Hypothesis Testing Framework

Coin Toss :

10 - match cricket series
Captain always calls "heads"
Won 7 tosses

Fair	Biased
22	3

100 - matches
"Heads"
Won 70 tosses

Fair	Biased
11	22

1000 - matches
"Heads"
Won 700 tosses

Fair	Biased
7	26

Framework to provide a quantifying metric to this intuition

Coin toss example:

(1) What was our default assumption about the coin?

"Coin is fair"

(2) When will we reject this assumption?

"If we have enough proof / evidence / data"

Judge in a Court → A suspect is brought in front of you

① What is our default assumption?

"He is innocent"

② When will you reject this assumption?

"When there is enough evidence"

Innocent until proven guilty

default assumption — evidence to reject

Machine Learning → Deploy (Eg: Youtube / Amazon)

RecSys

Amazon recommends some similar items

(existing ML model → "legacy code" model)

New model to improve performance

① What is the default assumption of the Product Owner?

"New model is not better than legacy model"

② When shall we reject this assumption?

"When we have enough evidence to show that new model is better"

Third umpire: ① on-field has called third umpire
② "soft signal"

① What is the default assumption of the 3rd umpire?

"On-field umpire's soft signal"

② When will he reject this assumption?

"When there is enough evidence to over rule the soft signal"

① Null Hypothesis & Alternate Hypothesis

"Coin fair"

"Person is innocent"

"Legacy is better"

"On-field"

default assumption

"Coin is biased"

"guilty"

"new model is better"

"On-field is wrong"

Burden of proof

H_0 → Null Hypothesis (default assumption)

H_a → Alternate

Recall Today example: H_0: person is innocent
H_a: person is guilty → Burden of proof

Data

Person has a knife in his pocket	Innocent people can carry a knife
Knife has blood stain	Cook / Chef / Butcher
Blood matches victims	This is too much!
CCTV → scene of crime	Highly unlikely that innocent people have these data points
Shirt has fingerprints of victim	
Strong motive to kill	

Probability of seeing such data under the assumption of innocence is very low

$P[\text{data} \mid \text{innocent}]$ is very low

$P[\text{data} \mid H_0]$ "p-value"

If p-value is small → reject H_0

(pronounce guilty)

Deep dive → Coin Toss

10 - matches, "Heads", won 7 Tosses

H_0 : coin is fair

H_a : Biased towards heads

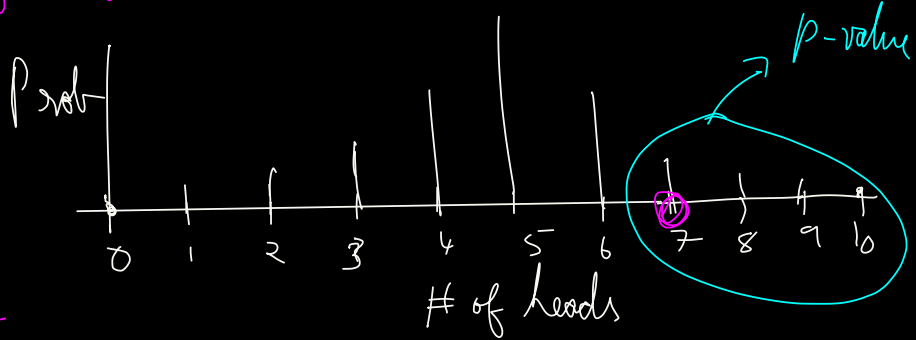
T : Test statistic (function of data / observation)

T = no. of heads → Random variable → distribution Binom

In p-value, we compute prob under the assumption of H_0

pmf under assumption of H_0

Probability of observing data as extreme as that observed ≥ 7



$$P[T \geq 7 \mid H_0] = 1 - P[X \leq 6 \mid H_0]$$

$$= 1 - \text{binom.cdf}(k=6, n=10, p=0.5)$$

$$= 0.172$$

p-value = 0.172

If p-value < "Significance level" (α)

Reject H_0

p-value $< \alpha$

If $\alpha = 0.05$ ⟶ Do not reject H_0

p-value 0.172 ↗

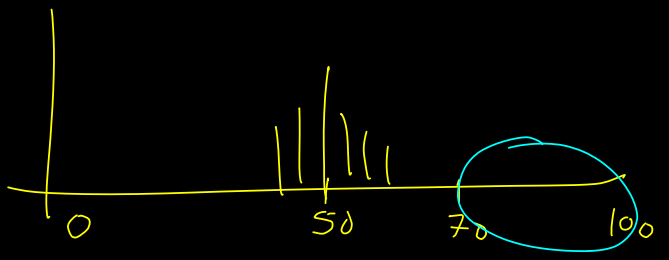
Continue to believe H_0 ⟶ Coin is fair

100 matches → 70 heads

H_0: Coin is fair

H_a: Coin is biased → H

T: no of heads → binomial



$$p\text{-value} \rightarrow P[T \geq 70 \mid H_0] = 1 - P[T \leq 69 \mid H_0]$$

$$= 1 - \text{binom.cdf}(k=69, n=100, p=0.5)$$

$$= 3.9 \times 10^{-5}$$

$\alpha = 0.05$

$p\text{-value} < \alpha$? → Yes

Reject H_0 → "Coin is biased"

1000 tosses → 700 heads

$$P[T \geq 700 \mid H_0] = 1 - \text{binom.cdf}(k=699, n=1000, p=0.5)$$

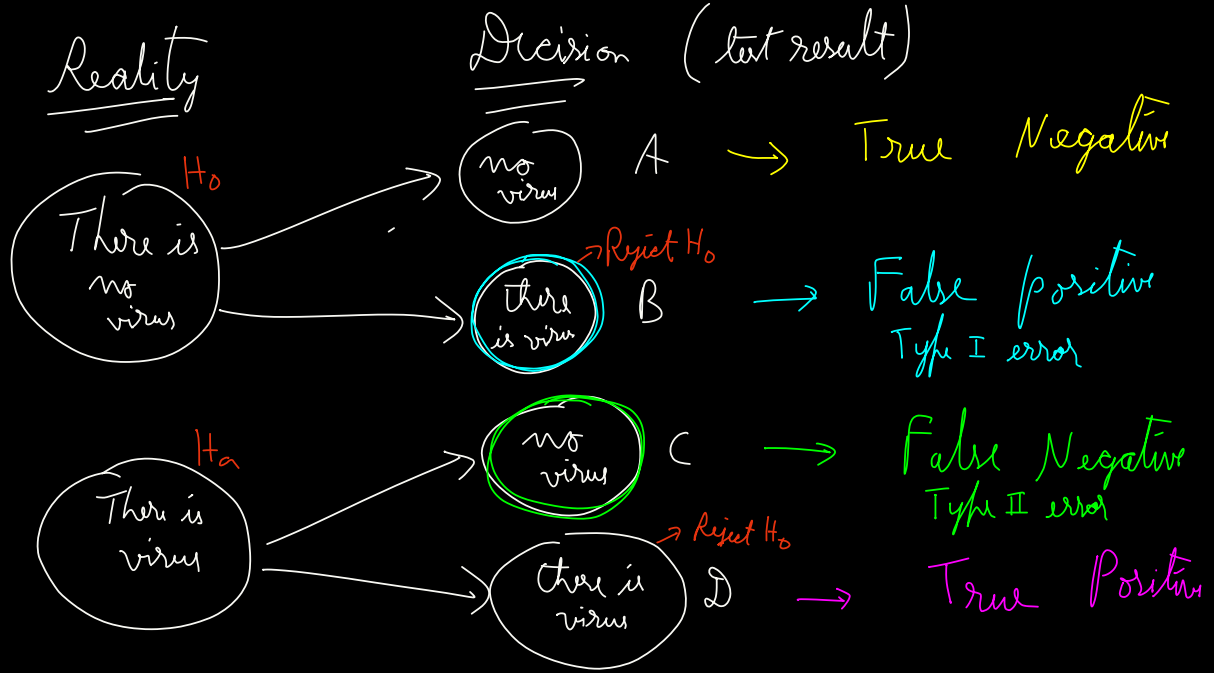
$$= 10^{-16}$$

$p\text{-value} < 0.05$ → Reject H_0

Test for virus

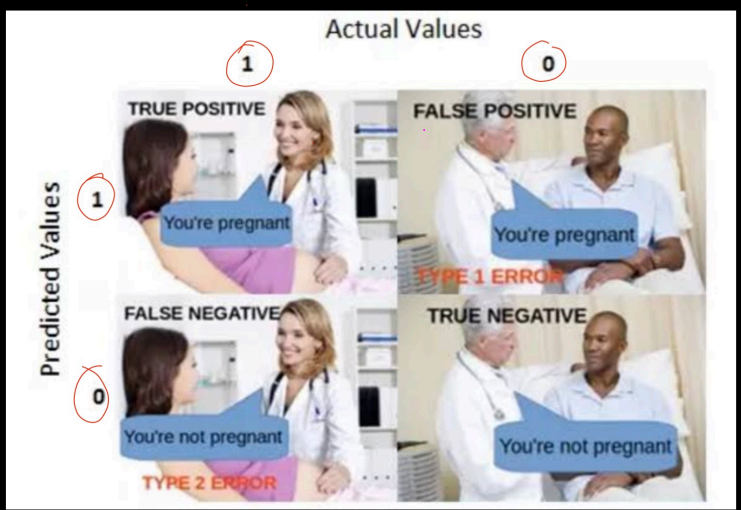
H_0: no virus (negative)

H_a: There is virus (positive)

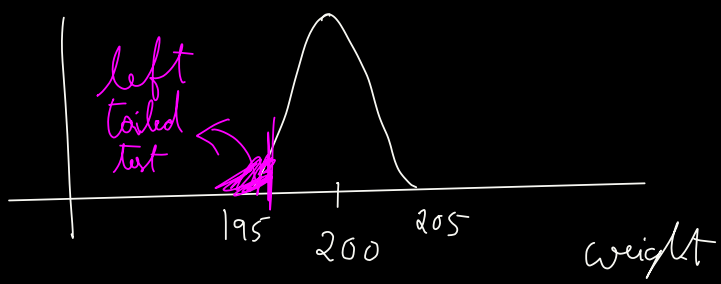


H_0 : not pregnant

H_a : pregnant



Burger : Company claims that their burger weighs 200 gm on average



Unhappy customer, still hungry, disprove the company's claim → weight < 200

H_0: $\mu = 200$ (μ → avg weight

H_a: $\mu < 200$ left side of the mean

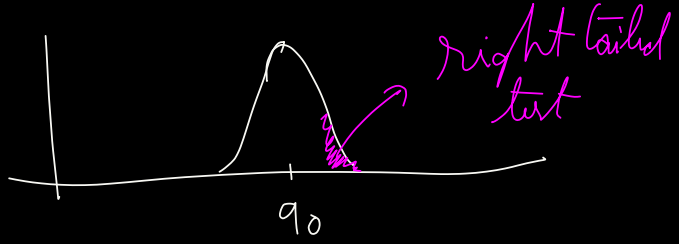
ML deployment : Legacy model → 90% accuracy on average

New model → prove that new model is better.

μ → avg accuracy of new model

$H_0 : \mu = 90$ (or $\mu \leq 90$)

$H_a : \mu > 90$



Height : India avg → 65 inch

Your state → avg height "μ"

$H_0 : \mu = 65$

$H_a : \mu \neq 65$