

Dans cette partie, je vais vous donner des petits exemples d'exécution des tests statistiques que l'on utilise le plus souvent.

On va utiliser la librairie `scipy` qui contient toutes les fonctions d'exécution des tests statistiques.

## Test de Chi<sup>2</sup>

Tests si deux catégories de variables sont indépendantes ou non.

Assumptions

- Observations sont indépendantes.
- 25 or more examples in each cell of the contingency table.

Interpretation

- H0: Les deux échantillons sont indépendants.
- H1: Il y a une dépendance entre les deux échantillons.

In [1]:

```
# Example of the Chi-Squared Test
from scipy.stats import chi2_contingency
table = [[10, 20, 30],[6, 9, 17]]
stat, p, dof, expected = chi2_contingency(table)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably independent')
else:
    print('Probably dependent')
```

```
stat=0.272, p=0.873
Probably independent
```

## Student T test

Tests si la moyenne de deux échantillons indépendants sont significativement différent.

Assumptions

- Les observations dans chaque groupe sont indépendants et ont une distribution identiques.
- Les observations dans chaque groupe ont une distribution normale.
- Les observations dans chaque groupe ont la même variance.

Interpretation

- H0: La moyenne des échantillons est égale.
- H1: La moyenne des échantillons est différente.

In [2]:

```
# Example of the Student's t-test
from scipy.stats import ttest_ind
data1 = [0.873, 2.817, 0.121, -0.945, -0.055, -1.436, 0.360, -1.478, -1.637, -1.869]
data2 = [1.142, -0.432, -0.938, -0.729, -0.846, -0.157, 0.500, 1.183, -1.075, -0.169]
stat, p = ttest_ind(data1, data2)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably the same distribution')
else:
    print('Probably different distributions')
```

stat=-0.326, p=0.748  
Probably the same distribution

## T-test de Student pairé

Tests si la moyenne de deux échantillons pairés sont significativement différent.

Assumptions

- Les observations dans chaque groupe sont indépendents et ont une distribution identiques.
- Les observations dans chaque groupe ont une distribution normale.
- Les observations dans chaque groupe ont la même variance.
- Les observations entre deux échantillons sont pairées.

Interpretation

- H0: La moyenne des échantillons est égale.
- H1: La moyenne des échantillons est différente.

In [3]:

```
# Example of the Paired Student's t-test
from scipy.stats import ttest_rel
data1 = [0.873, 2.817, 0.121, -0.945, -0.055, -1.436, 0.360, -1.478, -1.637, -1.869]
data2 = [1.142, -0.432, -0.938, -0.729, -0.846, -0.157, 0.500, 1.183, -1.075, -0.169]
stat, p = ttest_rel(data1, data2)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably the same distribution')
else:
    print('Probably different distributions')
```

stat=-0.334, p=0.746  
Probably the same distribution

## Anova

Test si la moyenne de deux ou plus échantillons indépendants sont significativement différent.

Assumptions

- Les observations dans chaque groupe sont indépendents et ont une distribution identiques.

- Les observations dans chaque groupe ont une distribution normale.
- Les observations dans chaque groupe ont la même variance.

#### Interpretation

- H0: La moyenne des échantillons est égale.
- H1: La moyenne des échantillons est différente.

In [4]:

```
# Example of the Analysis of Variance Test
from scipy.stats import f_oneway
data1 = [0.873, 2.817, 0.121, -0.945, -0.055, -1.436, 0.360, -1.478, -1.637, -1.869]
data2 = [1.142, -0.432, -0.938, -0.729, -0.846, -0.157, 0.500, 1.183, -1.075, -0.169]
data3 = [-0.208, 0.696, 0.928, -1.148, -0.213, 0.229, 0.137, 0.269, -0.870, -1.204]
stat, p = f_oneway(data1, data2, data3)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably the same distribution')
else:
    print('Probably different distributions')
```

stat=0.096, p=0.908  
Probably the same distribution

## Test Mann-Whitney

Tests si la distribution de deux échantillons indépendants est égale ou non.

#### Assumptions

- Les observations dans chaque groupe sont indépendantes.
- Les observations dans chaque groupe peuvent être classées.

#### Interpretation

- H0: La distribution est égale.
- H1: La distribution est différente.

In [ ]:

```
# Example of the Mann-Whitney U Test
from scipy.stats import mannwhitneyu
data1 = [0.873, 2.817, 0.121, -0.945, -0.055, -1.436, 0.360, -1.478, -1.637, -1.869]
data2 = [1.142, -0.432, -0.938, -0.729, -0.846, -0.157, 0.500, 1.183, -1.075, -0.169]
stat, p = mannwhitneyu(data1, data2)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably the same distribution')
else:
    print('Probably different distributions')
```

## Test de Wilcoxon

Tests si la distribution de deux échantillons appariés est égale ou non.

## Assumptions

- Les observations dans chaque groupe sont indépendant.
- Les observations dans chaque groupe peuvent être classées.
- Les observations dans entre deux échantillons sont pairées.

## Interpretation

- H0: La distribution est égale.
- H1: La distribution est différente.

In [6]:

```
# Example of the Wilcoxon Signed-Rank Test
from scipy.stats import wilcoxon
data1 = [0.873, 2.817, 0.121, -0.945, -0.055, -1.436, 0.360, -1.478, -1.637, -1.869]
data2 = [1.142, -0.432, -0.938, -0.729, -0.846, -0.157, 0.500, 1.183, -1.075, -0.169]
stat, p = wilcoxon(data1, data2)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably the same distribution')
else:
    print('Probably different distributions')
```

stat=21.000, p=0.508  
Probably the same distribution

# Test de Kruskal-Wallis

Tests si la distribution de deux ou plus échantillons indépendants est égale ou non.

## Assumptions

- Les observations dans chaque groupe sont indépendant.
- Les observations dans chaque groupe peuvent être classées.

## Interpretation

- H0: La distribution de tout les groupes est égale.
- H1: La distribution d'un ou plusieurs groupe est différente.

In [7]:

```
# Example of the Kruskal-Wallis H Test
from scipy.stats import kruskal
data1 = [0.873, 2.817, 0.121, -0.945, -0.055, -1.436, 0.360, -1.478, -1.637, -1.869]
data2 = [1.142, -0.432, -0.938, -0.729, -0.846, -0.157, 0.500, 1.183, -1.075, -0.169]
stat, p = kruskal(data1, data2)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably the same distribution')
else:
    print('Probably different distributions')
```

stat=0.571, p=0.450  
Probably the same distribution

# Test de Friedman

Tests si la distribution de deux ou plus échantillons dépendants est égale ou non.

## Assumptions

- Les observations dans chaque groupe sont indépendant.
- Les observations dans chaque groupe peuvent être classées.
- Les observations sont pairés entre les groupes

## Interpretation

- H0: La distribution de tout les groupes est égale.
- H1: La distribution d'un ou plusieurs groupe est différente.

In [5]:

```
# Example of the Friedman Test
from scipy.stats import friedmanchisquare
data1 = [0.873, 2.817, 0.121, -0.945, -0.055, -1.436, 0.360, -1.478, -1.637, -1.869]
data2 = [1.142, -0.432, -0.938, -0.729, -0.846, -0.157, 0.500, 1.183, -1.075, -0.169]
data3 = [-0.208, 0.696, 0.928, -1.148, -0.213, 0.229, 0.137, 0.269, -0.870, -1.204]
stat, p = friedmanchisquare(data1, data2, data3)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably the same distribution')
else:
    print('Probably different distributions')
```

stat=0.800, p=0.670

Probably the same distribution