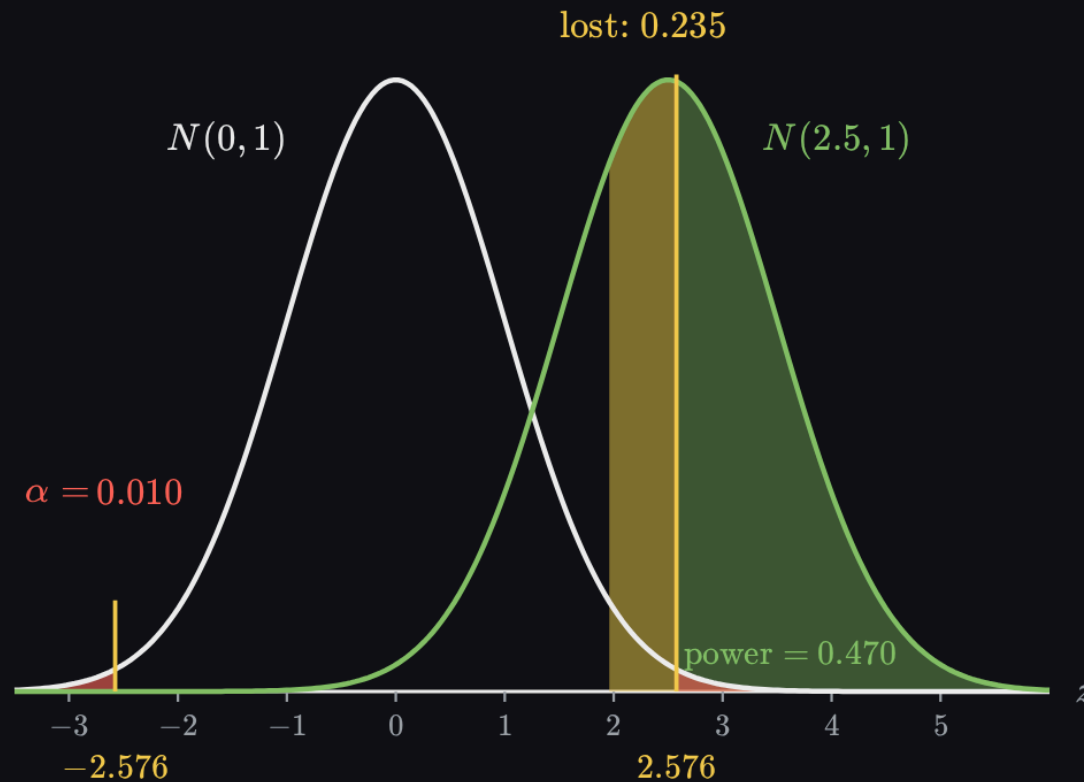
$$p = P(|Z| \geq 1.77 \mid H_0)$$

$$= 2\{1 - \Phi(1.77)\} \approx 0.077$$

$0.077 > 0.05$: fail to reject H_0 at 5%
equivalently $|z| = 1.77 < 1.96$

The p-value assumes H_0 ; it is not the probability that H_0 is true.

What does a stricter test cost?



H_0 true: $Z \sim N(0, 1)$

true mean 2.5 SE above: $Z \sim N(2.5, 1)$

5%: reject if $|Z| > 1.96$

power = $\Phi(2.5 - 1.96) = \Phi(0.54) \approx 0.705$

1%: reject if $|Z| > 2.576$

power $\approx \Phi(2.5 - 2.576) = \Phi(-0.076) \approx 0.470$

A smaller Type I error bound costs power, unless the sample size rises.

3.7 | TESTS BY DESIGN

The same designs as the intervals, with the standard error computed under the null

Design	H_0	SE under H_0	z	p_{val}	At 5%
One mean	$\mu = 65$	$s/\sqrt{n} = 0.283$	-3.54	0.0004	reject
Independent means	no anchoring effect	$\sqrt{7.5}$	-4.97	< 0.0001	reject
Paired means	equal exam means	$\sqrt{124.45/71}$	-1.36	0.17	do not reject
One proportion	$p = 0.5$	$\sqrt{p_0(1 - p_0)/n}$	-8.53	< 0.0001	reject
Two proportions	$p_X = p_Y$	pooled $\hat{\pi} \approx 0.390$	3.58	0.0003	reject
Paired proportions	no change (Table A)	$s_D/\sqrt{n}$	-1.30	0.20	do not reject

COMMON MISTAKE

For proportions the null fixes the SE: use p_0 , not $\hat{p}$, for one proportion, and pool only when testing equality.

3.8 | TESTS AND INTERVALS AGREE

An interval is every claim the matching test keeps

KEY FORMULA RESULT

(10)

Inverting a test

$$\left| \frac{\hat{\theta} - \theta_0}{e} \right| \leq c \iff \theta_0 \in [\hat{\theta} - ce, \hat{\theta} + ce]$$

IN WORDS a value is retained by the test exactly when it lies in the interval.

NEEDS the same data, SE e , reference distribution, tails and level.

WORKED EXAMPLE $\bar{x} = 64$, $SE\ 0.283$, *interval* $[63.446, 64.554]$

$\mu_0 = 64.2$: $z \approx -0.71$, $p_{\text{val}} \approx 0.48$, retained.

$\mu_0 = 65$: $z \approx -3.54$, $p_{\text{val}} \approx 0.0004$, rejected.

COMMON MISTAKE

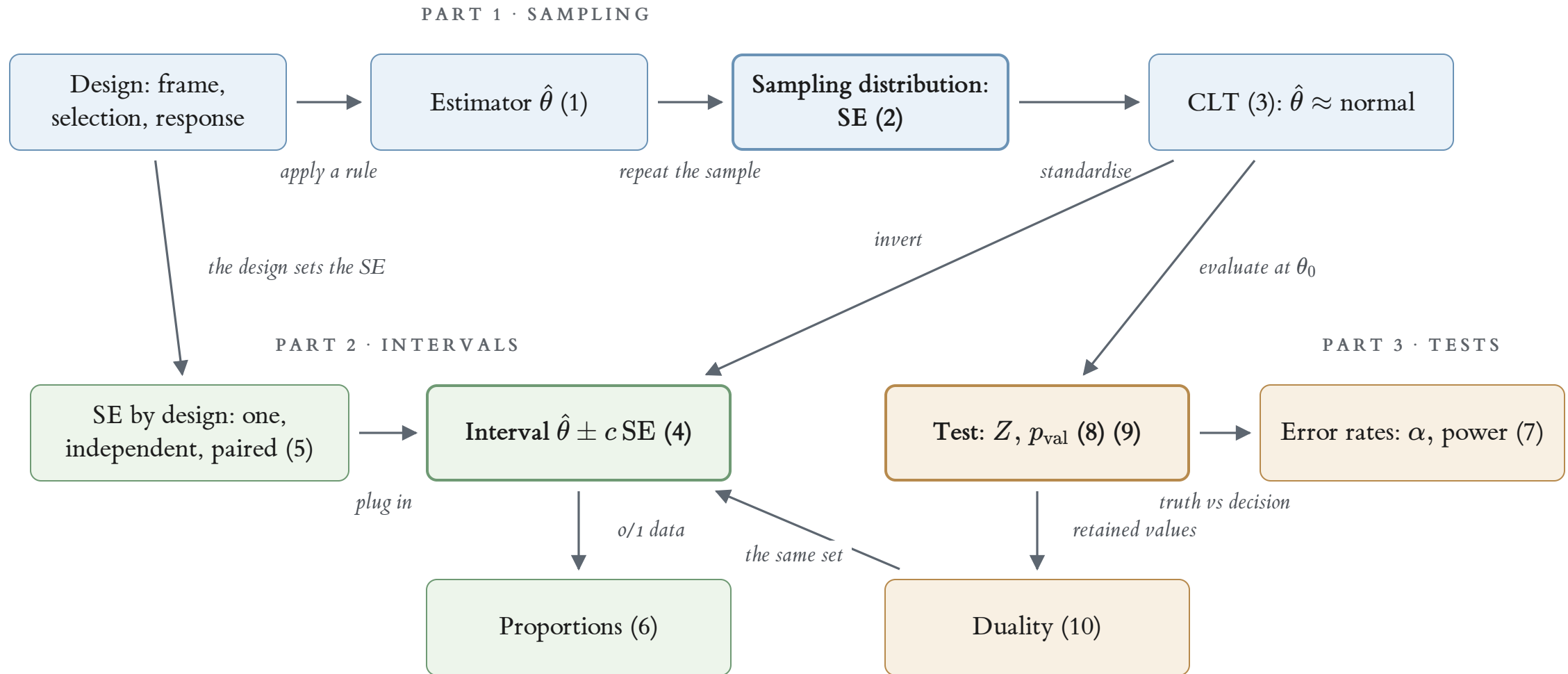
For a proportion the score test (SE from p_0) inverts to the Wilson interval, not the Wald interval: only a matching pair agrees.

Report the estimate and its interval, not just “significant”: significant is not important, and not rejecting is not equivalence.

TAKEAWAY A test counts null standard errors from the claim; the interval is every claim it would not reject.

REVIEW · HOW THE IDEAS CONNECT

One chain: design, estimator, standard error, then interval and test



REVIEW · SUMMARY SHEET

What to take away: the terms and the ten formulas

KEY TERMS

1 SAMPLING

Population the collection or distribution we want to make a statement about.

Iid independent draws, each with the same distribution.

Estimator $\hat{\theta} = g(X_1, \dots, X_n)$: a rule applied to the random sample.

Sampling distribution how an estimator varies across repeated samples.

Bias $E[\hat{\theta}] - \theta$: the average error over repeated samples.

Standard error the SD of an estimator's sampling distribution.

2 INTERVALS

Coverage the share of repeated samples whose interval contains the parameter.

Critical value $c = z_{1-\alpha/2}$: 1.645, 1.960, 2.576 for 90, 95, 99%.

3 TESTS

Type I error rejecting a true null hypothesis.

KEY FORMULAS

1 SAMPLING

(1) **Sample mean and variance** DEFINITION

$$\bar{X} = \frac{1}{n} \sum_i X_i; s^2 \text{ divides by } n-1$$

iid, finite moments, $n > 1$

(2) **Standard error of the mean** RESULT

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}, \quad \text{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

independence and identical variances

(3) **Central limit theorem** RESULT

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

iid, $0 < \sigma^2 < \infty$

2 INTERVALS

(4) **Confidence interval for a mean** RESULT

$$\bar{X} \pm c \text{SE}$$

coverage $1 - \alpha$; large samples use $s/\sqrt{n}$

(5) **Standard errors by design** RESULT

$$\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} \quad \text{or} \quad \frac{s_D}{\sqrt{n}}$$

independence between groups, or between pairs

(6) **Interval for a proportion (Wald)** RESULT

$$\hat{p} \pm c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

large iid sample, away from 0 and 1

3 TESTS

(7) **Level and power** DEFINITION

$$P(\text{reject } H_0 \mid H_0 \text{ true}) \leq \alpha$$

fix the level and direction before the data

(8) **Test statistic** DEFINITION

$$Z = \frac{\hat{\theta} - \theta_0}{\text{SE}_0(\hat{\theta})}$$

approximately $N(0, 1)$ under H_0

(9) **Two-sided p-value** DEFINITION

$$p_{\text{val}} = 2\{1 - \Phi(|z_{\text{obs}}|)\}$$

computed under H_0 ; one tail only if chosen in advance

(10) **Inverting a test** RESULT

$$\left| \frac{\hat{\theta} - \theta_0}{e} \right| \leq c \iff \theta_0 \in [\hat{\theta} - ce, \hat{\theta} + ce]$$

same SE, reference distribution, tails and level

REVIEW · BACK TO THE OPENING QUESTION

Same 5-point fall; only Table B's interval excludes zero

RESULT *the opening question*

The margins cannot decide it: the standard error comes from the people who changed. Table A, 15 changes: $[-0.126, 0.026]$. Table B, 5 changes: $[-0.093, -0.007]$.

REMEMBER

1. An estimate is one draw; its SE says how far another would land.
2. 95% describes the procedure, not one computed interval.
3. A p-value is computed assuming H_0 ; it is not the probability that H_0 is true.

BEFORE NEXT TIME

- Notes 3.1–3.8 with the worked solutions.
- The sampling, coverage, tails and duality figures in the notes.
- Rework the six tests in the table for yourself.

NEXT · LECTURE 4

Time series. When observations depend on their past, they are not iid: what changes for models, forecasts and standard errors?