1–sample t-test

Louisa / Salomé

Understanding the 1-sample t-test

Definition: A statistical test used to determine if the mean of a single sample is significantly different from a known or hypothesized population mean.

Unknown Population Variance: The standard deviation of the population is not known.

State the Hypotheses:

- **Null Hypothesis (H_0)**: The sample mean equals the population mean.
- **Alternative Hypothesis (H_1)**: The sample mean is different from the population mean

Calculate t-statistic

- $\bar{x}$: sample mean
- μ_0: population mean
- s: sample standard deviation
- n: sample size

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

If the t value > critical t value: reject the null hypothesis
If the t value < critical t value: fail to reject the null hypothesis

Example

Example
A group of 150 students has a mean math score of 5.0 with a std of 0.5, while the population average is 4.75. We want to test, whether the class has a significantly different score compared to the population.

Null Hypothesis (H_0): The class mean equals the population mean.
Alternative Hypothesis (H_1): The class mean is different from the population mean

Define alpha and degrees of freedom

- α: 0.05
- df = n-1 = 149
- Critical t-value: 1.9760

Calculate t-value
(5.0 - 4.75) / (0.5 / √150) = **6.13**

As the t-value is bigger than the critical t-value, the null-hypothesis can be rejected.

One Tail	0.05	0.025
Two Tails	0.1	0.05
df		
1	6.3138	12.7065
2	2.9200	4.3026
3	2.3534	3.1824
4	2.1319	2.7764
5	2.0150	2.5706
6	1.9432	2.4469
7	1.8946	2.3646
8	1.8595	2.3060
9	1.8331	2.2621
10	1.8124	2.2282
144	1.6555	1.9766
145	1.6554	1.9765
146	1.6554	1.9764
147	1.6553	1.9762
148	1.6552	1.9761
149	1.6551	1.9760
150	1.6551	1.9759
151	1.6550	1.9758
152	1.6549	1.9757
153	1.6549	1.9756
154	1.6548	1.9755
155	1.6547	1.9754

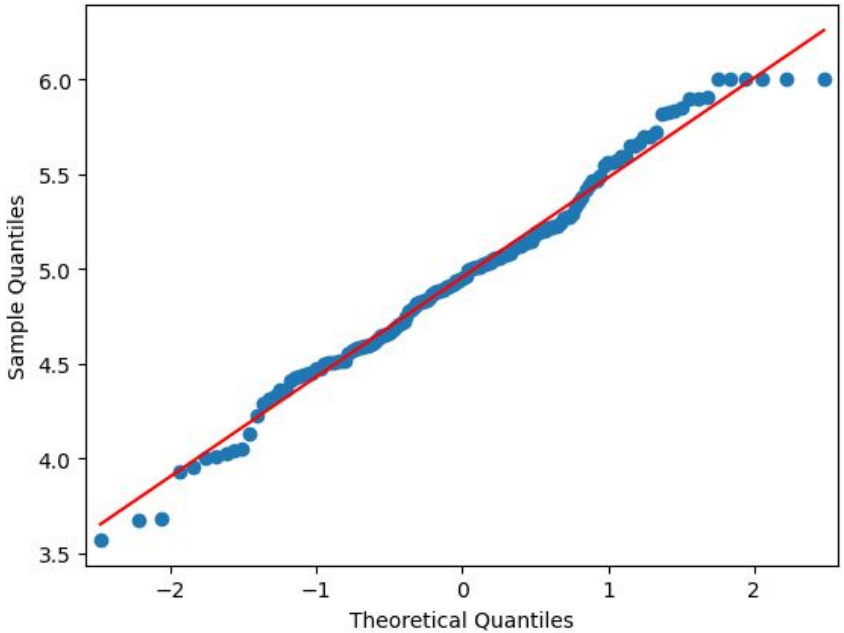
Application in python - generate dataset

```
import pandas as pd                                    # Import libraries
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
from scipy.stats import ttest_1samp                    # Set the random seed for reproducibility

np.random.seed(1729)                                   # Number of students

num_students = 150                                     # Generate math scores with mean 5.0, standard
                                                       deviation 0.5
math_scores = np.random.normal(loc=5.0,                # Ensure values stay within the range [1, 6] by
scale=0.5, size=num_students)                          clipping the values - if below or above - set
math_scores = np.clip(math_scores, 1, 6)               to limits

df = pd.DataFrame({                                    # Create a DataFrame with student names and
    'Student': [f'Student {i+1}' for i in              math scores
range(num_students)],
    'Math Score': math_scores})
print(df)                                              # Display the DataFrame
```

Check for Normality

```
# Plot qq Plot
fig=sm.qqplot(df['Math Score'], line='s')
plt.show()
```

```
We assume normality is given.
```



Perform T Test

```
population_mean = 4.75
t_statistic, p_value = ttest_1samp(df[ 'Math Score'], popmean=population_mean)

# Display the results
print("\nOne-Sample t-Test Results:" )
print(f"t-statistic: {t_statistic:.4f}")
print(f"p-value: {p_value:.4f}")

# Interpret the result
alpha = 0.05  # Significance level
if p_value < alpha:
    print("Reject the null hypothesis: The sample mean is significantly different from the population mean.")
else:
    print("Fail to reject the null hypothesis: There is no significant difference between the sample mean and the population mean." )
```

```
One-Sample t-Test Results:
t-statistic: 4.7660
p-value: 0.0000
Reject the null hypothesis: The sample mean is significantly different from the population mean.
```