# Statistical Analysis Report: Understanding T-Tests

**A Demonstration of Independent and Paired T-Tests**

**Submitted by:** Mausam Kar(24BAI10284)

## Github:

https://github.com/Mausam5055/Data-Science/tree/main/Assignment-4

## Table of Contents



### 1. Introduction to Hypothesis Testing

In statistics, a T-Test is a fundamental tool used for hypothesis testing. It allows us to determine if there is a significant difference between the means (averages) of two groups. The core question a T-Test answers is whether an observed difference between two groups is likely due to a real effect or if it's simply due to random chance.

There are two primary types of T-Tests, which depend on the nature of the data being compared:

- **Independent T-Test:** Used when comparing the means of two separate, unrelated groups (e.g., comparing the test scores of two different classrooms).
- **Paired T-Test (or Dependent T-Test):** Used when comparing the means of the same group at two different times or under two different conditions (e.g., comparing the blood pressure of patients before and after taking a medication).

This report demonstrates how to perform both types of tests and, more importantly, how to interpret their results.

### 2. Objectives

The core objectives of this analysis are:

- **Demonstrate T-Tests:** To show how to perform both Independent and Paired T-Tests using Python's scipy library.
- **Understand Key Outputs:** To explain the meaning of the T-statistic and the P-value, which are the primary outputs of a T-Test.
- **Interpret Results:** To correctly interpret the results of the tests to determine if the differences between group means are statistically significant.

### 3. Methodology and Python Implementation

The analysis was performed using Python and the scipy.stats module, which provides robust functions for performing a wide range of statistical tests.

**Complete Python Script:**

```
`
after  = [85, 89, 92, 87, 90]

t_stat_rel, p_val_rel = stats.ttest_rel(before, after)

print("🔹 Paired T-Test")
print(f"T-statistic: {t_stat_rel:.4f}")
print(f"P-value: {p_val_rel:.4f}")

# Interpretation
if p_val_rel < alpha:
    print("Result: The difference between the 'before' and 'after' scores is statistically significant.")
else:
    print("Result: The difference between the 'before' and 'after' scores is not statistically significant.")
```

### 4. Results and Interpretation

The script performs two separate tests. We will analyze each one. For both tests, we will use a standard significance level (alpha) of **α = 0.05**. This means we are willing to accept a 5% chance of concluding there is a difference when there isn't one.

#### 4.1. Independent T-Test

This test compares the means of two distinct, unrelated groups (group1 and group2).

**Script Output:**

🔹 Independent T-Test
T-statistic: -3.5444
P-value: 0.0077
Result: The difference between the groups is statistically significant.

**Interpretation:**

- **T-statistic (-3.5444):** This value measures the size of the difference between the group means relative to the variation within the groups. A larger absolute T-statistic indicates a larger difference between the groups. The negative sign simply indicates that the mean of the first group is smaller than the mean of the second group.

- **P-value (0.0077):** This is the most critical value for our conclusion. The P-value represents the probability of observing a difference as large as (or larger than) the one in our data, assuming that there is no real difference between the groups (i.e., the "null hypothesis" is true).

- **Conclusion:** Our P-value (0.0077) is less than our chosen alpha (0.05). Therefore, we **reject the null hypothesis**. We conclude that the observed difference between group1 and group2 is **statistically significant** and not likely due to random chance.

**4.2. Paired T-Test**

This test compares the means of a single group at two different points in time (before and after).

**Script Output:**

🔹 Paired T-Test
T-statistic: 4.5410
P-value: 0.0105
Result: The difference between the 'before' and 'after' scores is statistically significant.

**Interpretation:**

- **T-statistic (4.5410):** This value indicates that there is a substantial difference between the 'before' and 'after' scores, relative to the consistency of the differences for each pair.
- **P-value (0.0105):** This is the probability of seeing the observed average difference (or a greater one) between the 'before' and 'after' scores purely by chance, if the intervention had no real effect.
- **Conclusion:** Our P-value (0.0105) is less than our alpha (0.05). Therefore, we again **reject the null hypothesis**. We conclude that the change from the 'before' state to the 'after' state represents a **statistically significant** difference.

**5. Conclusion**

This report successfully demonstrated the application and interpretation of two fundamental statistical tests.

The key takeaways are:

1. **Test Selection is Crucial:** The choice between an Independent and a Paired T-Test depends entirely on the experimental design. Using the wrong test can lead to invalid conclusions.
2. **P-value Drives Decisions:** The P-value is the primary evidence used to make a statistical decision. By comparing it to a pre-determined significance level (alpha), we can conclude whether our results are statistically meaningful.
3. **Significant Results Found:** In both of our examples, the P-values were below 0.05, leading us to conclude that the differences observed in both the independent and paired scenarios were statistically significant.

Understanding how to perform and interpret these tests is a foundational skill in data analysis, allowing one to move beyond simple observation to data-driven decision-making.