Hypothesis Testing (Two Proportion Test, Two Sample t-Test)

# Two Proportion Test

## Overview

The two proportion test is used to compare the proportions of two independent groups to determine if there is a significant difference between them.

## Hypotheses

Null Hypothesis (H₀): p₁ = p₂ (The proportions are equal).  
Alternative Hypothesis (H₁): p₁ ≠ p₂ (The proportions are not equal).

## Assumptions

The samples are independent.  
The sample sizes are sufficiently large (np and n(1-p) > 5).

## Test Statistic

The test statistic for the two proportion test is a z-score, calculated as follows:  
z = (p̂₁ - p̂₂) / √[p̂(1 - p̂)(1/n₁ + 1/n₂)]  
where:  
- p̂₁ and p̂₂ are the sample proportions.  
- p̂ is the pooled sample proportion:  
p̂ = (x₁ + x₂) / (n₁ + n₂)

## Example

```python  
import statsmodels.api as sm  
  
# Example data  
count = [30, 20] # Number of successes in each group  
nobs = [100, 100] # Number of observations in each group  
  
# Perform two proportion z-test  
stat, p\_value = sm.stats.proportions\_ztest(count, nobs)  
print("Z-Statistic:", stat)  
print("P-Value:", p\_value)  
```

## Interpretation

If p-value < α (e.g., 0.05): Reject the null hypothesis, indicating a significant difference between the proportions.  
If p-value ≥ α: Fail to reject the null hypothesis, indicating no significant difference between the proportions.

# Two Sample t-Test

## Overview

The two sample t-test is used to compare the means of two independent groups to determine if there is a significant difference between them.

## Hypotheses

Null Hypothesis (H₀): μ₁ = μ₂ (The means are equal).  
Alternative Hypothesis (H₁): μ₁ ≠ μ₂ (The means are not equal).

## Assumptions

The samples are independent.  
The data in each group are approximately normally distributed.  
The variances of the two groups are equal (for a pooled t-test) or not equal (for Welch’s t-test).

## Test Statistic

The test statistic for the two sample t-test is calculated as follows:  
t = (X̄₁ - X̄₂) / √[s\_p²(1/n₁ + 1/n₂)]  
where:  
- X̄₁ and X̄₂ are the sample means.  
- s\_p² is the pooled variance:  
s\_p² = [(n₁ - 1)s₁² + (n₂ - 1)s₂²] / (n₁ + n₂ - 2)  
  
For Welch’s t-test:  
t = (X̄₁ - X̄₂) / √[(s₁²/n₁) + (s₂²/n₂)]

## Example

```python  
from scipy import stats  
  
# Example data  
group1 = [23, 25, 27, 22, 24]  
group2 = [35, 37, 36, 38, 39]  
  
# Perform two sample t-test (assume equal variance)  
t\_stat, p\_value = stats.ttest\_ind(group1, group2, equal\_var=True)  
print("T-Statistic:", t\_stat)  
print("P-Value:", p\_value)  
  
# Perform Welch's t-test (assume unequal variance)  
t\_stat, p\_value = stats.ttest\_ind(group1, group2, equal\_var=False)  
print("Welch's T-Statistic:", t\_stat)  
print("Welch's P-Value:", p\_value)  
```

## Interpretation

If p-value < α (e.g., 0.05): Reject the null hypothesis, indicating a significant difference between the means.  
If p-value ≥ α: Fail to reject the null hypothesis, indicating no significant difference between the means.

## Applications

Comparing the effectiveness of two treatments.  
Comparing the performance of two groups in an experiment.  
Evaluating differences in means across different populations.