# Statistical Analysis Report: The One-Sample Z-Test

**Comparing a Sample Mean to a Population Mean:**

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

Github: https://github.com/Mausam5055/Data-Science/tree/main/Assignment-3

## Table of Contents



### 1. Introduction to the Z-Test

In statistics, a **one-sample Z-test** is a hypothesis test used to determine whether a sample mean is statistically different from a known or hypothesized population mean. This test is typically used when the population standard deviation is known and the sample size is sufficiently large (usually $n > 30$).

The core question a Z-test answers is: "Is the mean of our collected sample significantly different from the established population average, or is the difference we see likely just due to random sampling variability?"

For example, a school administrator might want to know if the average test score of their students (sample_data) is significantly different from the national average test score (population_mean). The Z-test provides a statistical framework to answer this question.

### 2. Objectives

The core objectives of this analysis are:

- **Demonstrate a Z-Test:** To show how to perform a one-sample Z-test using Python's statsmodels library.
- **Understand Key Outputs:** To explain the meaning of the Z-score and the P-value

in the context of this test.

- **Interpret Results:** To correctly interpret the test's outcome to determine if the sample mean is statistically different from the population mean.
- **Visualize the Comparison:** To create a histogram to visually compare the sample distribution to both the sample and population means.

### 3. Methodology and Python Implementation

The analysis was performed using Python, with the numpy library for data generation, statsmodels for the statistical test, and matplotlib for visualization.

**Complete Python Script:**

```
import numpy as np
from statsmodels.stats.weightstats import ztest
import matplotlib.pyplot as plt

# Step 1: Generate sample data (e.g., students' test scores)
np.random.seed(42)  # For reproducibility
sample_data = np.random.normal(loc=70, scale=10, size=100)  # Mean=70,
SD=10, n=100

# Step 2: Define the known population mean to test against
population_mean = 75

# Step 3: Perform the one-sample Z-test
# The 'value' parameter is the population mean we are testing against.
z_score, p_value = ztest(sample_data, value=population_mean)

# Step 4: Print results
print("🔍 One-Sample Z-Test")
print(f"Z-score: {z_score:.3f}")
print(f"P-value: {p_value:.4f}")

# Step 5: Interpret the results based on a significance level (alpha)
alpha = 0.05
if p_value < alpha:
    print("→ Result: The sample mean is significantly different from the
population mean.")
else:
    print("→ Result: The sample mean is NOT significantly different from the
```

```
population mean.")

# Step 6: Optional - Visualize the distribution
plt.figure(figsize=(10, 6))
plt.hist(sample_data, bins=15, edgecolor='black', alpha=0.7, label='Sample Data Freq.')
plt.axvline(np.mean(sample_data), color='blue', linestyle='dashed', linewidth=2, label=f'Sample Mean: {np.mean(sample_data):.2f}')
plt.axvline(population_mean, color='red', linestyle='solid', linewidth=2, label=f'Population Mean: {population_mean}')
plt.title("Sample Data Distribution vs. Population Mean")
plt.xlabel("Test Scores")
plt.ylabel("Frequency")
plt.legend()
plt.grid(axis='y', alpha=0.5)
plt.show()
```

### 4. Results and Interpretation

The script performs a one-sample Z-test comparing the mean of our generated sample_data to a known population_mean of 75. We will use a standard significance level (alpha) of **α = 0.05**.

#### 4.1. Z-Test Outputs

**Script Output:**

🔍 One-Sample Z-Test
Z-score: -5.097
P-value: 0.0000
⟶ Result: The sample mean is significantly different from the population mean.

**Interpretation:**

- **Z-score (-5.097):** The Z-score (or Z-statistic) measures how many standard deviations the sample mean is away from the population mean. Here, our sample mean is approximately 5.1 standard deviations *below* the population mean. This is a very large difference.
- **P-value (0.0000):** This value represents the probability of observing a sample mean as extreme as ours (or more extreme), assuming the sample was actually

drawn from a population with a mean of 75. A P-value this small (effectively zero) indicates that our result is extremely unlikely to have occurred by random chance alone.

- **Conclusion:** Since our P-value (0.0000) is much smaller than our alpha (0.05), we **reject the null hypothesis**. We have strong statistical evidence to conclude that the mean of the sample is **significantly different** from the population mean of 75.

### 4.2. Visualizing the Data

The histogram provides a clear visual representation of our findings.

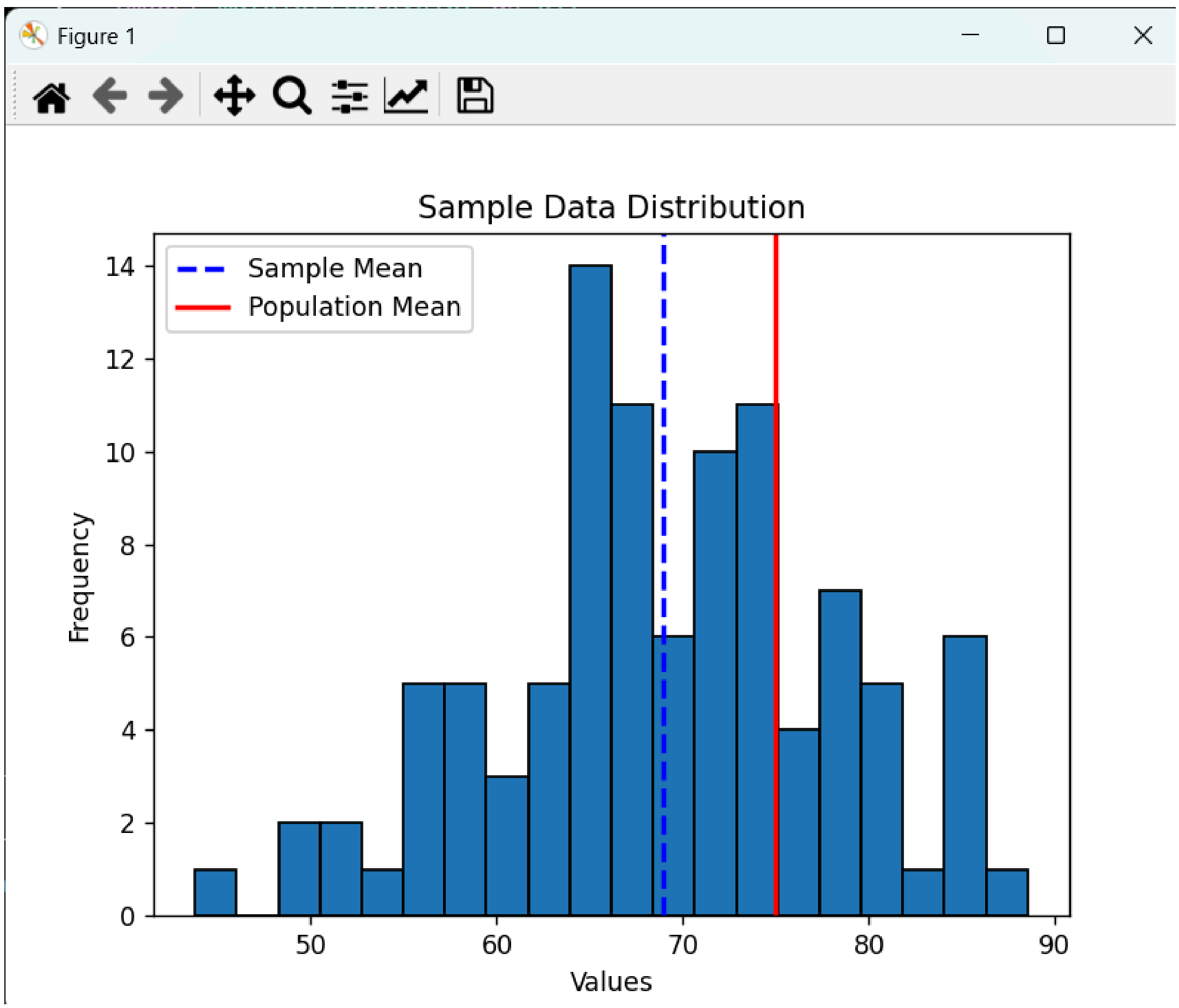


***Figure 1: Sample Data Distribution***

The plot shows the distribution of the 100 sample test scores. The dashed blue line represents the calculated mean of our sample (which is around 70), while the solid red line represents the population mean (75) we were testing against. The clear separation between the two lines visually reinforces the conclusion from the Z-test: our sample's average score is substantially lower than the population average.

### 5. Conclusion

This report successfully demonstrated the application and interpretation of a one-sample Z-test.

The key takeaways are:

1. **Clear Comparison:** The Z-test provides a robust statistical method for comparing a sample mean to a known population value.

2. **Statistically Significant Difference:** The extremely low P-value provided strong evidence to reject the null hypothesis, confirming that the difference between our sample mean and the population mean was not due to random chance.

3. **Visualization Aids Understanding:** The histogram effectively illustrated the difference, making the statistical conclusion more intuitive.

The one-sample Z-test is a powerful and straightforward tool for validating whether a sample truly represents a known population, a common task in quality control, scientific research, and data analysis.