

January 3, 2023

DSML: CC Fundamentals

Hypothesis Testing - 1

"Toughest topic in DSML"

Welcome back!
Happy
New Year!!

Class begins @ 9:05 p.m.



TERMINOLOGY
EVERYWHERE!

Overview of the module :

✓ 3 classes: H.T.

✓ 1 Correlation .

✓ 1 ANOVA .

| P. S .

| Experiment design .

2 Feature engineering .

→ | OpenCV for C.V.

→ | NLTK for NLP .

Cricket Series Example

The captain always calls heads

- i] 10 match series.
7 tosses were won.

Is the coin
fair?

Yes	No
32	10

- 2] 100 match series.
70 tosses were won.

Is the coin
fair?

20 31

- 3] 1000 match series.
700 tosses won.

Is the coin 2 41..
fair?

What is the Math framework which
will allow us to quantify this?

Cricket Series Example

i] What is our default assumption?

The coin is fair.

2] When should we reject this assumption?

When enough data that makes us conclude otherwise.

Judge in court

Assume that you are judging a murder case.

- 1] What is the default assumption about the suspect?

The person is innocent.

- 2] When will we reject the assumption?

When we have sufficient evidence to prove otherwise.

Machine Learning Model Deployment :

An ML algorithm (legacy) is in production.

You and your team have built a new model, and want to replace the legacy model.

- 1] What is the default assumption of the product owner?

The new model is similar to legacy.

- 2] When should the product owner reject this assumption?

If there is enough data to show that the new model performs better.

Third Umpire

Suppose you are the third umpire.

The on-field umpire has given a soft-signal.

i] What is the default assumption?

The person is innocent.

2] When should the assumption be rejected?

When there is enough evidence to show otherwise

Fingerprint scanner.

We unlock our phones using a fingerprint scanner. A finger is placed on the scanner.

1] What should the default assumption be?

The fingerprint does not belong.

2] When should the assumption be rejected?

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?

When there is sufficient evidence for a plane.

TERMINOLOGY TIME!

H_0 : The null hypothesis -

- The coin is fair.
- The new model is not better than the legacy.
- The person is innocent.
- The on-field umpire is correct.
- The fingerprint does not belong.
- There is no plane.

Judge in court H_0 : Person is innocent.

We shall only reject H_0 when we have enough data to conclude guilty.

Data:

- ① The person owns a knife. X
- ② The knife has blood stain on it. X
- ③ Blood matches ^{and blood} the victim. ~
- ④ The fingerprints ^{and blood} of the victim are on the suspect's shirt. ✓
- ⑤ Eye-witness account.

Verdict:

Probability of seeing data as extreme as what was observed, under the null hypothesis, is very low.

$P[\text{data}]$
P-value.

$\underline{H_0 \text{ is true.}}$ is very low.

↳ They all say guilty.

Deep dive: Coin toss example

* How to quantify our suspicion
that the coin is biased?

Ans: Use all our Probability
skills !!

Coin Toss

H_0 : Coin is fair. $\Rightarrow P(H) = 0.5$.

Case 1: 10 match series, 7 heads were observed.

Would you believe that the coin is fair? Verdict: $P = 0.17$ so H_0 cannot be rejected.

(a) What is the random variable?

T: Number of heads. TEST STATISTIC.

(b) What is its distribution?

Binomial.

(c) What is the observed value?

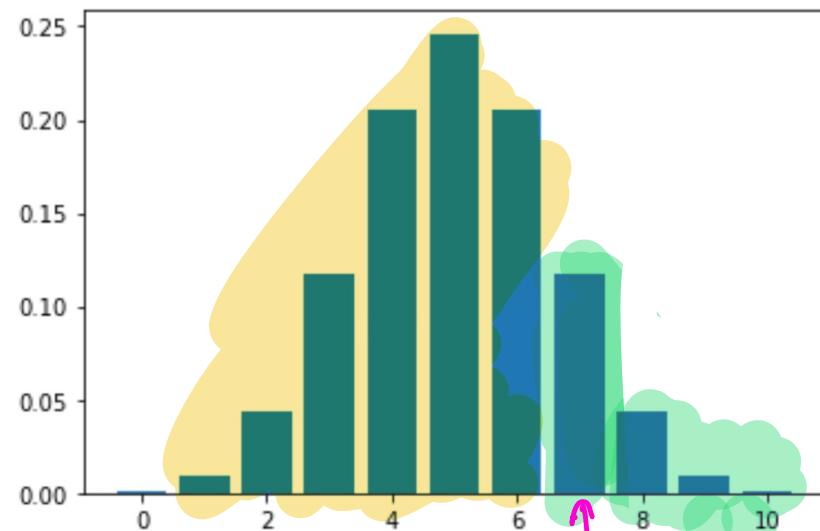
7.

→ (d) What is the probability of our observation assuming H_0 is true?

$P[T = 7 \mid H_0 \text{ is true}]$.

(d) What is the probability of our observation assuming H_0 is true?

Typically, we reject the null hypothesis if $P < 0.05$.



$$Q: (a) P[T = 7 | H_0] \rightarrow$$

$$\checkmark (b) P[T \geq 7 | H_0]. \rightarrow$$

$$\text{binom.pmf}(k=7, n=10, p=0.5)$$

$$0.12$$

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

$$0.17$$

Coin Toss

H_0 : coin is fair $\Rightarrow P(H) = 0.5$.

Test statistic :

1] 10-match series, 7 heads.

$$0.17 > 0.05$$

so \rightarrow

$P[T \geq 7 | H_0 \text{ is true}] = 0.17$, H_0 is not rejected.

✓ 2] 100-match series, 70 heads.

$P[T \geq 70 | H_0 \text{ is true}] = \frac{2.317 \times 10^{-5}}{H_0 \text{ is rejected}}$.

✓ 3] 1000-match series, 700 heads.

$P[T \geq 700 | H_0 \text{ is true}] = \frac{5.06 \times 10^{-38}}{H_0 \text{ is rejected}}$.

Terminologies

H_0 : Null hypothesis : default assumption.
"status quo"

p-value : Probability of the data given
the null hypothesis is true.

Test statistic : The random variable
of interest.

* Significance level : the threshold value
below which we reject the Null
hypothesis. ($P = 0.05$).

Radar Example

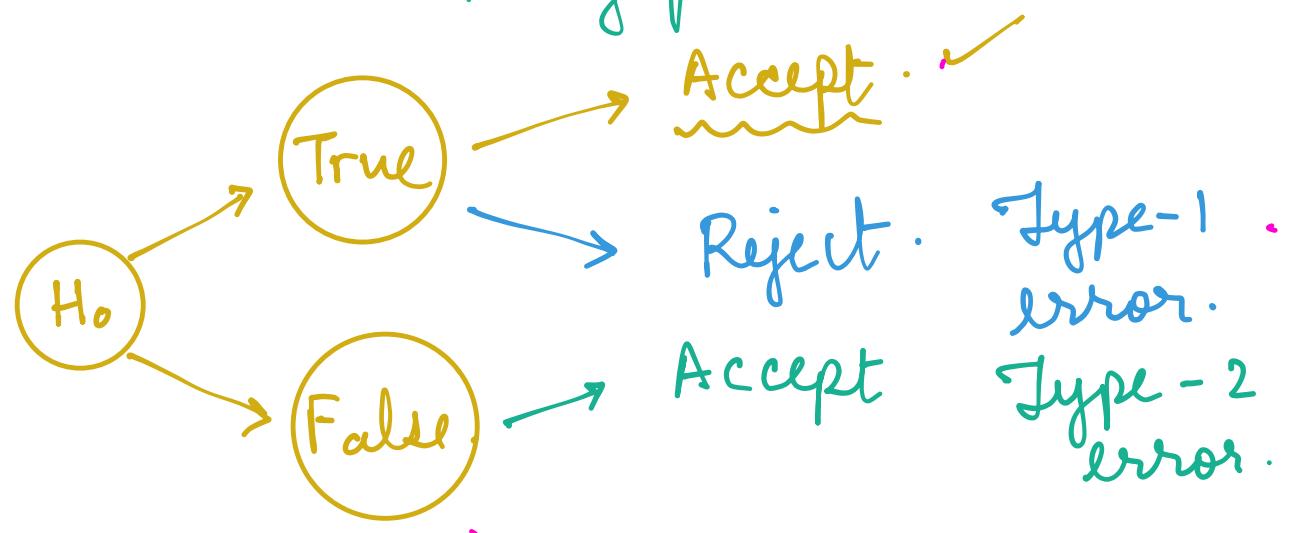
H_0 : There is no plane.

"Null hypothesis"

H_a : There is a plane /

UFO/bogey.

"Alternate hypothesis"



		Decision	
		H_0	$\neg H_0$
H_0	T	True positive. $(1-\beta)$	False negative. (β)
	F	False positive. (α)	True negative. $(1-\alpha)$

Null hypothesis vs. Alternate hypothesis.

H_0 : Null hypothesis

H_a : 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"

ML deployment.

H_0 : "New same as legacy"

H_a : "New model is better".

Practice Problems: Identify the correct Null and Alternative hypotheses.

Burger company

A company selling burgers claims that its burgers weigh 200 gms. A hungry customer wants to disprove it.

$$H_0 : \text{weight} = 200$$

$$H_a : \text{weight} < 200.$$

$$H_0 : \text{weight} = 200$$

$$\rightarrow H_a : \text{weight} \neq 200.$$

$$H_0 : \text{weight} \geq 200$$

$$H_a : \text{weight} < 200.$$

$$\rightarrow H_0 : \text{weight} = 200$$

$$\rightarrow H_a : \text{weight} > 200. \checkmark$$

$$- H_0 : \text{weight} \leq 200$$

$$H_a : \text{weight} > \underline{200}:$$

Practice Problems: Identify the correct Null and Alternative hypotheses.

AI Chip Startup.

An AI chip startup claims that it beats the GPU in CV tasks. The training time for ResNET on GPU₁ is 15 minutes.

→ T: Training time for new chip.

$$H_0: T \geq 15.$$

$$H_a: T < 15.$$

Practice Problems: Identify the correct Null and Alternative hypotheses.

Height from your state.

The average height of Indians is 65 inches.

You want to verify that this is true for your state. You want to show that people in your state are taller.

T: Average height for my state.

$$H_0 : T \leq 65.$$

$$H_a : T > 65.$$

Practice Problems: Identify the correct Null and Alternative hypotheses.

Google Pay

You want to convince your parents that paying by GPay is faster than paying cash and waiting for change. Let M denote the ratio between time for Gpay to time for cash.

$$T : M = \frac{\text{Time for gpay}}{\text{Time for cash.}}$$

$$H_0: M \geq 1.$$

$$H_a: M < 1.$$

Steps for Hypothesis testing.

- 1] Set up / choose the null and alternate hypotheses .
- 2] Choose the correct test statistic .
- 3] Write down the correct expression for the p - value (conditional prob .).
- 4] Compute the p - value .
- 5] If p - value < significance level .
reject the null hypothesis .