

Hypothesis Testing

Heads up

1. The name of the test.
2. What the test is checking.
3. The key assumptions of the test.
4. How the test result is interpreted.
5. Python API for using the test.

1. **Normality Tests**

Shapiro-Wilk Test

D'Agostino's K^2 Test

Anderson-Darling Test

2. **Correlation Tests**

Pearson's Correlation Coefficient

Spearman's Rank Correlation

Kendall's Rank Correlation

Chi-Squared Test

3. **Parametric Statistical Hypothesis Tests**

Student's t-test

Paired Student's t-test

Analysis of Variance Test (ANOVA)

4. **Nonparametric Statistical Hypothesis Tests**

Mann-Whitney U Test

Wilcoxon Signed-Rank Test

Kruskal-Wallis H Test



Shapiro-Wilk Test

Tests whether a data sample has a Gaussian distribution.

Assumptions

- Observations in each sample are independent and identically distributed (iid).

Interpretation

- H_0 : the sample has a Gaussian distribution.
- H_1 : the sample does not have a Gaussian distribution.



D'Agostino's K^2 Test

Tests whether a data sample has a Gaussian distribution.

Assumptions

- Observations in each sample are independent and identically distributed (iid).

Interpretation

- H_0 : the sample has a Gaussian distribution.
- H_1 : the sample does not have a Gaussian distribution.



Anderson-Darling Test

Tests whether a data sample has a Gaussian distribution.

Assumptions

- Observations in each sample are independent and identically distributed (iid).

Interpretation

- H_0 : the sample has a Gaussian distribution.
- H_1 : the sample does not have a Gaussian distribution.



Pearson's Correlation Coefficient

Tests whether two samples have a linear relationship.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample are normally distributed.
- Observations in each sample have the same variance.

Interpretation

- H_0 : the two samples are independent.
- H_1 : there is a dependency between the samples.

Spearman's Rank Correlation

Tests whether two samples have a monotonic relationship.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample can be ranked.

Interpretation

- H_0 : the two samples are independent.
- H_1 : there is a dependency between the samples.



Kendall's Rank Correlation

Tests whether two samples have a monotonic relationship.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample can be ranked.

Interpretation

- H_0 : the two samples are independent.
- H_1 : there is a dependency between the samples.

Chi-Squared Test

Tests whether two categorical variables are related or independent.

Assumptions

- Observations used in the calculation of the contingency table are independent.
- 25 or more examples in each cell of the contingency table.

Interpretation

- H_0 : the two samples are independent.
- H_1 : there is a dependency between the samples.



Student's t-test

Tests whether the means of two independent samples are significantly different.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample are normally distributed.
- Observations in each sample have the same variance.

Interpretation

- H_0 : the means of the samples are equal.
- H_1 : the means of the samples are unequal.

Paired Student's t-test

Tests whether the means of two paired samples are significantly different.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample are normally distributed.
- Observations in each sample have the same variance.
- Observations across each sample are paired.

Interpretation

- H_0 : the means of the samples are equal.
- H_1 : the means of the samples are unequal.

Analysis of Variance Test (ANOVA)

Tests whether the means of two or more independent samples are significantly different.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample are normally distributed.
- Observations in each sample have the same variance.

Interpretation

- H_0 : the means of the samples are equal.
- H_1 : one or more of the means of the samples are unequal.



Mann-Whitney U Test

Tests whether the distributions of two independent samples are equal or not.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample can be ranked.

Interpretation

- H_0 : the distributions of both samples are equal.
- H_1 : the distributions of both samples are not equal.



Wilcoxon Signed-Rank Test

Tests whether the distributions of two paired samples are equal or not.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample can be ranked.
- Observations across each sample are paired.

Interpretation

- H_0 : the distributions of both samples are equal.
- H_1 : the distributions of both samples are not equal.



Kruskal-Wallis H Test

Tests whether the distributions of two or more independent samples are equal or not.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample can be ranked.

Interpretation

- H_0 : the distributions of all samples are equal.
- H_1 : the distributions of one or more samples are not equal.

