# Practical 4 Hypothesis Testing

Formulate null and alternative hypotheses for a given problem.

Conduct a hypothesis test using appropriate statistical tests (e.g., t-test, chi-square test). Interpret the results and draw conclusions based on the test outcomes.

# T-test (Comparing Salaries Between Marketing and Sales Departments):

The goal is to test if the average salaries in the **Marketing** and **Sales** departments are the same or different.

# Null Hypothesis (H₀):

* The null hypothesis assumes that there is **no significant difference** between the average salaries of employees in the Marketing and Sales departments.
* **Mathematically**: **H₀: μ₁ = μ₂** Where:
  + μ₁ is the average salary in the Marketing department
  + μ₂ is the average salary in the Sales department

# Alternative Hypothesis (H₁):

* The alternative hypothesis assumes that there is a **significant difference** between the average salaries in the Marketing and Sales departments.

# Mathematically: H₁: μ₁ ≠ μ₂

This is a two-tailed test, meaning we're testing for any difference (higher or lower) in the average salaries of the two departments.

# Two-sample t-test

(ttest\_ind) to compare the means of the two independent groups (Marketing and Sales). This test will help us determine if there is enough evidence to reject the null hypothesis.

Code:

import numpy as np from scipy import stats

# Data: Salaries of employees in Marketing and Sales departments marketing\_salaries = [50000, 55000, 60000, 52000, 58000, 49000]

sales\_salaries = [48000, 51000, 54000, 53000, 50000, 46000]

# Perform the two-sample t-test

t\_stat, p\_value = stats.ttest\_ind(marketing\_salaries, sales\_salaries)

# Calculate degrees of freedom for t-test n1 = len(marketing\_salaries)

n2 = len(sales\_salaries)

df\_ttest = n1 + n2 - 2 # For independent two-sample t-test

print(f"T-statistic: {t\_stat}") print(f"P-value: {p\_value}")

print(f"Degree of Freedom for t-test: {df\_ttest}")

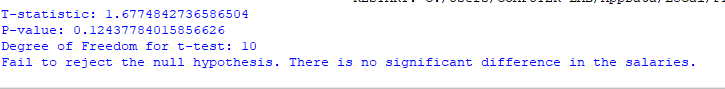
# Hypothesis interpretation alpha = 0.05

if p\_value < alpha:

print("Reject the null hypothesis. There is a significant difference in the salaries.") else:

print("Fail to reject the null hypothesis. There is no significant difference in the salaries.")

# Interpreting the Results:

****

**Conclusion:**

Based on the **p-value**:

* Since **p-value > alpha**, we **fail to reject** the null hypothesis.
* Therefore, we conclude that there is **no significant difference** in the average salaries between the Marketing and Sales departments.

Code for Chi-Square Test

from scipy.stats import chi2\_contingency

# Data: Frequency of employees in each department across two regions (North, South) data = [

[30, 20, 50], # North Region (Marketing, Sales, HR)

[20, 30, 40] # South Region (Marketing, Sales, HR)

]

# Perform the Chi-Square test

chi2\_stat, p\_value, dof, expected = chi2\_contingency(data)

print(f"Chi-square statistic: {chi2\_stat}") print(f"P-value: {p\_value}")

print(f"Degree of Freedom for Chi-Square Test: {dof}") print(f"Expected Frequencies: \n{expected}")

# Hypothesis interpretation alpha = 0.05

if p\_value < alpha:

print("Reject the null hypothesis. The distribution of employees is different across regions.")

else:

print("Fail to reject the null hypothesis. The distribution of employees is the same across regions.")

