# Hypothesis Testing

A statistical hypothesis test is a method of Statistical Inference used to decide whether the data at hand sufficiently support a particular hypothesis. Hypothesis testing allow us to make probabilistic statements about pop parameters.

## Null and Alternate Hypothesis

1. **Null hypothesis ($H_0$):** In simple terms, the null hypothesis is a statement that assumes there is no significant effect or relationship btn the variable being studied. It serves as the starting point for hypothesis testing and represent the status quo or the assumption of no effect until proven otherwise. The purpose of hypothesis testing is to gather evidence (data) to either reject or fail to reject the null hypothesis in favour of the alternative hypothesis, which claims there is a significant effect or relationship.

eg → Lays $\xrightarrow{\text{weight}}$ 100gm

↳ our assumption is Lays $\xrightarrow{\text{weight}}$ May be more than 100gm. or less than.

* but Null hypothesis Statement is ⇒ Lays weight ⇒ 100 gm

2. Alternative hypothesis ($H_1$ or $H_a$): The alternative hypothesis, is a statement that contradicts the null hypothesis and claims there is a significant effect or relationship both the variables being studied. It represents the research hypothesis or the claim that the researcher want to support through statistical analysis.

Alternative hypothesis ⟶ opposite to Null hypo

ex:— In Null hypothesis

lays weight ≥ 100 gm

but In Alter hypo

lays weight ≠ 100gm.

$H_0 : \mu = 100\ gm$ → Hypothesis style

$H_a : \mu \neq 100\ gm$

Important Points

- How to decide what will be Null hypothesis and what will be Alternative Hypothesis (Typically the Null hypothesis says nothing new happen)
- We try to gather evidence to reject the null hypothesis
- It's important to note that failing to reject the null hypothesis doesn't necessarily mean that the null hypothesis is true; it just means that there is not enough evidence to support the alternative hypothesis.

Ex:- ~~Lawyer~~ Hypothesis test are similar to jury trials.

$H_0$ → Criminal didn't do any crime or murder

$H_1$ → Criminal did murder

lawyer have to gather some evidence to proove that $H_0$ (Criminal didn't do murder) is wrong.

But lawyer haven't gather proper evidence. So, judge announce that $H_0$ (Criminal didn't do murder) is true.

↳ But this is not properly true bcz lawyer had not gather evidence.

## Steps involved in Hypothesis Testing

### Rejection Region Approach

1. Formulate a Null and Alternative hypothesis.

ex:- yt video shooting technique

Old shooting technique → 6 min avg watch time

New shooting technique → 13 min " " "

$H_0$: $\mu = 6$ min (old technique avg watch time = new tech avg watch time)

$H_a$: $\mu > 6$ min (new tech is greater watch time than old tech)

2. Select a significance level (This is the prob of rejecting the null hypothesis when it is actually true, usually set at 0.05 or 0.01)

3. Check assumption (example distbn)

→ normally distbn

→ If pop $\sigma$ given ⟶ z-test

If not ⟶ t-test.

4. Decide which test is appropriate (Z-test, T-test, chi-square test, Anova)

Z-test → pop σ

t-test → not pop σ

Chi-Squares test → for categorical

5. State the relevant test statistics

t-test → t score

Z-test → Z-score

6. Conduct the test

7. Reject or not reject the Null hypothesis.

8. Interpret the result.

## Performing a Z test

Example 1

Suppose a company is evaluating the impact of a new training program on the productivity of its employees. The company has data on the avg productivity of its employees before implementing the training program. The average productivity was 50 units per day with a ~~company~~ known pop standard deviation of 5 units. After implementing the training program, the company measures the productivity of a random sample of 30 employees. The sample has an average productivity of 53 units per day

$\mu = 50 \quad \sigma = 5$

$n = 30 \quad \bar{x} = 53$

Steps →

1) $H_0: \mu = 50 \qquad H_a: \mu > 50$

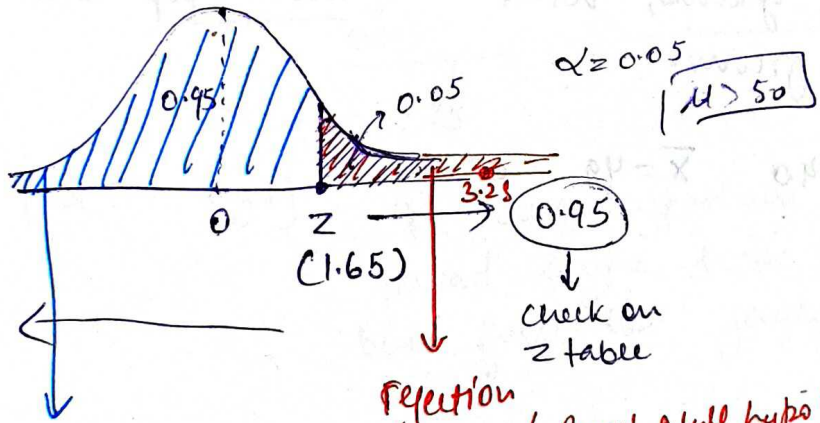
2) $\alpha = 0.05 \rightarrow 5\%$.

3) Normality Valid / pop std($\sigma$) known → Z-test

4) Z test

5) Z score

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{53 - 50}{5/\sqrt{30}} = 3.28$$




* Reject Null Hypothesis

## Example 2

Suppose a snack food company claims that their Lays wafer packets contain an average weight of 50 grams per packet. To verify this claim, a consumer watchdog organization decides to test a random sample of Lays wafer packets. The organization wants to determine whether the actual average weight differs significantly from the claimed 50 grams. The organization collects a random sample of 40 Lays wafer packets and measures their weights. They find that the sample has an average weight of 49 grams, with a known pop standard deviation of 4 grams.

$\mu = 50$ , $n = 40$ $\bar{x} = 49$ $\sigma = 4$

1) $H_0: \mu = 50$

$H_a: \mu \neq 50$

2) $\alpha = 0.05$

3) Normality — $\sigma$ ✓ → Z test

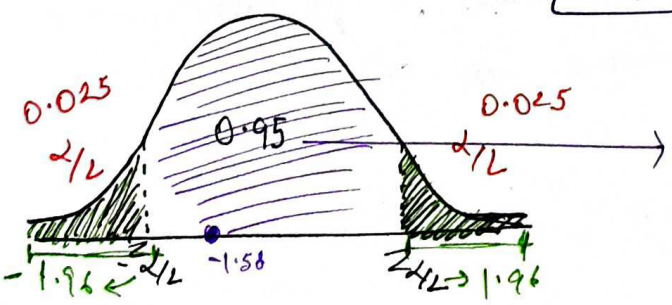
4) Z test

5) Z - score

6) $z = \frac{49 - 50}{4/\sqrt{40}}$ $z = \frac{-\sqrt{40}}{4}$ $z = \boxed{-1.58}$

$\alpha = 5\%$

$\boxed{\mu > 50}$ $\mu \neq 50$



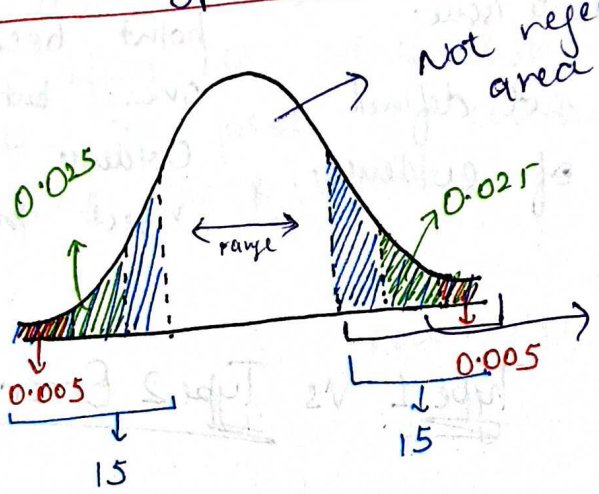
# Rejection Region

**Significance level** :- denoted as $\alpha$ (alpha), is a predetermined threshold used in hypothesis testing to determine whether the null hypothesis should be rejected or not. It represents the probability of rejecting the null hypothesis when it is actually true, also known as Type 1 error.

$\alpha = 5\%$

$\alpha = 1\%$

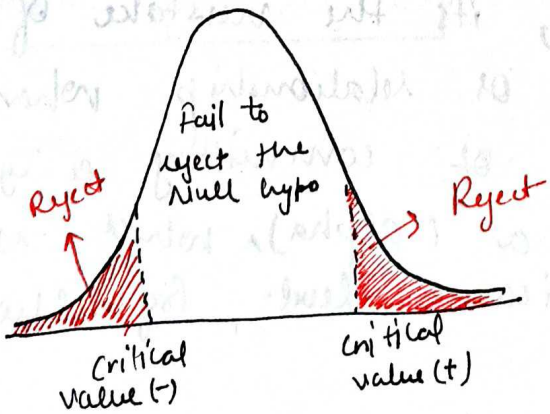
$\alpha = 30\%$



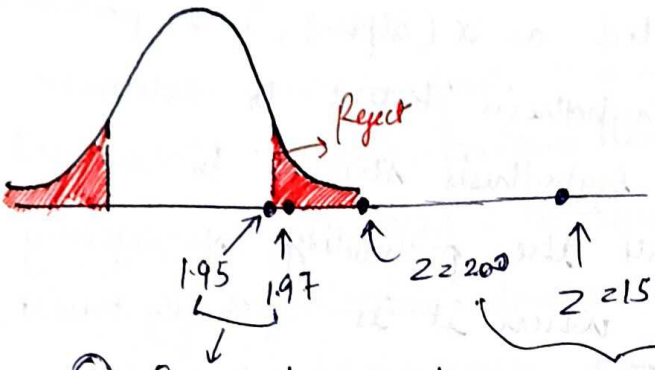

[$\alpha \uparrow$ - range ↓]

* Null hypothesis true hoga par jayad hypo hone se True hypo bhi false ho jayega.

The rejection region is the region of values that corresponds to the rejection of the null hypothesis at some chosen prob level.



## Problem with ~~Z score test~~ Rejection Region Approach




① Only 0.01 se reject this create issue.

② we can not defined Streyth of evidence.

There is no difference beth both point because point in reject area but Z=15 has more evidence than Z=2.00 to reject point.

## Type 1 Vs Type 2 Error

Type-1 (false Positive) error occurs when the sample results lead to the rejection of the null hypothesis when it is in fact true.

- $H_0$ → no crime
- ~~Reject~~ Reject null hypothesis ~~Reject~~ (But in Real man not did crime)

In other words, it's the mistake of finding a significant effect or relationship when there is none. The prob of committing a Type I error is denoted by α (alpha), which is also known as the significance level. By Choosing a

significance level, researchers can control the risk of making Type I error.

Type II (False Negative) error occurs when based on the sample results, the null hypo is not rejected when it is in fact false.

This means that the researcher fails to detect a significant effect or relationship when one actually exists. The prob of committing a Type II error is denoted by β (beta).

(β)

Power → (1 − β)

[Truth about the population]

| | $H_0$ true | $H_0$ false |
|---|---|---|
| Reject $H_0$ | Type I error | correct decision |
| Accept $H_0$ | correct decision | Type II error |

$\alpha = 0.05$

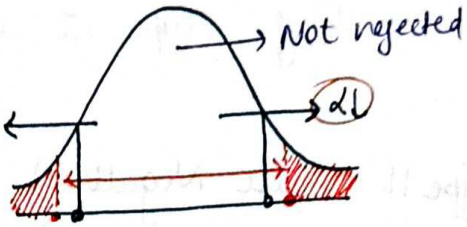
↓

good Value because of trade off

0.05 is stable Value

## Trade-off between Type 1 and Type 2 errors

$\alpha \downarrow$ Type II $\uparrow$



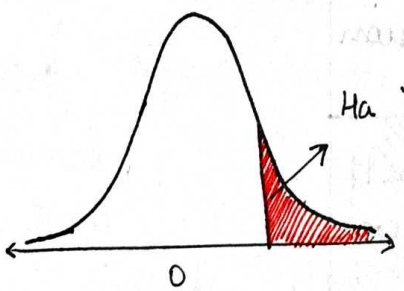

Type I error $\downarrow\downarrow$ – Type II error $\uparrow\uparrow$

Type II error $\downarrow\downarrow$ – Type I error $\uparrow\uparrow$

## One Sided vs Two Sided test

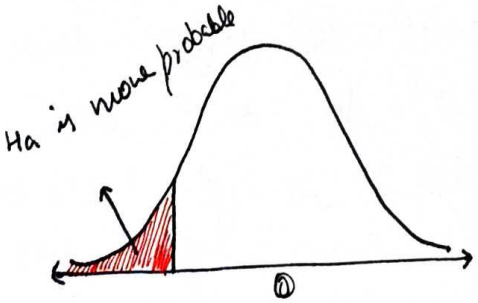
**One sided (one-tailed) test:** A one-sided test is used when the researcher is interested in testing the effect in a specific direction (either greater than or less than the value specified in the null hypothesis). The alternative hypothesis in a one-sided test contains an inequality (either "$>$" or "$<$")

Example: A researcher wants to test whether a new medication increase the average recovery rate compared to the existing medication.




Right-tail test

$H_a: \mu > \text{value}$

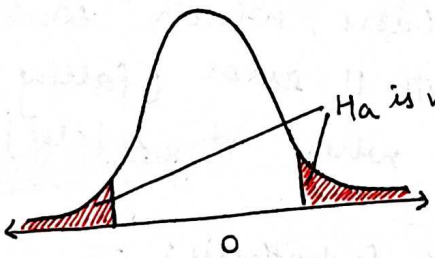



Left-tail test

$H_a: \mu < \text{value}$

**Two sided (two tailed) test:** A two-sided test is used when the researcher is interested in testing the effect in both direction (i.e, whether the value specified in the null hypo is different, either greater or lesser). The alternative hypothesis in a two-sided test contains a 'not equal to' sign ($\neq$).

Example: A researcher wants to test whether a new medication has a different average recovery rate compared to the existing medication.




Two tail test

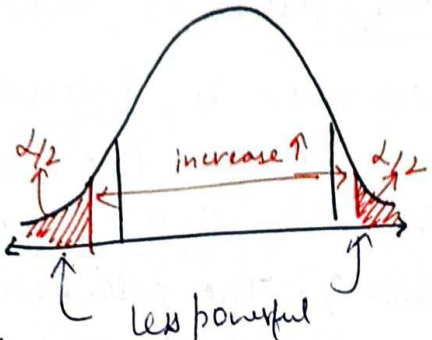
Ha: $\mu \neq$ value

## Advantages and Disadvantages

Two-tailed test (two-sided):

Advantages:

1. **Detects effects in both directions:** Two-tailed tests can detect effects in both directions, which makes them suitable for situations where the direction of the effect is uncertain or when researchers want to test for any differences betn the group of variab
2. **More conservative:** Two-tailed tests are more conservative because the significance level ($\alpha$) is split betn both tails of the distbn. This reduces the risk of Type I error in cases where the direction of the effect is uncertain




$\alpha = 0.05$ ↓ split into two sides (left or right) (and increase the middle size)

Disadvantages:

1. Less Powerful: Two-tailed tests are generally less powerful than one-tailed test because the significance level ($\alpha$) is divided beth both tails of the distbn. This means the test requires a larger effect size to reject the null hypothesis, which could lead to a higher risk of Type II error (failing to reject the null hypothesis when it is false).

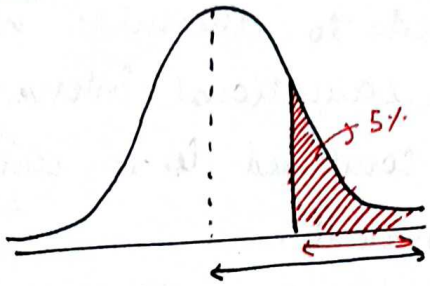
2. Not appropriate for direction hypothesis: Two-tailed tests are not ideal for cases where the research question or hypothesis is directional, as they test for differences in both directions, which may not be of interest or relevance.

says $\mu \neq 100$ but dont know $\mu > 100$?
$\mu < 100$?

# One tailed test (one-sided):

## Advantages

1. **More Powerful:**

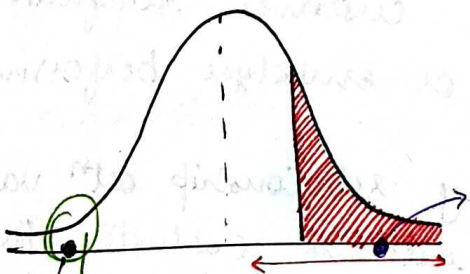



More Powerful to reject Null hypothesis than two tailed test.

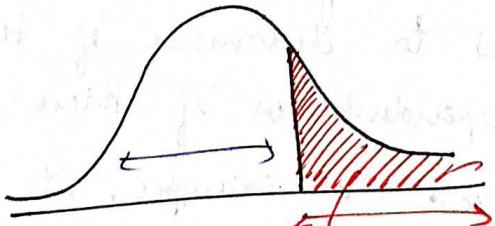
2. **Directional hypothesis:** One tailed test are appropriate when there is strong theoritical or practical reason to test for an effect in a specific direction.

## Disadvantages:

1. **Missed effects:**




2. **Increased risk of Type I error:**



Maximum chance to fall in Red area

→ max or large area of Reject Null hypo

# When can be Hypothesis Testing Applied?

1. Testing the effectiveness of interventions or treatments: Hypo testing can be used to determine whether a new drug, therapy, or educational intervention has a significante effect compared to a control group or an existing treatment.

2. Comparing means or proportions: Hypo testing can be used to compare means or proportion betn two or more groups to determine if there's a significanct diff. This can be applied to compare average customer satisfaction scores, conversion rates, or employee performance ~~after~~ across diff grps

3. Analysing relationship betn variables: Hypothesis testing can be used to evaluate the association betn variably, such as the corelation betn age and income or the relationship betn advertising spend and sales.

4. Evaluating the goodness of fit: Hypo testing can help asses if a particular theoritical distbn (eg. normal, bionomial or poisson) is a good fit for the observed data.

5. Testing the independence of categorical variables: Hypo testing can be used to determine if two categorical variables are independent or if there's a significant association betn them. For example, it can be used to test if there's relationship betn the type of product and the likelihood of it being returned by a customer.

6. A/B testing: In marketing product development and website design, hypo testing is often used to compare the performance of two diff version (A and B) to determine which one is more effective in term of conversion rates, user engagement or other metrics.

## Hypothesis Testing ML Application

1. Model comparison
2. Feature selection
3. Hyperparameter tuning
4. Assessing model assumption.