Hypothesis Testing

A guide by Satyajit Pattnaik

Steps in a data-driven decision making

Formulate a hypothesis

Find the right test

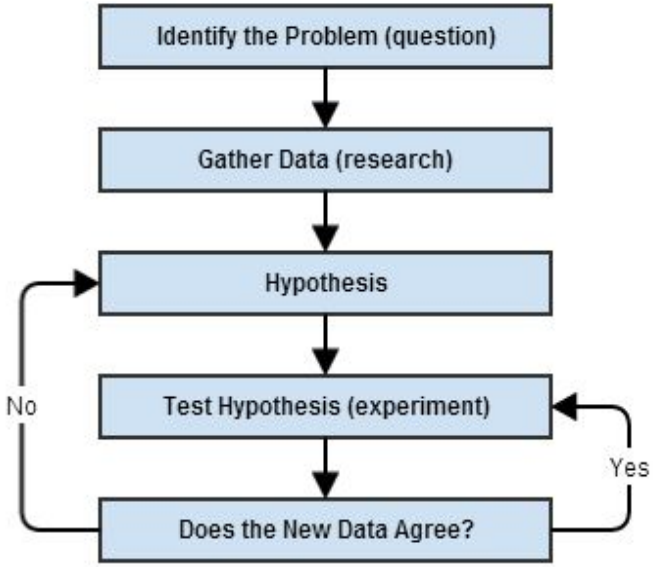
Execute the test

Making a decision

for illustration purposes only

What is a Hypothesis?

*A hypothesis is an idea that can be tested



*not a formal definition

House rent in Hong Kong is very expensive
→ **Statement**

On an average, house rent in Hong Kong is expensive & more than 1,00,000 INR
→ **Hypothesis**

Satyajit Pattnaik

Not a hypothesis

Salary

Data Analyst vs Data Scientist

No data → cannot be tested → not a hypothesis

Satyajit Pattnaik

Example

As per Glassdoor, Microsoft has 27 salaries mentioned by various employees, and the average salary of a Data Scientist in Microsoft is 16.7 lakhs per annum

*source: glassdoor.co.in

Satyajit Pattnaik

Null vs Alternative Hypothesis

Null Hypothesis (H_0)

A statement about a population parameter

We test the likelihood of this statement being true in order to decide whether to accept or reject our alternative hypothesis.

Can include =, <=, or >= sign

Alternative Hypothesis (H_A or H_1)

A statement that directly contradicts the null hypothesis

We determine whether or not to accept or reject this statement based on the likelihood of the null (opposite) hypothesis being true.

Can include a $\neq$, >, or < sign

To be tested

Satyajit Pattnaik

For illustration purposes only

Example

As per Glassdoor, Microsoft has 27 salaries mentioned by various employees, and the average salary of a Data Scientist in Microsoft is 16.7 lakhs per annum

Data Scientist
Microsoft
27 salaries

₹ 16,73,507/yr

₹230K ₹3,350K

*source: glassdoor.co.in

$H_0 : \mu_0 =$ Rs. 16.7 lakhs

$H_1 : \mu_0 \neq$ Rs. 16.7 lakhs

Accept if: μ_0 is close enough to the true mean

Reject if: μ_0 is too far from the true mean

glassdoor

Satyajit Pattnaik

Good luck!

I hope you had some good understanding on hypothesis testing.

Satyajit Pattnaik