Cricket series example

The Captain always calls heads

1) 10-match series
   Won 7 tosses

Would you believe
it is a fair coin?

2) 100-match series
   Won 70 tosses

Would you believe
it is a fair coin?

3) 1000-match series
   Won 700 tosses

Would you believe
it is a fair coin?

The important question we need to address is as follows:

What is the framework that can provide a quantifying metric to this intuition of ours?

Cricket series example

1) What is our default assumption?

The coin is fair

2) When shall we reject this assumption?

We shall reject only when we have enough data that makes us conclude otherwise

Judge in court

Assume that you are a judge in a court

A person is brought in front of you as a murder suspect

1) What is our default assumption?

The person is innocent

2) When shall we reject this assumption?

We shall reject only when we have enough data that makes us conclude that he is guilty

Machine Learning Deployment

A machine learning model (legacy) is in production for a few years, and is doing fairly well

You and your team have built a new model, and want to claim that it is better

1) What is the default assumption of the product owner?

The new model is not better than the legacy model

2) When shall we reject this assumption?

When enough data is given that the new model outperforms the legacy model significantly

Third umpire

Suppose you are the third umpire

The on-field umpire has called for your help, and given a soft signal

1) What is our default assumption?

The on-field umpire is correct

2) When shall we reject this assumption?

We shall reject only when we have enough data that makes us conclude that the on-field umpire’s decision can be changed

Fingerprint scanner

We unlock our phones using fingerprint scanner

A finger is now placed on the scanner

1) What should the default assumption be?

The fingerprint does not belong

2) When should the default assumption be rejected?

The default assumption should be rejected only when the data (fingerprint) is very conclusive that it belongs to the owner

Radar example

A radar has to detect a plane

1) What should the default assumption be?

There is no plane

2) When should the default assumption be rejected?

The default assumption should be rejected only when the data is very conclusive that there is a plane

Null Hypothesis H_0 **The technical term for our default assumption**

The coin is fair

The new model is not better than the legacy model

The person is innocent

The on-field umpire is correct

The fingerprint does not belong

There is no plane

All these are examples of setting up the **Null Hypothesis**

Terminologies

H_0 **Null Hypothesis**

Judge in court H_0 The person is innocent

We shall reject only when we have enough data that makes us conclude that he is guilty

Data:

The person has a knife in his pocket	Innocent people can carry knife
The knife has blood stains	Maybe he is a cook/chef
Blood matches that of the victim	Ok, this is too much
His shirt has fingerprints of the victim	Highly unlikely that an innocent man has all these data points

Verdict: Guilty! (Reject the null hypothesis)

Probability of seeing data as extreme as what was observed, under the assumption that he is innocent, is very low

$P\left[\text{data} \mid \text{H0 is true}\right]$ is very low This is called $p-$value

If $p-$value is very low, we reject H0

Terminologies

H_0 **Null Hypothesis**

$p-$**value**

Deep dive: coin toss

Put a quantitative metric on our suspicion that coin is biased

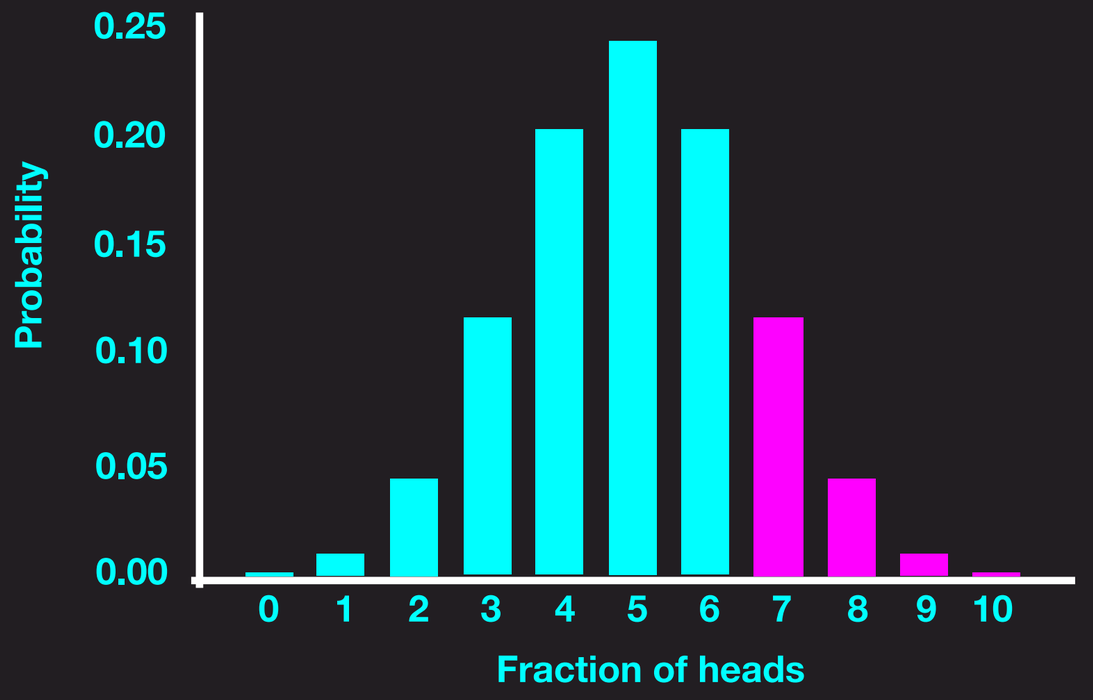
Deep dive: coin toss

Put a quantitative metric on our suspicion that coin is biased

H_0: coin is fair. Probability of heads = 0.5

1) 10-match series 7 Heads

Would you believe it is a fair coin?



Let T = number of heads — Test Statistic

Is T a random variable? Yes

What is its distribution? Binomial

What is the observed value of T? $T_{\text{obs}} = 7$

What is $p-$value?

$p-$value is the probability of observing data as extreme the observed test statistic under the assumption that the null hypothesis is true

The $p-$value is therefore given by $P\left[T \geq 7 \,\middle|\, H_0 \text{ is true}\right]$

How do we compute P-value here?

```
binom.pmf(k=7, n=10, p=0.5) + binom.pmf(k=8, n=10, p=0.5) + binom.pmf(k=9, n=10, p=0.5) + binom.pmf(k=10, n=10, p=0.5)
```

```
1 - binom.cdf(k=6, n=10, p=0.5)
```

$= 0.172$

Deep dive: coin toss

Put a quantitative metric on our suspicion that coin is biased

H_0: coin is fair. Probability of heads = 0.5

Test Statistic

Let T = number of heads

1) 10-match series 7 Heads

$$P\left[T \geq 7 \middle| H_0 \text{ is true}\right] = \texttt{1 - binom.cdf(k=6, n=10, p=0.5)} = 0.172$$

2) 100-match series 70 Heads

$$P\left[T \geq 70 \middle| H_0 \text{ is true}\right] = \texttt{1 - binom.cdf(k=69, n=100, p=0.5)} = 0.000039$$

3) 1000-match series 700 Heads

$$P\left[T \geq 700 \middle| H_0 \text{ is true}\right] = \texttt{1 - binom.cdf(k=699, n=1000, p=0.5)} = 0$$

When do we reject the null hypothesis H0?

When the $p-$value is very low

Typically used threshold is 0.05 (This can change based on business needs)

This threshold is denoted by α and is called Significance Level

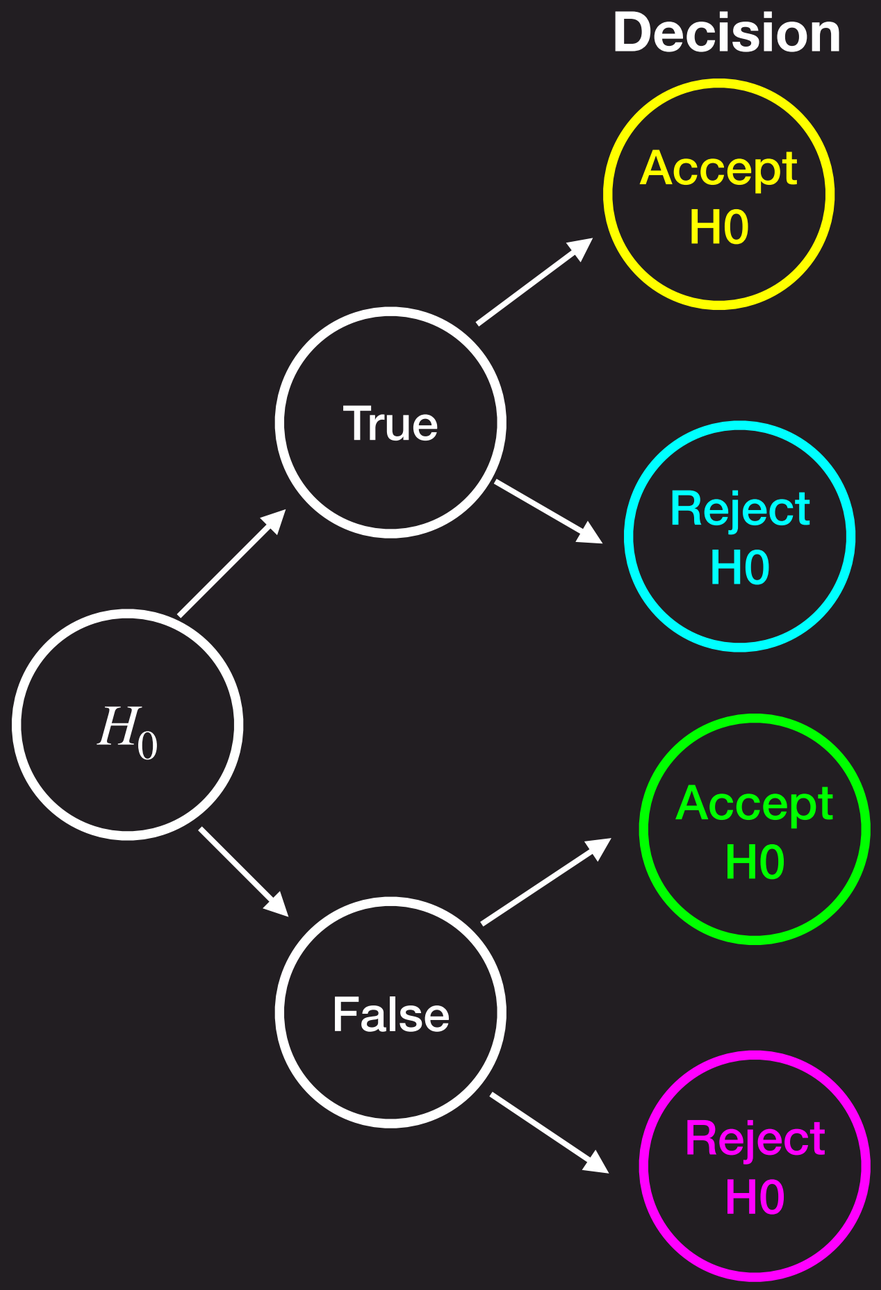
Terminologies

H_0 **Null Hypothesis**

$p-$**value**

Test Statistic

Significance Level



Which of these decisions are errors?

		Decision: Accept	Decision: Reject
H_0	True	True negative $1-\alpha$ Confidence level	False positive α Significance level
	False	False negative β	True positive $1-\beta$ Power

Note: Statisticians do not say “Accept”. They say “fail to reject”

Null Vs Alternate

H0 indicated “null” hypothesis

In the event of rejection of H0, we need the right “alternate hypothesis”

Court example

H_0 “person is innocent”

H_a “person is guilty”

Coin toss with 70% heads

H_0 “coin is fair”

H_a “coin is biased towards heads”

ML deployment

H_0 “The new model is not better than the legacy model”

H_a “The new model is better than the legacy model”

Burger company

A company selling burgers claims that its burger weighs 200 grams.

An unsatisfied customer, still feeling hungry after eating, wants to disprove the claim

Which is the right setup?

H_0: weight = 200

H_a: weight < 200

H_0: weight = 200

H_a: weight ≠ 200

H_0: weight = 200

H_a: weight > 200

H_0: weight >= 200

H_a: weight < 200

H_0: weight <= 200

H_a: weight > 200

AI Chip startup

An AI chip startup claims that it beats the GPU in computer vision tasks

The training time for ResNet is 15 minutes on the GPU

What is the right way for the startup to claim that it is better?

Which is the right setup?

H_0: training time = 15

H_a: training time < 15

H_0: training time >= 15

H_a: training time < 15

H_0: training time = 15

H_a: training time ≠ 15

H_0: training time = 15

H_a: training time > 15

H_0: training time <= 15

H_a: training time > 15

Google Pay

You want to convince your parents that paying by Google Pay is faster than paying by cash and waiting for change

Your parents counter by saying that internet may be slow and Google Pay is slower

Let μ be the proportion of times Google Pay was faster than cash

Which is the right setup?

H_0: $\mu = 0.5$
H_a: $\mu < 0.5$

H_0: $\mu = 0.5$
H_a: $\mu \neq 0.5$

H_0: $\mu = 0.5$
H_a: $\mu > 0.5$

H_0: $\mu >= 0.5$
H_a: $\mu < 0.5$

H_0: $\mu <= 0.5$
H_a: $\mu > 0.5$

Height from your state

The average height of Indians is 65 inches

You want to verify whether this is true for people from your state

Let μ be the average height of people from your state

Which is the right setup?

H_0: $\mu = 65$
H_a: $\mu < 65$

H_0: $\mu = 65$
H_a: $\mu \neq 65$

H_0: $\mu = 65$
H_a: $\mu > 65$

H_0: $\mu >= 65$
H_a: $\mu < 65$

H_0: $\mu <= 65$
H_a: $\mu > 65$

Left Vs right tailed

Burger company

H_0: weight = 200

H_a: weight < 200

Left-tailed

AI Chip startup

H_0: training time = 15

H_a: training time < 15

Left-tailed

Google Pay

H_0: μ = 0.5

H_a: μ > 0.5

Right-tailed

Height from your state

H_0: μ = 65

H_a: $\mu \neq 65$

Two-tailed

1) Setup the Null and Alternate Hypothesis

2) Choose the right test statistic

3) Left-tailed Vs Right-tailed Vs Two-tailed

4) Compute p-value

5) If p-value is less than alpha, then reject the null hypothesis