18 June 2023 15:39

Recap:
- → Statistics
- → Z-Score
- → Use case of Std
- → Normal distribution / Gaussian dist / Bell Curve
- → Empirical formula 68-95-99.7 Rule

Agenda:
- Hypothesis testing
- Types of variable
- Basics of Machine learning

Theoritical part

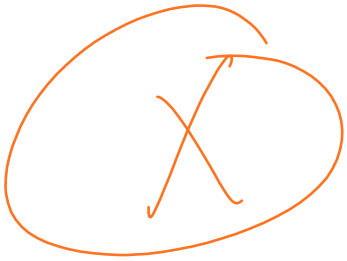
Hypothesis Testing

Here we evaluate two Mutual exclusive statements on a population data, using sample data. then we conclude that this population belong to a particular statement or not.

⇒ Roll a dice

✓ E_1 = even number [2, 4, 6]

✓ E_2 = divisible by 3 [3, 6]



↑ Hypothesis → Assumption

* Headache

Medicine (A) → 3 Hrs to Relief

Medicine (B) → 2 Hrs to Relief

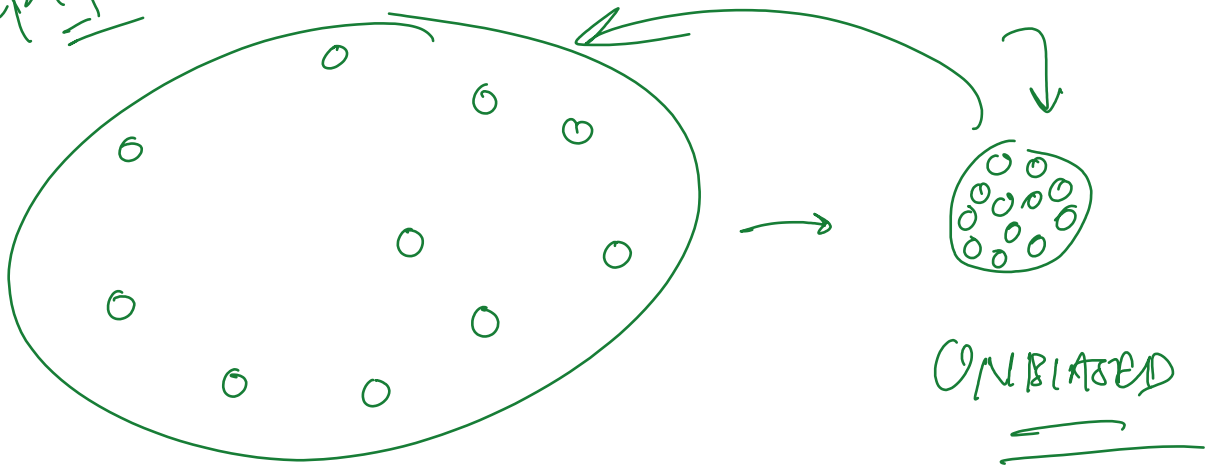
⇒ Trials → Sample data (500)

→ Sample data (1000)

→ Sample data (100 k) = failed

① Null Hypothesis : Medicine A is best — Rejected

② Alternate Hypothesis :- Medicine B is best — Accepted



⇒ Steps

① Make Initial Assumption (Ho Hypothesis (NULL)

② Gather Evidences

③ Rejected or Accept the Null Hypothesis, Based on the Testing Results.

18 June 2023 16:10

When we are testing Null Hypothesis (H_0) against alternate hypothesis (H_a), there are 4 Cases

(i) H_0 accepted when H_0 is true [Correct]

(ii) H_0 rejected when H_0 is true [Error Type I]

(iii) H_0 accepted when H_0 is false [Error Type II]

(iv) H_0 rejected when H_0 is false [Correct]

* Type II error is More dangerous

P- Value (Probability value) :-

If P value for any Assumption (H_1) is more than 5% then we go and accept the Null Hypothesis.

Range of probability [0-1]

if P value is >= 0.05 (5%)

then we accept our Claim (H_0).

Variables in Statistics

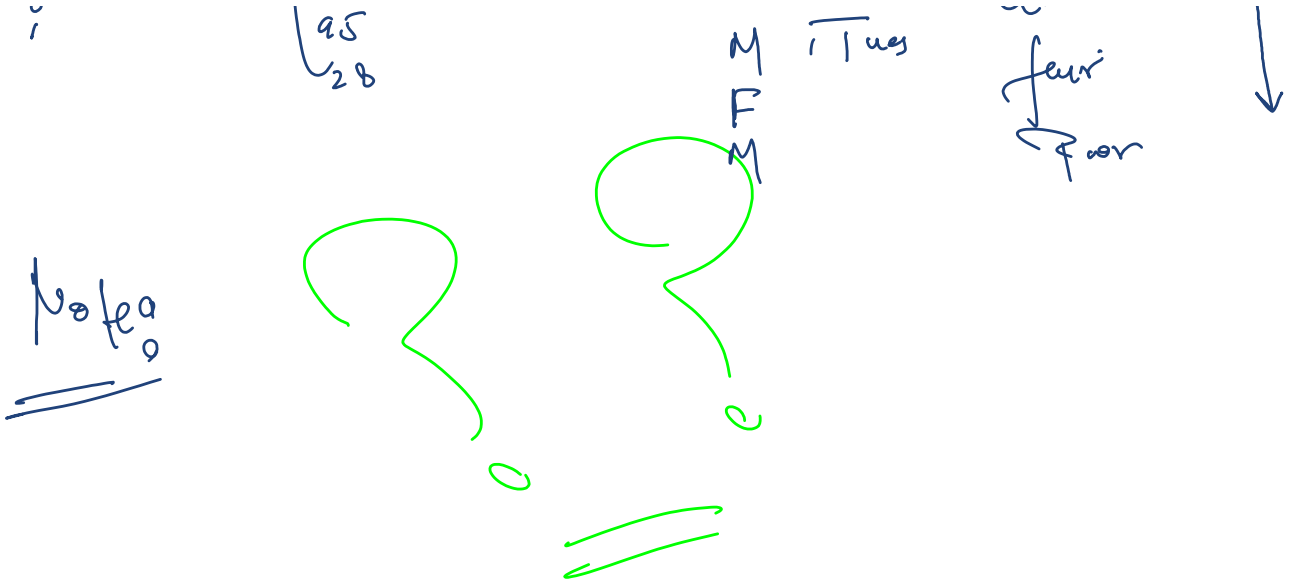
- A variable is any characteristic, number, or quantity that can be measured or counted.

	total_bill	tip	sex	smoker	day	time	size
2	16.99	1.01	Female	No	Sun	Dinner	2
3	10.34	1.66	Male	No	Sun	Dinner	3
4	21.01	3.5	Male	No	Sun	Dinner	3
5	23.68	3.31	Male	No	Sun	Dinner	2
6	24.59	3.61	Female	No	Sun	Dinner	4
7	25.29	4.71	Male	No	Sun	Dinner	4
8	8.77	2	Male	No	Sun	Dinner	2
9	26.88	3.12	Male	No	Sun	Dinner	4
10	15.04	1.96	Male	No	Sun	Dinner	2
11	14.78	3.23	Male	No	Sun	Dinner	2

→ Variables

Types of Variable

- Numerical
  - Discrete (int): 0, 1, 0, 1, 0, 1
  - Continuous (float): 96, 84, 24.8, 95, 28
- Categorical
  - Nominal: M, Sun; F, M; M, Tues
  - Ordinal: Excellent, Good, Fair



Date Time Variables

Date only → 18 June, 2025

Time only → 4:27 Pm IST

Date Time both → 18 June 2023, 4:27 Pm

Mixed Variables

Stu ID	Percentage (Performance)
23	A++
12	85
34	failed
96	failed
26	90
38	90

Mixed

Machine Learning

* It is a Subset of Artificial Intelligence.
* In machine learning we feed data to ML models, And model derive patterns from data which they use to predict the future insights.

Learn the Pattern] Training your model.

Ready for Prediction

New data

Prediction → Correct → Accept

→ Incorrect → Repeat the training process → discard the model


Prediction → Incorrect

18 June 2023 17:09

Types of ML: