



Hypothesis Testing

By : L Vinay Rajiv Reddy

What is hypothesis testing ?

- Data alone isn't enough.
- Interpretation from the data does matters.

-

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa
10	5.4	3.7	1.5	0.2	Iris-setosa
11	4.8	3.4	1.6	0.2	Iris-setosa
12	4.8	3.0	1.4	0.1	Iris-setosa
13	4.3	3.0	1.1	0.1	Iris-setosa
14	5.8	4.0	1.2	0.2	Iris-setosa
15	5.7	4.4	1.5	0.4	Iris-setosa
16	5.4	3.9	1.3	0.4	Iris-setosa
17	5.1	3.5	1.4	0.3	Iris-setosa

Data Set : Popular Trains in SCR

NAME	TYPE	DIVISION	DISTANCE TRAVELLED
Palnadu	Exp.	Guntur	350KM
Kakatiya	Pass.	Sec-bad	128KM
Pinakini	Exp.	Vijayawada	399KM
Golconda	Exp.	Guntur	458 KM
LPI – BZA	Exp.	Vijayawada	349KM
Repalli	Pass.	Guntur	401KM
Narayanadri	Exp.	Guntakal	676KM
SC – GNT	Exp.	Sec-bad	458KM
Cocanada	Exp.	Vijayawada	652 KM
NDKD - MCL	Pass.	Guntur	35KM

- HYPOTHESIS – Investigation from the data

- H_0 – NULL HYPOTHESIS

- H_1 – ALTERNATIVE HYPOTHESIS

- Mutually Exclusive Events

- Eg:

1) **H_0** : India is going to win 99 medals in upcoming Tokyo Olympics.

H_1 : India is not going to win 99 medals in upcoming Tokyo Olympics.

2) **H_0** : Guntur Division has more no. of popular trains.

H_1 : Guntur Division doesn't have more no. of popular trains.

3) **H_0** : Vijayawada is the only division to have all Express trains.

H_1 : Vijayawada is not only division to have all Express trains.

STEPS:

- 1) Initial Assumption.
- 2) Collecting the evidences (population sample).
- 3) Analyzing to reject NULL or not. (based on P value)
- 4) If null is proven to be false, we go for H_1 .

How to perform

- While analyzing the data/ evidences, we obtain P Value/ Significant value.
- If $p \leq 0.05$ (Reject H_0)
- Else consider H_0
- **Methods to perform :**
 - 1) Chi Square Test
 - 2) T Test
 - 3) Anova Test

- If (2 Categorical Features)

➔ Chi Square Test

- If (1 Categorical Features and 1 Continuous Features)

➔ T Test

- If (Categorical data having more than 3 unique values)

➔ Anova Test

- **Errors:**

- 1) Type -1 Error :

- If H_0 was rejected by analysis, but indeed it's only true

- 2) Type-2 Error :

- If H_0 is analyzed as True, but indeed it's not true.