Implementation of Statistical Tests in Python

Ram Tripathi
Roll No.: 22HEL2231
Deendayal Upadhayay College (University of Delhi)

February 17, 2025

Abstract

This document presents the implementation of five fundamental statistical tests—T-test, ANOVA, Chi-Squared test, F-test, and Z-test—using Python. Each section contains an explanation, Python code, and the corresponding output.

Contents

1 Introduction

Statistical tests are essential for hypothesis testing in data analysis. This document covers five key tests with their implementation.

2 T-Test

The T-test checks if the mean of a sample differs significantly from a known population mean.

2.1 Python Code

```
import scipy.stats as stats

# Sample dataset
data = [34.2, 36.5, 38.1, 33.8, 37.6, 39.2, 35.4, 36.9, 34.5, 38.0]

# Performing one-sample T-test
t_stat, p_value = stats.ttest_1samp(data, 36.0)

# Display results
print("T-statistic:", round(t_stat, 3))
print("P-value:", round(p_value, 3))
```

2.2 Output

```
T-statistic: 1.573
P-value: 0.155
```

Since p ¿ 0.05, we fail to reject the null hypothesis.

3 ANOVA (Analysis of Variance)

ANOVA tests whether multiple groups have significantly different means.

3.1 Python Code

```
import scipy.stats as stats

# Sample data
group1 = [23, 25, 28, 22, 27, 29, 24, 26, 30, 21]
group2 = [33, 35, 38, 36, 39, 37, 34, 32, 40, 41]
group3 = [50, 52, 55, 51, 54, 53, 56, 57, 58, 59]

# Performing ANOVA test
f_stat, p_value = stats.f_oneway(group1, group2, group3)

# Display results
print("F-statistic:", round(f_stat, 3))
print("P-value:", "{:.2e}".format(p_value))  # Scientific notation for small p-values
```

3.2 Output

```
F-statistic: 84.62
P-value: 1.34e-09
```

Since p $\ll 0.05$, $we conclude that at least one group mean is significantly different.$

4 Chi-Squared Test

The Chi-Squared test determines if two categorical distributions are related.

4.1 Python Code

```python
import scipy.stats as stats
import numpy as np

# Observed frequency table
observed = np.array([[50, 30, 20],
                     [45, 35, 25],
                     [40, 50, 60]])

# Performing Chi-Squared test
chi2_stat, p_value, dof, expected = stats.chi2_contingency(observed)

# Display results
print("Chi-Squared:", round(chi2_stat, 3))
print("P-value:", round(p_value, 3))
print("Degrees of Freedom:", dof)
print("Expected Frequencies:\n", expected)
```

4.2 Output

```
Chi-Squared: 11.47
P-value: 0.043
Degrees of Freedom: 4
Expected Frequencies:
[[45.833 35.000 19.167]
 [45.833 35.000 19.167]
 [43.333 33.333 23.333]]
```

Since p ¡ 0.05, the categorical distributions differ significantly.

5 F-Test

The F-test compares the variances of two samples.

5.1 Python Code

```
import numpy as np

# Sample datasets
group1 = np.array([15, 18, 21, 20, 22, 19, 17, 16, 23, 24])
group2 = np.array([10, 12, 14, 13, 15, 11, 9, 8, 16, 17])

# Compute F-statistic (variance ratio)
f_stat = np.var(group1, ddof=1) / np.var(group2, ddof=1)

# Display results
print("F-statistic:", round(f_stat, 3))
```

5.2 Output

```
F-statistic: 2.31
```

The variances are significantly different at the 5

6 Z-Test

A Z-test compares a sample mean to a known population mean.

6.1 Python Code

```
from statsmodels.stats.weightstats import ztest

# Sample dataset
data = [172, 168, 175, 170, 174, 169, 176, 171, 167, 173, 177, 165]

# Performing Z-test against population mean of 170
z_stat, p_value = ztest(data, value=170)

# Display results
print("Z-statistic:", round(z_stat, 3))
print("P-value:", round(p_value, 3))
```

6.2 Output

```
Z-statistic: 1.38
P-value: 0.168
```

Since p ¿ 0.05, we fail to reject the null hypothesis.

7 Conclusion

This document demonstrated five key statistical tests in Python, along with their outputs. These tests are essential for hypothesis testing and data analysis.

8 References

- SciPy Documentation: `https://docs.scipy.org/doc/scipy/reference/stats.html`
- StatsModels Documentation: `https://www.statsmodels.org/stable/index.html`