Inferential Statistics

↓

'Infer' ⟶ Conclude

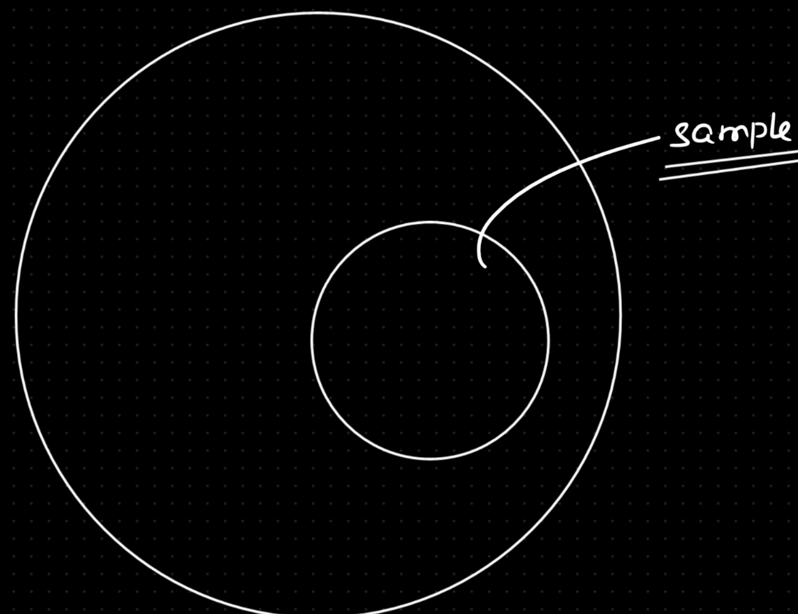
Accept / Reject

(Assumptions

Null — $H_0 \longrightarrow \mu_{sample} = \mu_{population}$

Alternate — $H_1 \longrightarrow \mu_{sample} \neq \mu_{population}$

Population



A company packages flour in bags labeled as containing 1 kg. The quality control team wants to ensure that the average weight of the flour in the bags is indeed 1 kg, as claimed on the label.

$H_0 \longrightarrow \mu_{sample} = 1\ kg$

$H_1 \longrightarrow \mu_{sample} \neq 1\ kg$

Reject / fail to Reject

A factory produces widgets, and the machines are set to produce these widgets at a speed of 50 units per hour. The factory manager wants to verify that the average production speed is indeed 50 units per hour.

Two tail test

$H_0 \longrightarrow \mu_{sample} = 50$ units per hour

$H_1 \longrightarrow \mu_{sample} \neq 50$ units per hour

A company claims that a new marketing campaign will increase customer sign-ups by 20%. The baseline sign-up rate before the campaign was 5%. After running the campaign, the marketing team wants to test whether the average sign-up rate has indeed increased to 25% as claimed.

$H_0 \longrightarrow \mu_{\text{sign-up after Campaign}} = 5\%$

$H_1 \longrightarrow \mu_{\text{sign-up after Campaign}} > 5\%$

one-tail test

Normal Distribution

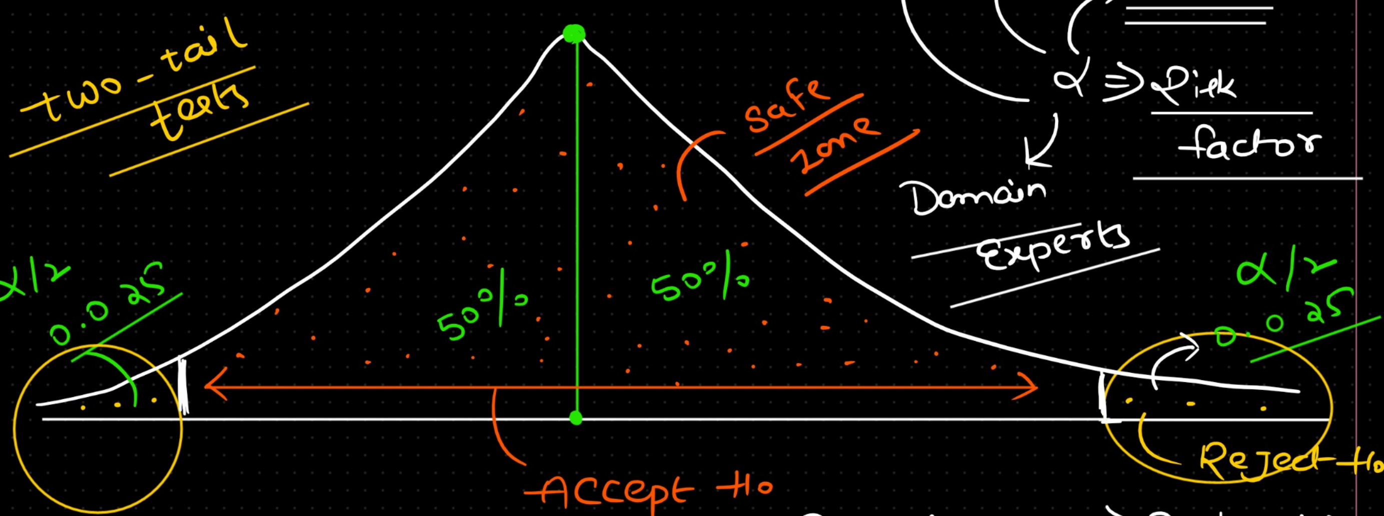
Error

Significance Level

0.05

$\alpha \Rightarrow$ Risk factor

Domain Experts



P-value $\longrightarrow$ Probability of accepting H_0

P-value $> \alpha$

P-value $< \alpha$

Experiments

widely asked interview question?

Decision	Reality	
	H_0 is true	H_0 is false
Reject H_0	Type I error (α)	Correct
Fail to Reject H_0	Correct	Type II error (β)

fact

Reality — $H_0 \rightarrow$ true — Patient is diabetic

Reject H_0 — Patient is not diabetic

$\rightarrow$ Type I Error

Reality — $H_0 \rightarrow$ false

fail to Reject H_0 — Type 2 Error

$\alpha = 0.05 \longrightarrow$ we are willing to take 5% risk of Rejecting H_0 when it is true in reality.

Note: Type 1 error is more dangerous, we fix α & try to minimize β.

A factory produces light bulbs that are supposed to last an average of 800 hours. You take a sample of 50 light bulbs and find that the average lifespan is 790 hours. The standard error (SE) for this sample is 7.5 hours.
Population mean (μ): 800 hours
Sample mean ($\bar{x}$): 790 hours
Standard Error (SE): 7.5 hours
Significance level (α): 0.05 (which corresponds to 1.96 SD)

$H_0 \rightarrow \mu_{sample} = 800$ hours

$H_1 \rightarrow \mu_{sample} \neq 800$ hours

$$z = \frac{\bar{x} - \mu}{SE}$$

$$z = \frac{790 - 800}{7.5}$$

$$z \approx -1.33$$

Normal Distribution

↳ Statistical Test (Experiments)

$$z\text{-test} = \frac{\bar{x} - \mu}{SE}$$

P-value = 0.1835

$\alpha = 0.05$

P-value > α

↳ ~~accept Ho~~

fail to Reject Ho

Average lifespan of light bulb is ≈ 800 hrs only

A nutrition company claims that their protein bars contain 20 grams of protein on average. You take a sample of 40 bars and find that the average protein content is 18.8 grams. The standard error (SE) for this sample is 0.8 grams.

$$\begin{cases} H_0 \longrightarrow \mu = 20\,gm \\ H_1 \longrightarrow \mu \neq 20\,gm \end{cases}$$

$\alpha = 0.05$