The z Test and Decision Errors

Lectures 20–21

QBA201 – Quantitative Business Analysis

Shi Qiu – Fall 2026

From estimating a mean to checking a target

Recall the confidence interval

An interval estimates an unknown mean. Today the bottling team already has a target: **500 mL**.

The decision

A sample averages **503 mL**. Should the team investigate the process?

A sample average can differ from its target even when the process is centered correctly. We need a way to judge the gap.

Write the two claims before looking at the sample

Let μ be the population mean fill. The benchmark is $\mu_0 = 500$ mL.

$$H_0 : \ \mu = 500 \ \mathrm{mL} \qquad H_1 : \ \mu \neq 500 \ \mathrm{mL}$$

Which changes matter?

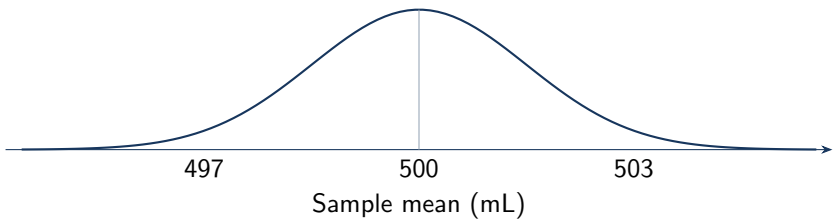
Underfilling and overfilling both matter. Look for departures in either direction.

How cautious should the rule be?

Set the significance level $\alpha = 0.05$ before sampling. The rule will flag about 5% of on-target samples under the model.

What sample averages look like under the null

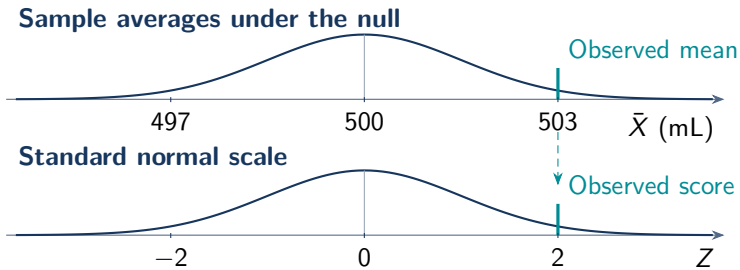
Assume independent normal bottle fills with known population SD $\sigma = 12$ mL. For $n = 64$ bottles, the spread of sample averages is $\text{SE} = \sigma/\sqrt{n} = 1.5$ mL.



If H_0 holds, sample averages cluster around 500 mL. The curve shows averages across repeated samples.

Put the observed gap on a common scale

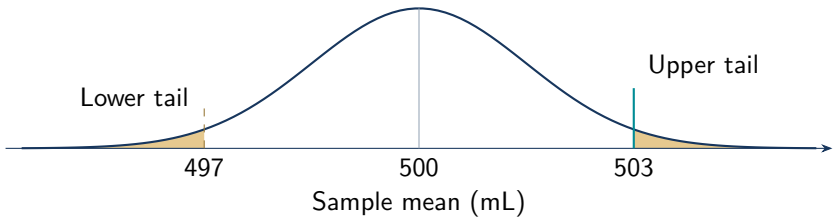
The observed sample mean $\bar{x} = 503$ mL is three mL above the null center.



$$z_{\text{obs}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{503 - 500}{1.5} = 2.$$

How unusual is a gap this large?

The question is two-sided. At least as far from 500 mL means **497 or below, or 503 or above**.



Give that area a name

The **p-value** is the probability, assuming H_0, of a result at least as extreme as the observed one in the direction or directions of H_1.

Use the evidence to make the decision

Let $\Phi(z)$ be the standard normal area to the left of z.

$$p = 2[1 - \Phi(2)] = 0.0455 < \alpha = 0.05.$$

Report the conclusion

Reject H_0. The sample gives evidence that the population mean fill differs from 500 mL.

State the limit

If $p > \alpha$, fail to reject. The data would be inconclusive at that level.

Whether an adjustment is worth making also depends on the size and cost of the fill difference.

Set the same rule using a cutoff

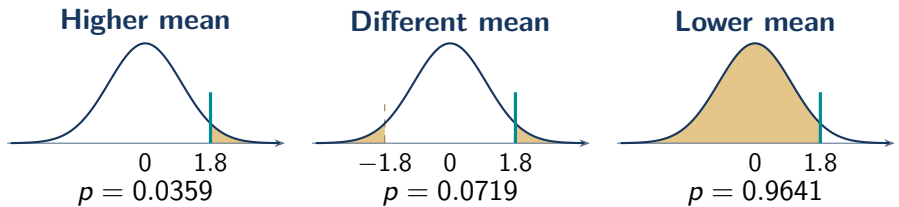
To reject in 5% of on-target samples, leave 2.5% in each tail. The boundaries are $z = \pm 1.96$.

Center the range on	What to check
Target mean, 500	503 falls outside $500 \pm 1.96(1.5) = [497.06, 502.94]$
Sample mean, 503	500 falls outside the 95% interval $[500.06, 505.94]$

The sample cutoff, the standardized cutoff and the matching two-sided confidence interval agree with the p-value decision.

A different business question changes the tail

Let c be a benchmark. Compare separate preplanned questions for the same $z_{\text{obs}} = 1.8$. For a one-sided test, use the null boundary $\mu = c$.



Higher: $H_0 : \mu \leq c$, $p = 1 - \Phi(z)$ Different: $H_0 : \mu = c$, $p = 2[1 - \Phi(|z|)]$ Lower: $H_0 : \mu \geq c$, $p = \Phi(z)$

Now evaluate the rule over many decisions

Decision	Process on target	Process has shifted
Investigate (reject H_0)	Type I: false alarm	Detect the shift
Continue (fail to reject)	Correct nonrejection	Type II: miss the shift

Why it matters

False alarms can stop a sound process. Missed shifts can leave underfills or overfills in production.

The level controls false alarms. To calculate the chance of a miss, specify a true shift.

A fixed rule has a false-alarm rate

Start with a rule, before any sample

Suppose the factory instead fixes this alarm: investigate when $|\bar{X} - 500| \geq 3$ mL.

$$\mathrm{P}_{H_0}(\text{alarm}) = \mathrm{P}(|Z| \geq 2) = 0.0455.$$

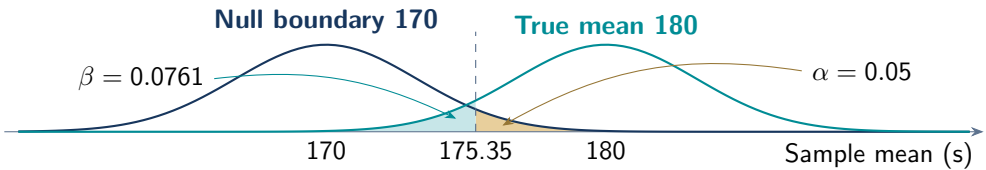
Keep the two questions separate

For a fixed boundary, the area evaluates the rule. For an observed result, the p-value evaluates how unusual that result is under the null.

Here the numbers agree because the fixed boundary happens to coincide with the earlier observed gap.

Keep one cutoff when the true mean changes

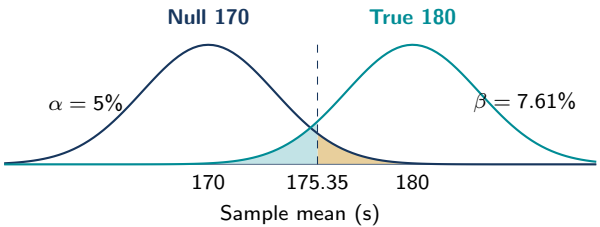
A billing team monitors delays. Test $H_0 : \mu \le 170$ against $H_1 : \mu > 170$ seconds. Use SE $= 3.25$ seconds and $\alpha = .05$.



Cutoff $= 170 + 1.6449(3.25) \approx 175.35$, $\beta(180) = 0.0761$, power $= 1 - \beta = 0.9239$.

Two ways to change how the rule performs

Keep the true delayed mean at 180 seconds. The last row uses four times as many independent observations, with the same population SD.



Policy	α	SE (s)	Power
Original rule	5%	3.250	92.39%
Stricter rule	1%	3.250	77.35%
Larger sample	5%	1.625	$>$ 99.99%

A stricter rule moves the cutoff. A larger valid sample narrows the curves. Decide how costly alarms and missed delays are before selecting a policy.

Choose the z test from the business question

The null hypothesis sets the benchmark. The question tells us what to measure and where its standard error comes from.

Business target	Sample statistic	Null SE uses
One mean	Mean	Known population SD
One proportion	Proportion	Benchmark proportion
Two independent means	Difference of means	Known SDs from both groups
Two independent proportions	Difference of proportions	Pooled under equality

In every case, measure the gap in standard errors. Then use the alternative to choose the normal tail or tails.

Set up a one-sample proportion test

Let p denote the population return proportion. A retailer tests $H_0 : p \geq .10$ against $H_1 : p < .10$. Among 400 independently sampled orders, 28 were returned.



At the boundary $p_0 = .10$, the expected counts are $400(.10) = 40$ returns and $400(.90) = 360$ other orders. Both exceed 5.

Use the null proportion in the test SE. The usual proportion confidence interval uses the observed proportion.

Test whether the return proportion is lower

The sample proportion is $\hat{p} = 28/400 = .07$.

$$\mathrm{SE}_0 = \sqrt{\frac{.10(.90)}{400}} = .015, \qquad z = \frac{.07 - .10}{.015} = -2.00.$$

Read the tail chosen before sampling

Alternative	p-value at $z = -2$	At 5%
$p < .10$ (planned)	Left: .0228	Reject
$p > .10$	Right: .9772	Fail to reject
$p \neq .10$	Both: .0455	Reject

The planned test supports a population return proportion below 10%.

Set up a two-sample test for means

A business compares average handling costs at two warehouses. Let μ_A and μ_B be their population mean costs. Each sample is random, comes from a normal population, and is independent of the other.

Warehouse	Sample size	Sample mean (AED)	Known population SD (AED)
A	64	108	12
B	64	102	16

$$H_0 : \mu_A - \mu_B = 0, \qquad H_1 : \mu_A - \mu_B \neq 0.$$

We observe one gap of AED 6. To judge it, account for sampling variability in both warehouses.

Test whether the two mean costs differ

Combine the sampling variances

$$\mathrm{SE}_0 = \sqrt{\frac{12^2}{64} + \frac{16^2}{64}} = 2.50 \text{ AED}, \qquad z = \frac{(108 - 102) - 0}{2.50} = 2.40.$$

Read the normal tails

The two-sided p-value is $2[1 - \Phi(2.40)] = .0164$. Reject at 5%. The data support different population mean costs.

For separate preplanned alternatives, $\mu_A > \mu_B$ uses the right tail (.0082) and $\mu_A < \mu_B$ uses the left tail (.9918).

The two sampling variances add because the samples are independent.

Set up a two-sample test for proportions

A retailer compares purchase proportions p_A and p_B at two sites. Sample visitors independently within and across sites.

Site	Visitors	Purchases	Sample proportion
A	300	90	.30
B	300	60	.20

$$H_0 : p_A = p_B, \qquad H_1 : p_A \neq p_B, \qquad \hat{p}_{\text{pool}} = \frac{90 + 60}{300 + 300} = .25.$$

Equality implies one common population proportion. Pool the counts to estimate it. Each site then has 75 expected purchases and 225 expected nonpurchases.

Test whether the two purchase proportions differ

Use the pooled standard error under equality

$$\mathrm{SE}_0 = \sqrt{.25(.75)\left(\frac{1}{300} + \frac{1}{300}\right)} \approx .03536, \qquad z = \frac{(.30 - .20) - 0}{\mathrm{SE}_0} \approx 2.83.$$

Read the normal tails

Two-sided $p \approx .0047$. Reject at 5%. The data support different population purchase proportions.

A preplanned claim that $p_A > p_B$ uses the right tail (.0023). A claim that $p_A < p_B$ uses the left tail (.9977).

Use the two observed proportions in the numerator and the common null estimate in the standard error.

Try the sequence on a new claim

A business asks whether comparable jobs pay less than \$180,000 on average. Use $\alpha = .05$. An independent normal sample of 100 jobs gives $\bar{x} = \$174{,}600$, with known $\sigma = \$30{,}000$.

Pause and choose the tail

What are the hypotheses? What number measures the gap in standard errors?

Check the decision

$H_0 : \mu \geq 180{,}000$, $H_1 : \mu < 180{,}000$. The standard error is \$3,000. $z = -1.8$ and the left-tail area is 0.0359. Reject H_0 at 5%.

Next: how would the analysis change if the \$30,000 were a sample SD?