DAV-3

HYPOTHESIS TESTING

(Class starts @ 9:10 PM)

Lecture 7: Advanced Hypothesis Testing

Class starts at 9:05 PM

Agenda

1. Recap of One way ANOVA + Levene's Test + SW Test
2. Two-way ANOVA ✓
3. KS Test + QQ plot ✓
4. A/B Testing ✓
5. Parametric Vs Non-parametric }

Recap of 1 way ANOVA

Num Var Vs Cat Var → >2 groups

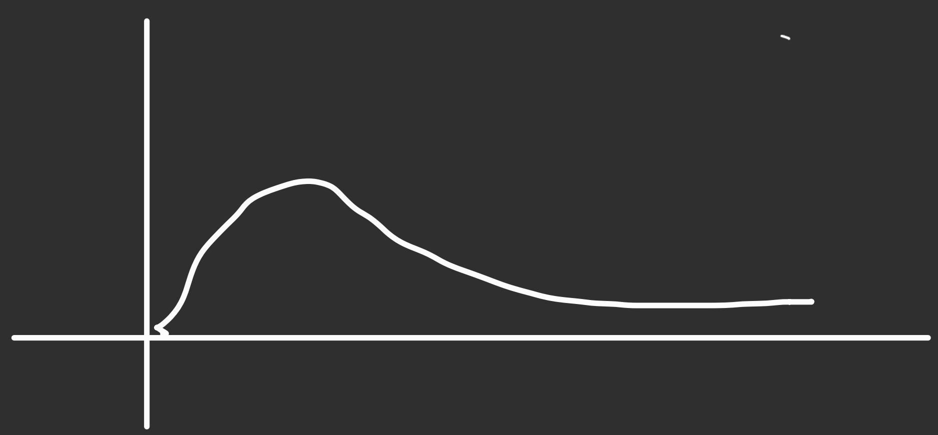
ANOVA
one way

Test statistic ⇒

$$F\text{ratio} \Rightarrow \frac{\text{Var b/w group}}{\text{Var with group}}$$

F ratio ⇒ Two Chi sq Statictic put together

→ Two dof



Q) In ANOVA what is the measuring statistic we follow?

H_0: Mean of all group are same

H_a: Mean of any one of the group is dif

or

Atleast one of the group has different mean

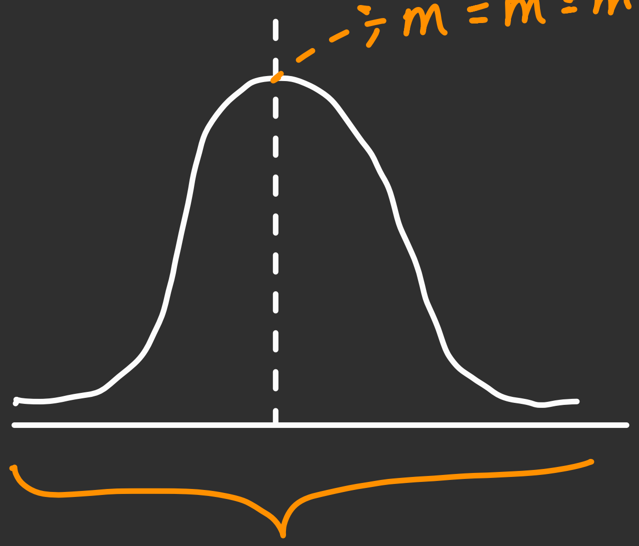
Assumption of ANOVA

1. Normal distribution (QQ plot / Shapiro Wilk Test)
   - QQ plot → Viz
   - Shapiro Wilk Test → HT
2. Equal variance within group (Levene's Test)
3. Sample should be randomly drawn

ANOVA

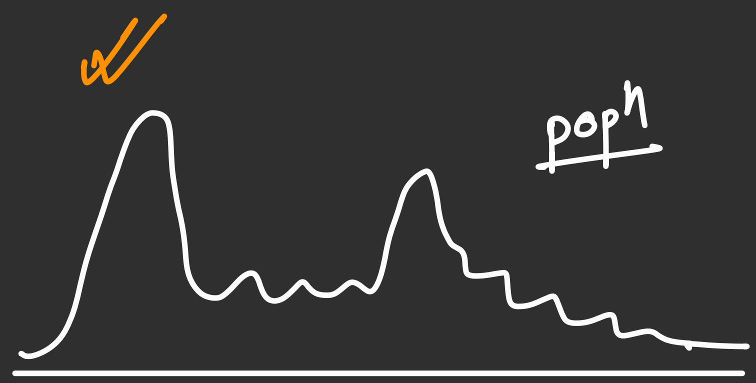
One way ANOVA → last class

Two way ANOVA → today



"log-normal"

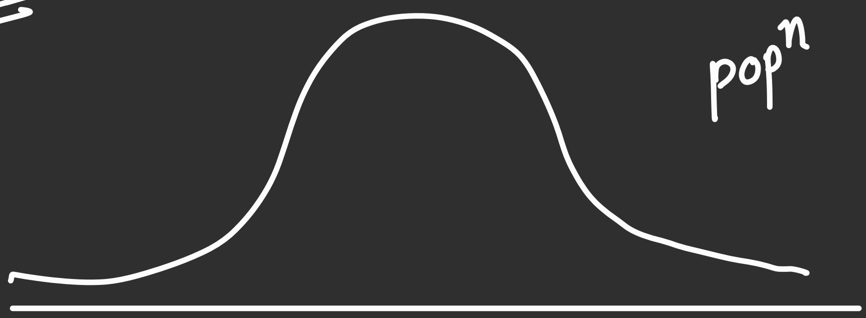
pop^n mean



$\Downarrow$

transformation

CLT $\longrightarrow$ pop^n mean



Shapiro-wilk Test → Parametric Test

(Normality Test ⇒ Hypothesis Test) Po

H_0: My dist is normal

H_a: My dist is not Normal

QQ plot

Note: It doesn't work well for sample size $\boxed{> 200}$ & $\boxed{< 50}$

↳ sample > 50 & < 200

Parametric Test → They always assume normality of data

Non-Parametric → They do not assume normality assumption

Levene's Test

Levene's Test (Aka Test of Variability)

Assumption of ANOVA : "There should be equal variance in diff groups/ categories in our data".

$$\text{f ratio} = \frac{(\text{var b/w group})}{\text{var within grop}}$$

"Hypothesis Test" ⟶ Assump

H_0: Equal var

H_a: Not Equal var

Q) Can I apply Levene's Test for only 2 groups?

Yes

Two Way ANOVA

Two way ANOVA

Data scientist @ Beverage company (Sales ↑)

dep ind var

$y = f(x)$

Location, Flavour, Unit, Sales

Annova

Independent Variables

Location, Flavour

num

Dependent Variables

Sales

Factors

cat

Location ⟷ Flavour ⟷ Sales

cat num

- East
- West
- North
- South

- Lemon
- Cola
- Orange

Two Types of Effect

Annova

1. Main Effect : 1. Flavour Vs Sales → H_0 and H_a
   2. Location Vs Sales → H_0 & H_a

Chi

2. Interaction Effect : 1. $\boxed{\text{Flavour Vs Location}}$ (Vice Versa) → H_0 and H_a

Interaction

3 categorical → 3 Main effects

Numeric

3

Q) Is there any significant difference between Sales and Flavour?

Q) Is there any significant difference between Sales and Location?

4

Sales & Gender

Q) is there any interaction between Location and Flavour on sales?

Chi

Location & Gender

Flavour & Gender

pvalue is significant $\longrightarrow$ H_a

(significant interaction)

pvalue $< \alpha$

Reject (H_0) / Accept (H_a)

else

Accept (H_0) / Reject (H_a)

(pvalue $\downarrow$)

One Way ANOVA

Num Vs Cat

1 ind var

>2

Two-Way ANOVA

Num Vs $\underbrace{\text{Cat1 Vs Cat2}}$

2 ind var

>2 >2

p value is significant → Ha

pvalue $< \alpha$:

Reject H_0 / Accept H_a

p value

Kolmogorov Simirnov Test

KS Test → It tells you whether the two distributions are similar

$M_1: x_1, x_2, x_3, x_4 \cdots\cdots x_{100}$

$M_2: x_1, x_2, x_3, x_4 \cdots\cdots x_{90}$

Recovery time of two Medicine

popln sd

Q) Is there any significant difference between recovery time between two medicines?

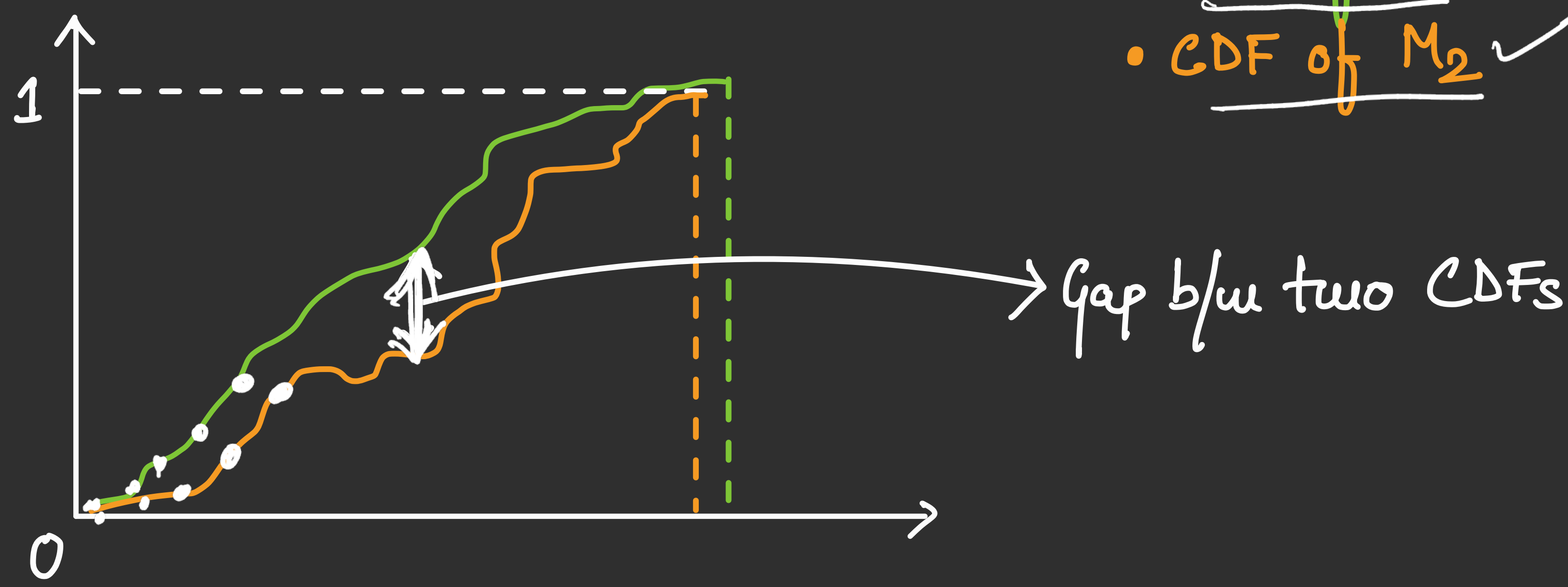
2 sample T test

paired t test X

2 sample T-Test

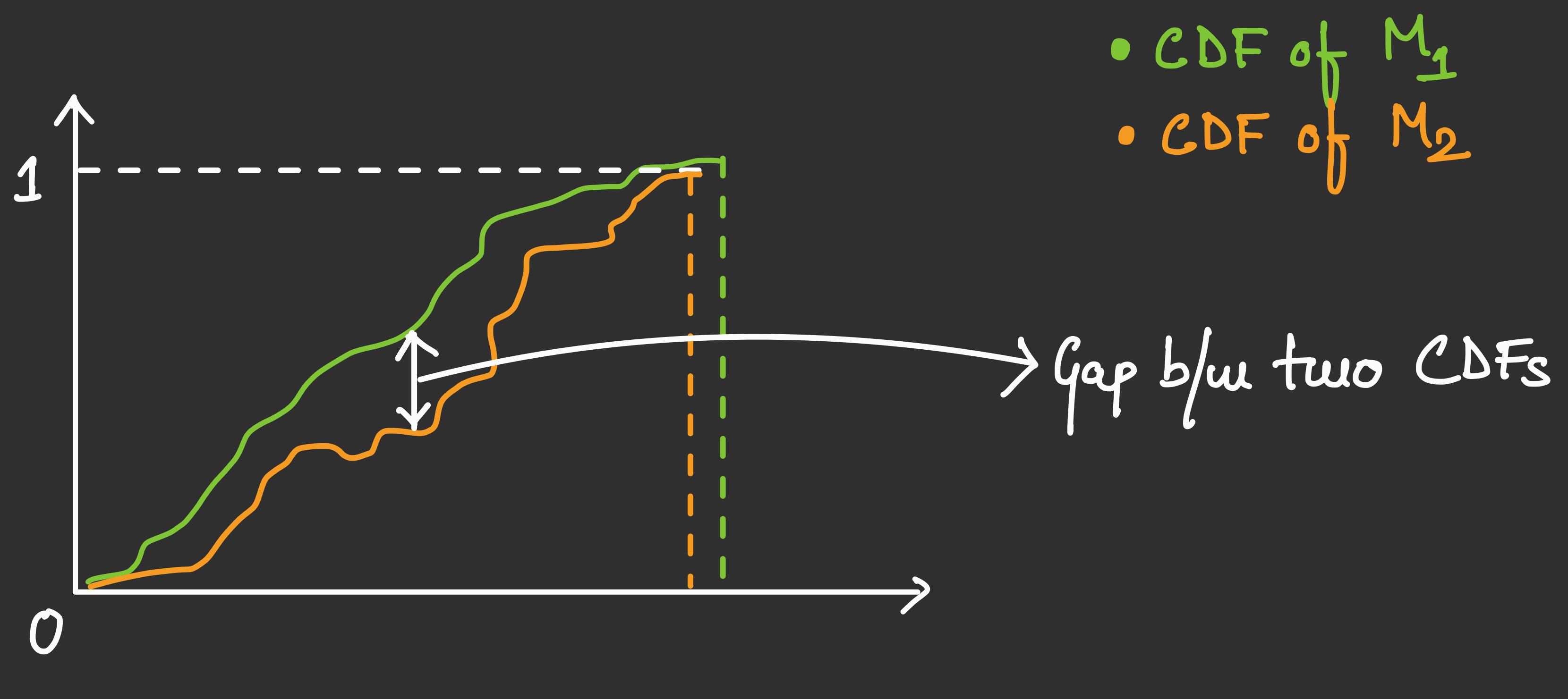
1. Population parameter is known
2. Statistic is mean
3. Data is normal

- CDF of M_1 ✓
- CDF of M_2 ✓



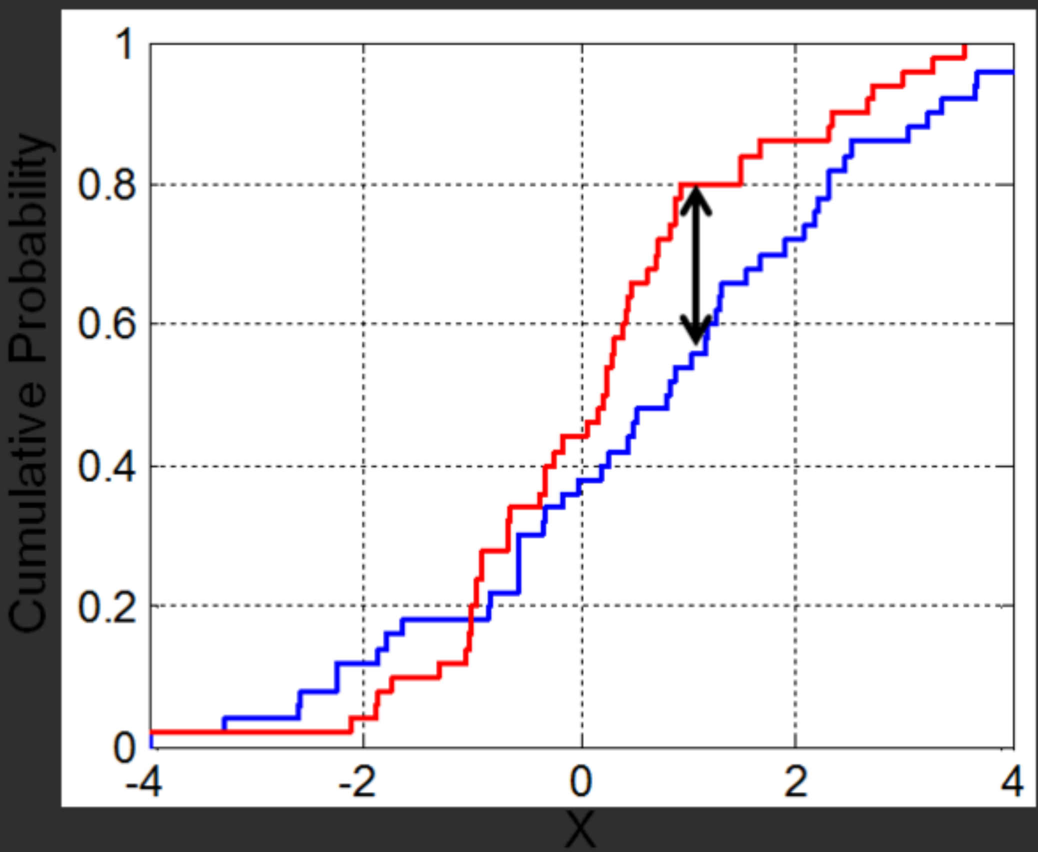
$$KS_{st} \Rightarrow \max(\text{gap b/w Two CDFs})$$

$$\max \left\| \left[CDF_1 - CDF_2 \right] \right\|$$



① Point wise CDF comparision

② It is a distribution free Test (Non parametric Test)



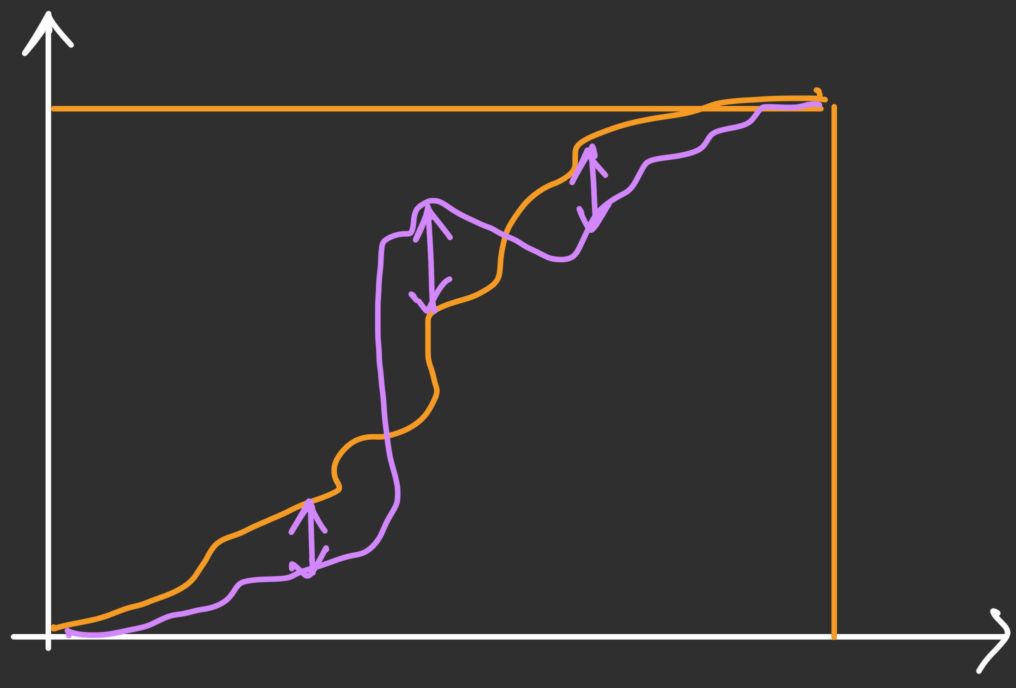
H_0: Both the distribution are same

H_a: Both the distribution are not same

Under $H_0 \Rightarrow K_{st} \longrightarrow 0$

Under $H_a \Rightarrow K_{st} \longrightarrow$ (large +ve/-ve Number)

We use KS Test to compare two distributio



$\underline{\underline{x - y}}$

$\underline{\underline{-ve}}$

Can I use KS Test
as a normality Test?

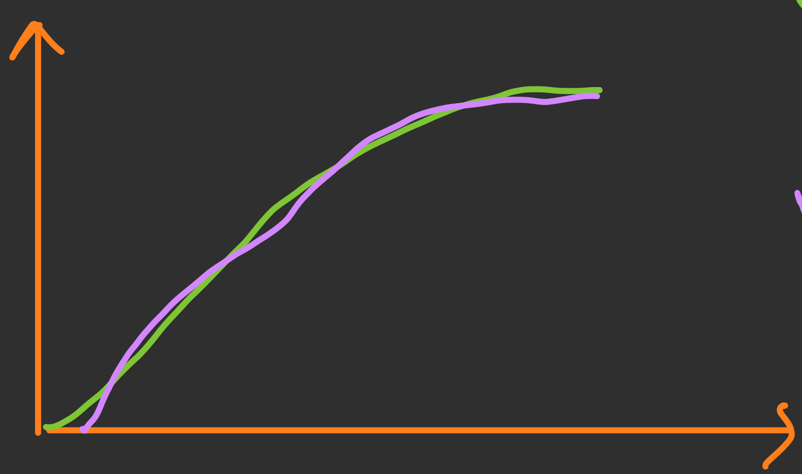
dist 1 Vs dist 2
(Normal)

Difference b/w Q/Q plot & KS Test

Q/Q Plot ⟶ Comparision based on percentile

KS Test ⟶ Comparision based on CDFs

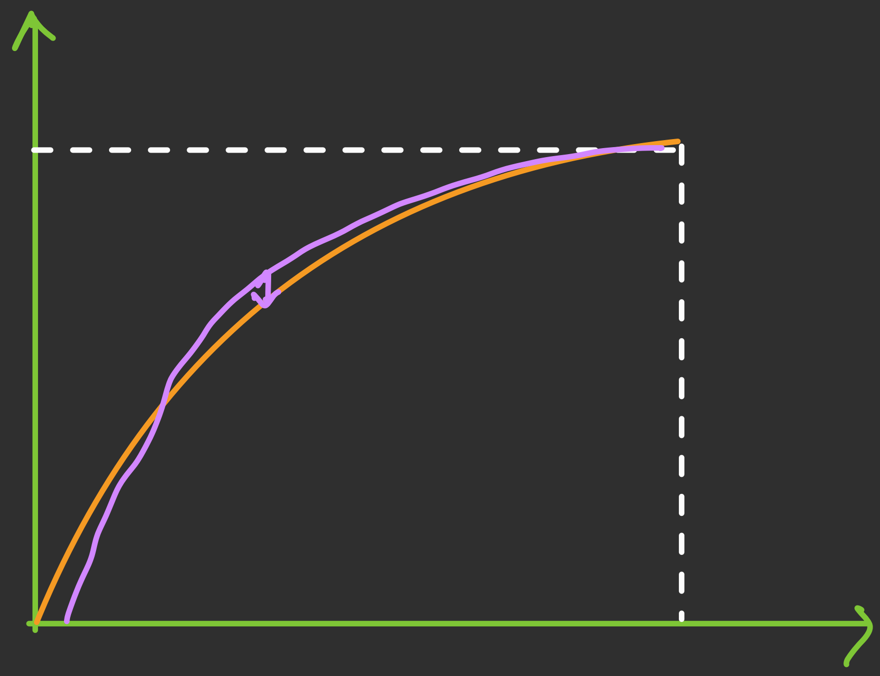
↳ Normality Test



KS Test

"Non parametric Test"

"Biggest flaw"



normal

↳ cdf

A/B Testing

A/B Testing

A/B Testing

It is not a new Test, but a framework

A/B/C Testing

youtube (Ad)

10,000

5000

A 5000 (50%) → Control group → 1 ad

$[x_1, x_2, x_3 \cdots x_n]$

user engagement

B 5000 (50%) → Treatment group → 2 ads

$\underbrace{[y_1, y_2, y_3 \cdots y_n]}_{\text{watch time}}$

5000

long time

"A/B Testing / Experimentation"

2 ads

① Treatment Group : Group with new feature (Group B)

1 ad

② Control Group : Group with No new feature (Group A)

A/B Testing

A/B Test

1. For A/B Test, we use $\boxed{\text{T-test}}$
2. There is no such thing, as A/B/C Test
3. How long are you going to run the Experiment (Time bound)

   (1w, 1m, 3m, 20d, 15d, 10d)

   domain expert
4. Know your evaluation metrics

A/B Testing Summary

1. **Purpose of A/B Testing**

- A/B testing aims to make data-driven decisions by comparing two (or more) versions of a variable to see which version leads to better outcomes, such as higher conversion rates, click-through rates, or user engagement.

2. **Control and Treatment Groups**

- In a basic A/B test, two groups are created:
  - Control Group (A): Exposed to the current or standard version.
  - Treatment Group (B): Exposed to the new or modified version.
- By comparing these two groups, we can isolate the impact of the change made in the treatment group.

3. **Randomization**

- Random assignment of users (or other subjects) to the control and treatment groups is crucial to eliminate biases. Randomization ensures that each group is statistically equivalent, allowing for a fair comparison.

4. **Statistical Significance**

- The results of an A/B test need to be statistically significant to conclude that any observed differences are unlikely due to random chance. Typically, a p-value (e.g., $p < 0.05$) is used to determine significance.

5. **Primary Metric (Key Performance Indicator)**

- The primary metric, or KPI, is the specific metric used to measure the success of the test. This could be conversion rate, average revenue per user, time spent on a page, etc., depending on the test's objective.

6. **Sample Size and Power**

- Choosing the right sample size is essential to ensure that the test has enough statistical power to detect a meaningful difference. Small sample sizes can lead to inconclusive results, while overly large sizes may detect even trivial differences.

7. **Hypothesis Formulation**

- A/B tests start with a hypothesis, such as "Changing the color of the call-to-action button will increase click-through rates." This hypothesis guides the design of the test and the interpretation of results.

8. **Test Duration**

- The test should run long enough to collect sufficient data across a representative sample of users, capturing any cyclical or seasonal effects. Ending a test too early may lead to unreliable conclusions.

9. **Avoiding Multiple Testing Pitfalls**

- Testing too many variations simultaneously (A/B/n testing) or running consecutive tests on overlapping user groups can lead to false positives and complicate analysis. It’s important to account for multiple testing effects.

10. **Interpretation of Results**

- A statistically significant result shows that one version outperformed the other, but it’s also important to consider effect size, user feedback, and practical significance to determine if the change is worth implementing.

Limitations of A/B Testing

- A/B testing may not capture long-term effects (e.g., novelty effects may wear off).
- It is not suited for all types of changes, especially when sample sizes are small or when there are complex interactions between variables.
- For significant business changes, other methods like multivariate testing or sequential testing might be more appropriate.

Parametric

vs

Non-Parametric

Test

Parametric Vs Non-parametric Test

Parametric	Non-Parametric
Check the underlying distribution of data before Test	No need to check the underlying distribution of data before test
T-test, ANOVA	KS Test, KW Test

Parametric HT

Assumptions:

- Parametric tests make specific assumptions about the population distribution from which the data is drawn.
- Common assumptions include normality (data follows a normal distribution) and variance is constant across groups or conditions.
- Parametric tests are typically used when the data reasonably follows the assumed distribution and other assumptions are met.
- Parametric tests tend to be more powerful (i.e., better at detecting true effects) than non-parametric tests when the assumptions are met.
- This is especially true when the sample size is large.

Non-Parametric HT

Assumptions:

- Non-parametric tests make fewer or no assumptions about the population distribution.
- They are distribution-free or rely on fewer assumptions, such as independence of observations.
- Non-parametric tests are useful when the assumptions for parametric tests are violated.
- They are also suitable for data types that don't fit well with parametric assumptions, such as ordinal or skewed data.
- Non-parametric tests are generally less powerful than parametric tests when data conforms to parametric assumptions.
- However, they can be more robust and appropriate when dealing with non-normally distributed data.