Hypothesis Testing

Hypothesis Testing

By codeswithpankaj

What is Hypothesis Testing?

Hypothesis Testing is a **statistical method** used to make **decisions** or draw **conclusions** about a **population** based on **sample data.**

Think of it like a court trial:

You have a **claim** (hypothesis).

→

You collect **evidence** (data).

→

You decide whether to **reject** the claim or **not reject** it.

Real-life Example:

A medicine company claims their new drug **reduces headache pain** in less than **30 minutes** on average.

→

You test this claim using **data** from **patients**.

Pankaj Chouhan | www.codeswithpankaj.com

Null Hypothesis (H_0) vs Alternative Hypothesis (H_1)

Null Hypothesis (H_0)

The "status quo" or "no effect" assumption.
It is what we try to **disprove**.

Usually contains $=$, $\geq$, or $\leq$.

Alternative Hypothesis (H_1)

What we actually **believe** or want to **prove**.

Contains $\neq$, $>$, or $<$.

Example:

H_0: Average time to reduce headache $= 30$ minutes
(no improvement)

H_1: Average time to reduce headache < 30 minutes
(drug is better)

Key Point: We never "prove" H_1.
We only say "**Reject H_0**" or "**Fail to reject H_0**".

Significance Level (α) - The Risk Threshold

α (alpha) is the probability of making a **Type I error** (explained below).

- Common values: **0.05 (5%)** or **0.01 (1%)**.

If p-value $< \alpha$

→ **Reject H_0**

(statistically significant result)

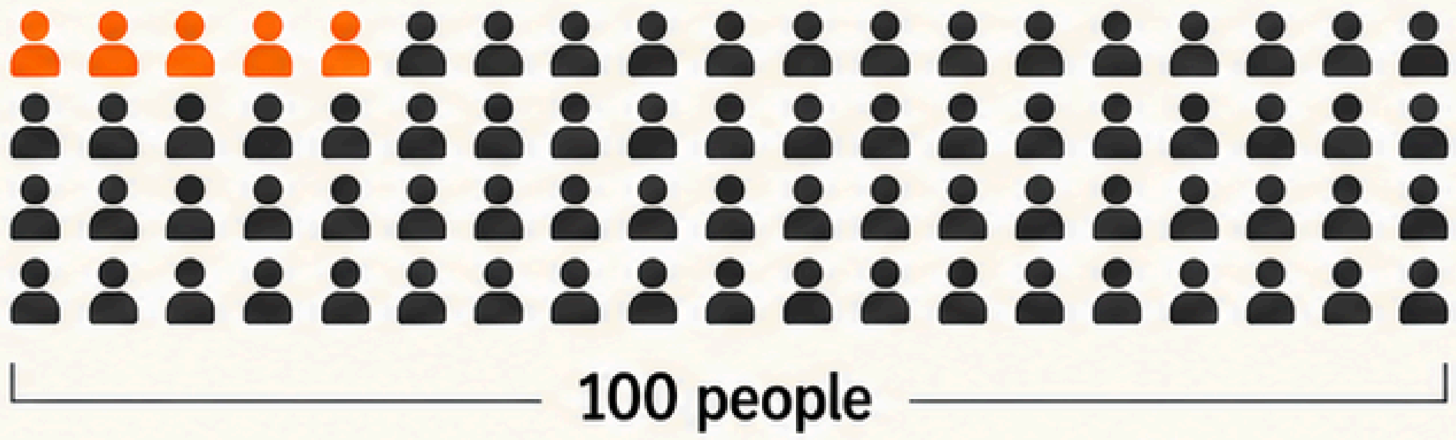
OR

If p-value $\geq \alpha$

→ **Fail to reject H_0**

Simple Analogy:

α = 5% means you are willing to be wrong **5 times out of 100** when you reject H_0.



Pankaj Chouhan | www.codeswithpankaj.com

Type I and Type II Errors

Error Type	Meaning	Real-life Example (Court)	Probability
Type I	Reject H_0 when H_0 is actually True	Sending innocent person to jail	α
Type II	Fail to reject H_0 when H_0 is actually False	Letting guilty person go free	β (beta)

Power of Test $= 1 - \beta$ (probability of correctly rejecting false H_0)

Mnemonic:

Type I = False Positive (you said there is an effect, but there isn't)

Type II = False Negative (you said there isn't an effect, but there is)

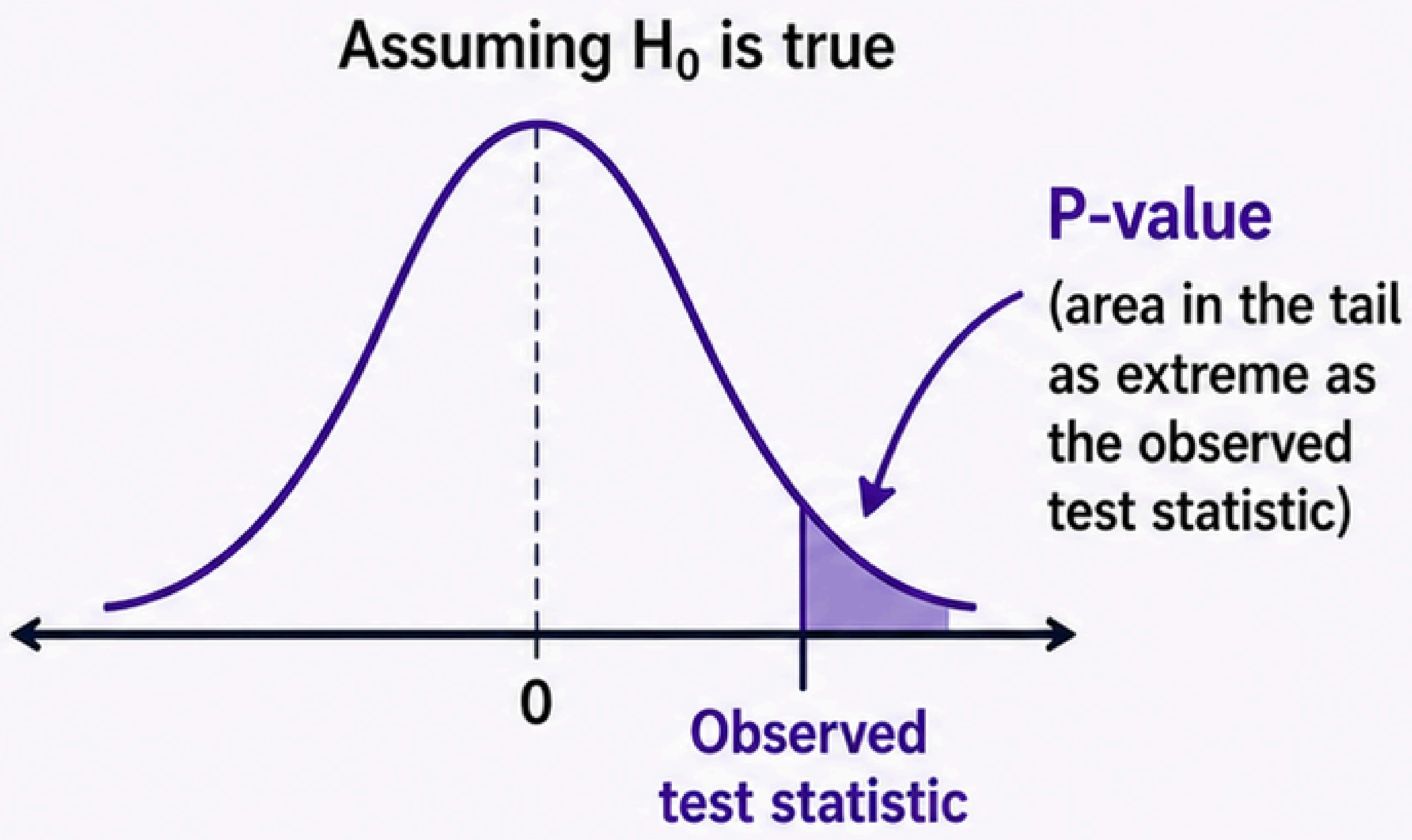
Test Statistic and P-value

Test Statistic:

A number calculated from sample data (e.g., t-score, z-score, chi-square).

P-value:

Probability of getting a test statistic as extreme as (or more extreme than) the one observed, assuming H_0 is true.



Rule of Thumb:

- Very small p-value (e.g., < 0.05) → Strong evidence against H_0.
- Large p-value → Data is consistent with H_0.

Statistical Significance vs Practical Significance

Statistical Significance

Result is unlikely due to chance (based on **p-value**).

Practical Significance

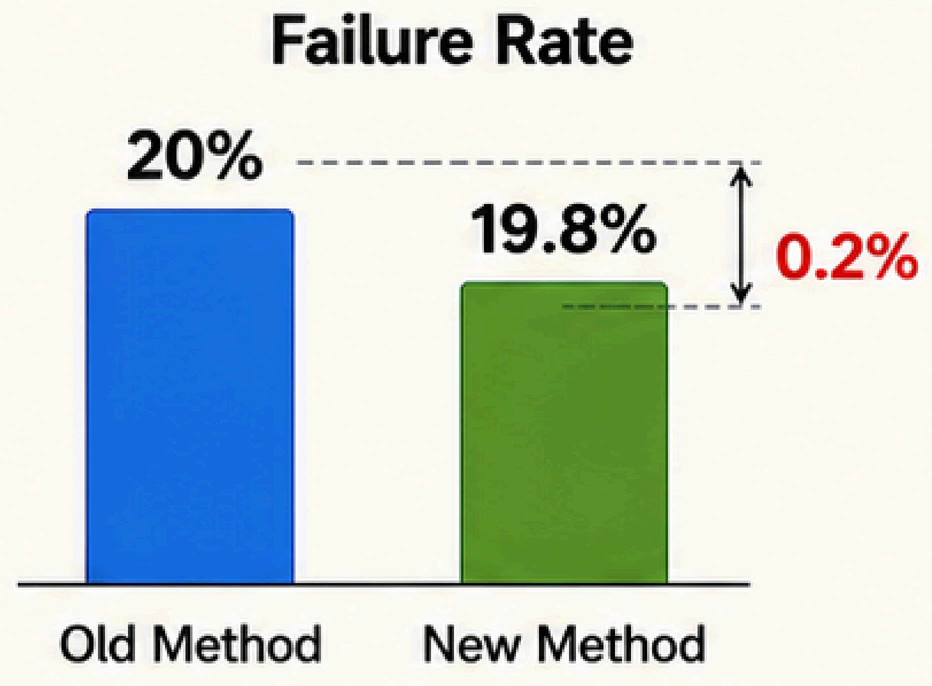
Result is meaningful in real life (**effect size** matters).

Classic Example

A new teaching method reduces exam failure rate from **20%** to **19.8%**.

With huge sample, **p-value < 0.01** → **Statistically significant.**

But saving only **0.2%** students? → **Not practically significant.**



Always check effect size (Cohen's d, correlation, etc.) along with p-value.

- **p-value** tells you if result is real (not by chance).
- **Effect size** tells you how big the effect is (in practical terms).
- Both together tell you what **truly matters.**

One Sample T-Test

Used to test whether the mean of a single sample is significantly different from a known population mean.

Problem:

A school claims average marks of students in a new batch is 75.
You take a sample of **30** students and get **mean = 72**.

Is the claim true?

Hypotheses (Two-tailed Test)

$H_0: \mu = 75$ (The average marks are 75)

$H_1: \mu \neq 75$ (The average marks are not 75)

Test Results (from Python)

Test Statistic (t)	-2.0433
P-value	0.0511

Sample size (n) = 30, Sample mean ($\bar{x}$) = 72.00, Population mean (μ_0) = 75

Decision (α = 0.05)

P-value (0.0511) > α (0.05) → Fail to reject H_0

Interpretation:

There is no statistically significant evidence to say that the average marks are different from 75.

Test Statistic

The test statistic for one sample t-test is:

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

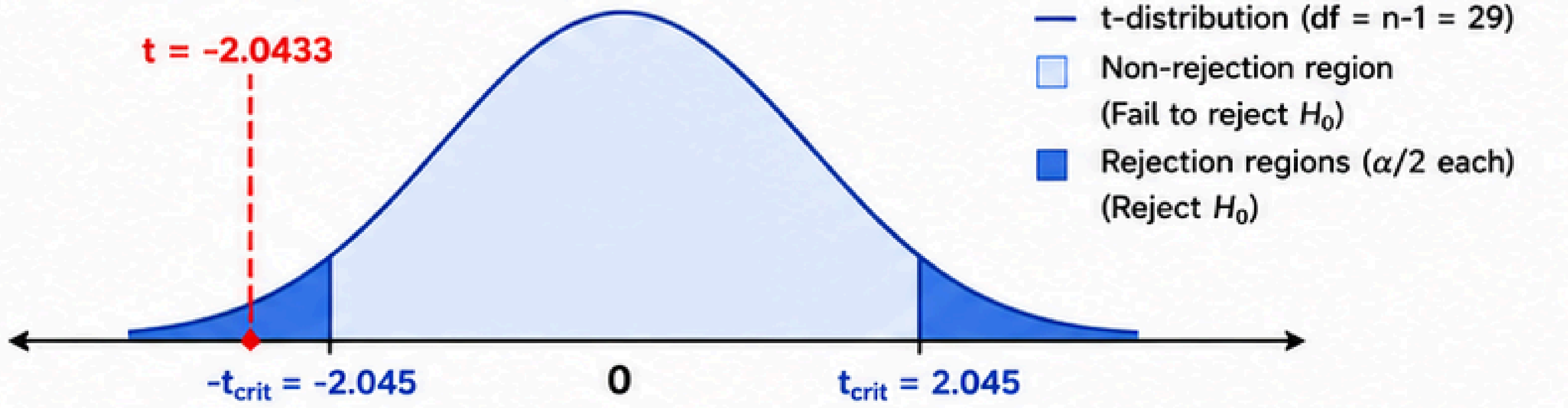
$\bar{x}$ = sample mean

μ_0 = population mean (claimed mean)

s = sample standard deviation

n = sample size

Visualization (Two-tailed Test)



Python Code

```
import numpy as np
from scipy import stats

# Sample data (marks of 30 students)
np.random.seed(42)
sample_marks = np.random.normal(loc=72,
                scale=8, size=30)

# Perform One Sample T-test
t_stat, p_value = stats.ttest_1samp(sample_marks,
                                    popmean=75)
```

```
# Hypotheses
# H0: mu = 75
# H1: mu != 75

alpha = 0.05

if p_value < alpha:
    print("Result: Reject H0 → Average marks are
          significantly different from 75")
else:
    print("Result: Fail to reject H0 → No
          significant difference from 75")
```

Key Takeaway: Since p-value (0.0511) > 0.05, we **fail to reject H_0**.
The data does **not** provide enough evidence that the average marks are different from 75.