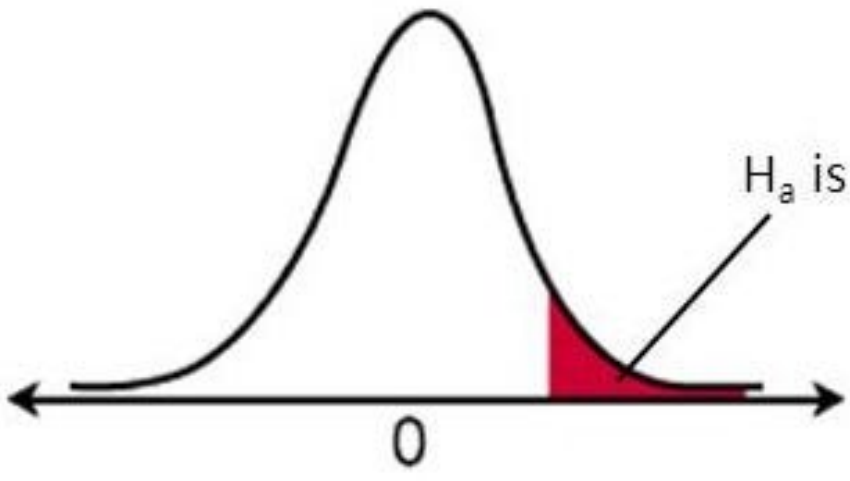
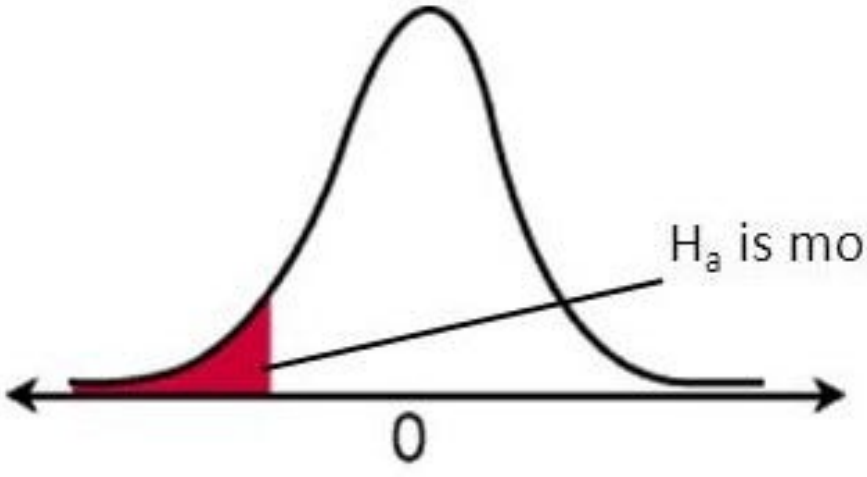
Analytics Vidhya

A Guide to Hypothesis Testing



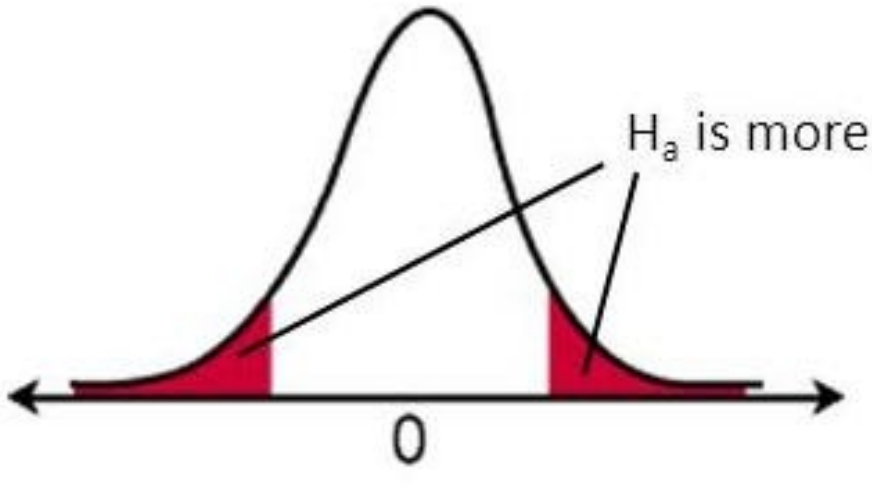
Right-tail test

$H_a: \mu > \text{value}$



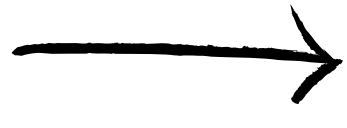
Left-tail test

$H_a: \mu < \text{value}$



Two-tail test

$H_a: \mu \neq \text{value}$



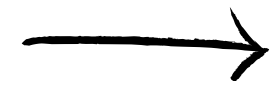
Analytics Vidhya

What is Hypothesis Testing?

- In its simplest form, hypothesis testing is a procedure used to determine if there is enough evidence in a sample of data to infer that a certain condition is true for the entire population.

    - **Null Hypothesis (H0)**: Represents a statement of no effect or no difference and is what you aim to reject.

    - **Alternative Hypothesis (H1)**: Represents what you want to prove.



Steps in Hypothesis Testing?

- **Set Up Hypotheses**: Define your H0 and H1.

- **Choose Significance Level (α)**: Commonly set at 0.05, this represents the probability of rejecting H0 when it is true.

- **Select Test & Compute Test Statistic**: Depending on data and distribution, select appropriate tests like t-test, chi-square, etc.

- **Make a Decision**: If the p-value is less than α, reject the null hypothesis.

Analytics Vidhya

Common Test in Machine Learning

- **T-test**: Compares the means of two groups.
- **Chi-square test**: Tests relationships between categorical variables.
- **ANOVA** (Analysis of Variance): Compares means among more than two groups.
- **F-test**: Compares variances of two populations.

Analytics Vidhya

Python Code to Implement Hypothesis Testing using T - Test

```python
# Import necessary libraries
import numpy as np
from scipy import stats

# Example data: scores of two groups of students
group_A = np.array([90, 85, 88, 78, 82, 80, 77, 92, 89, 84])
group_B = np.array([73, 70, 74, 65, 68, 66, 72, 71, 67, 69])

# Perform t-test
t_stat, p_value = stats.ttest_ind(group_A, group_B)

# Display results
print(f"t-statistic: {t_stat:.2f}")
print(f"p-value: {p_value:.4f}")

# Check significance
alpha = 0.05  # significance level
if p_value < alpha:
    print("Reject the null hypothesis: There's a significant difference
          between the means of the two groups.")
else:
    print("Fail to reject the null hypothesis: There's no significant
          difference between the means of the two groups.")
```