A white robot hand with red and blue stripes

Description automatically generated



**AI/ML**

**LAB MANUAL**

**Disclaimer: The content is curated from online/offline resources and used for educational purpose only**

**4.3 Analyzing Student Performance Using T-Tests and Z-Tests**

**Objective:**

To determine whether there are statistically significant differences in student performance based on gender, study time, or other factors using T-Tests and Z-Tests.

**Equipment Required:**

* Computer with Python and Jupyter Notebook installed
* Dataset: Student Performance (<https://archive.ics.uci.edu/dataset/320/student+performance>)
* Python Libraries: pandas, numpy, matplotlib, seaborn, scipy, statsmodels

**Prerequisites:**

* Basic understanding of Python programming
* Knowledge of statistics, including hypothesis testing
* Familiarity with pandas for data manipulation
* Understanding of data visualization using matplotlib and seaborn

**Problem Statement:**

This lab activity aims to analyze the impact of various factors, such as gender and study time, on student performance using statistical hypothesis testing. By applying **T-Tests and Z-Tests**, we will evaluate whether differences in student scores are statistically significant. The experiment will provide insights into whether these factors play a measurable role in academic performance and will help in making data-driven educational decisions.

**Code/Procedure:**

**# Step 1: Load Required Libraries**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from scipy import stats

**# Step 2: Load Dataset**

df = pd.read\_csv("student\_performance.csv", sep=';') # Modify path if needed

**# Step 3: Data Preprocessing**

**# Checking for missing values**

print(df.isnull().sum())

**# Display dataset information**

print(df.info())

print(df.describe())

**# Step 4: Exploratory Data Analysis (EDA)**

**# Visualizing the distribution of scores by gender**

plt.figure(figsize=(10,5))

sns.boxplot(x='sex', y='G3', data=df)

plt.title("Distribution of Final Grades by Gender")

plt.show()

**# Visualizing the impact of study time on final grades**

plt.figure(figsize=(10,5))

sns.boxplot(x='studytime', y='G3', data=df)

plt.title("Impact of Study Time on Final Grades")

plt.show()

**# Step 5: Hypothesis Testing**

**# H0: No significant difference in performance based on gender**

**# H1: A significant difference exists**

def perform\_ttest(score\_column):

    male\_scores = df[df['sex'] == 'M'][score\_column]

    female\_scores = df[df['sex'] == 'F'][score\_column]

    t\_stat, p\_value = ttest\_ind(male\_scores, female\_scores, equal\_var=False)  # Welch’s t-test

    return t\_stat, p\_value

**# Conduct t-tests**

g3\_t, g3\_p = perform\_ttest('G3')

**# Display results**

print(f"Final Grade (G3) T-test: t-statistic = {g3\_t:.3f}, p-value = {g3\_p:.5f}")

**# Step 6: Z-Test (Comparing study time groups)**

low\_study = df[df['studytime'] <= 2]['G3']

high\_study = df[df['studytime'] > 2]['G3']

z\_stat, z\_p\_value = ztest(low\_study, high\_study)

print(f"Study Time Z-test: z-statistic = {z\_stat:.3f}, p-value = {z\_p\_value:.5f}")

**# Step 7: Interpretation**

alpha = 0.05  # Significance level

def interpret\_results(p\_value, test\_name):

    if p\_value < alpha:

        print(f"There is a significant difference ({test\_name}, p < 0.05). Reject H0.")

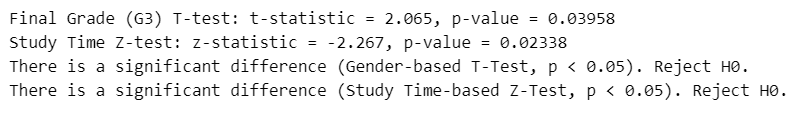
    else:

        print(f"No significant difference ({test\_name}, p >= 0.05). Fail to reject H0.")

interpret\_results(g3\_p, "Gender-based T-Test")

interpret\_results(z\_p\_value, "Study Time-based Z-Test")

**Output:**



**Conclusion**

If the p-value is \*\*less than 0.05\*\*, we reject the null hypothesis, indicating a significant difference.

If the p-value is \*\*greater than 0.05\*\*, we fail to reject the null hypothesis, indicating no significant difference.

The results from T-Tests and Z-Tests will help in understanding the impact of gender and study time on academic performance.