Tests paramétriques et non-paramétriques

Table des matières

[2. Test d’ajustement de deux distributions 1](#_Toc497731294)

[1. Définitions 1](#_Toc497731295)

[1. Répartition expérimentale : 1](#_Toc497731296)

[2. Répartitions théoriques : 1](#_Toc497731297)

[3. Ecart entre les deux distributions : 1](#_Toc497731298)

[4. Test de Pearson (Test d’hypothèse page 19) 1](#_Toc497731299)

[3. Test d’appartenance à la même population 1](#_Toc497731300)

[1. Tests Paramétriques 1](#_Toc497731301)

[1. Comparaison de deux moyennes d’échantillons : « test T » 1](#_Toc497731302)

[2. Comparaison de deux variances d’échantillons : « test F » 2](#_Toc497731303)

[3. Comparaison de deux distributions d’échantillons 3](#_Toc497731304)

[2. Tests non-paramétriques 3](#_Toc497731305)

[1. Comparaison de deux médianes d’échantillons : « test Kruskal-Wallis » 3](#_Toc497731306)

[2. Comparaison de moyennes et variances de deux échantillons : « test Mann-Whitney » 4](#_Toc497731307)

[3. Test de normalité : 5](#_Toc497731308)

# Test d’ajustement de deux distributions

Objectif : Déterminer si une distribution expérimentale correspond à une distribution théorique. Ce test permet d’identifier si une distribution est normale pour être étudier par des tests comme celui d’appartenance de deux populations à une seule et même distribution. Pour cela nous effectuerons un test du Khi deux entre une population observées et une population théorique.

Remarque : Une hypothèse tel que « les deux distributions proviennent de la même population » ne peut être que rejetée. Il n’est possible que de conclure d’une possible véracité de l’hypothèse (ex : l’explication sur [scipy](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html) et sur [stakexchange](https://stats.stackexchange.com/questions/83163/statistical-test-to-tell-whether-two-samples-are-pulled-from-the-same-population)).

## Définitions

### Répartition expérimentale :

On répartit les observations suivant k classes (si le caractère est continu) ou k valeurs (si le caractère est discret). On dispose alors des effectifs des k classes : n1, n2,....., nk. Cela, se concrétise par exemple dans le cas continu à réfléchir par intervalle de valeur. On a bien sûr la relation :

### Répartitions théoriques :

En se basant sur la base expérimentale nous définirons une loi théorique ayant la même moyenne et même variance dans le cas d’une loi normal et définirons les probabilités théoriques pour chaque intervalle défini expérimentalement et définirons des effectifs théoriques

### Ecart entre les deux distributions :

Où k est le nombre d’intervalle et r le nombre de paramètres estimés pour la distribution théorique.

### Test de Pearson ([Test d’hypothèse page 19](https://www.math.u-psud.fr/~pansu/web_ifips/Tests.pdf))

1ère Etape :

On définit les hypothèses :

2ème Etape :

On mesure comme précisé dans la partie précédente l’écart entre les deux distributions.

3ème Etape :

On détermine la borne supérieure autorisé à un risque α pour l’écart mesuré par le Khi-deux, la borne inférieure étant 0. Avec le degré de liberté et le risque on détermine :

4ème Etape :

* : est acceptée
* : est refusée ( est acceptée)

# Test d’appartenance à la même population

Objectif : vérifier l’hypothèse H1 pour chaque type d’alerte et chaque prix du MA.

Pour répondre à cet objectif il est nécessaire d’étudier les distributions continues de variables (ex : Prix des écarts positifs) en fonction de variables discrètes (ex : Signalement de mode dégradé par RTE).

## Tests Paramétriques

Remarque : nous faisons une hypothèse de distribution normale basée sur l’observation faites dans la parties description descriptive. En se basant sur Le document « [Test d’hypothèse](https://www.math.u-psud.fr/~pansu/web_ifips/Tests.pdf)» pour les explications complètes et le site de statistique [suivant](http://www.cons-dev.org/elearning/stat/parametrique/5-3/5-3.html) nous pourrons dans un premier temps faire une hypothèse de normalité des distributions puis si la cardinalité d’une population est jugé trop faible faire un test de Spearman (non paramétrique).

### Comparaison de deux moyennes d’échantillons : « test T »

* Cas 1 : nous sommes en situation d’alerte avec une moyenne .
* Cas 2 : Nous ne sommes pas en situation d’alerte avec une moyenne .
* Cardinalités : et
* Hypothèse implicite du test : loi bi-normale

1ère Etape :

Nous allons tester l’hypothèse , l’égalité des moyennes des deux populations contre l’hypothèse . Nous avons donc les définitions suivantes des hypothèses :

2ème Etape :

La statistique qui nous intéresse ici est la différence

En se plaçant sous l’hypothèse nous avons la loi de probabilité suivante :

La fonction discriminante est donc  en posant :

3ème Etape :

Nous recherchons l’écart autorisé par rapport à la moyenne de notre fonction discriminante, en se basant sur le risque α, tel que :

4ème Etape :

On mesure . En se basant sur les étapes précédentes et cette mesure, on peut valider ou invalider  :

* : est acceptée
* : est refusée ( est acceptée)

### Comparaison de deux variances d’échantillons : « test F »

* Cas 1 : nous sommes en situation d’alerte avec un écart-type .
* Cas 2 : Nous ne sommes pas en situation d’alerte avec un écart-type .
* Hypothèse implicite du test : loi bi-normale

1ère Etape :

Nous allons tester l’hypothèse , l’égalité des écart-types des deux populations contre l’hypothèse . Nous avons donc les définitions suivantes des hypothèses :

2ème Etape :

En se basant sur les observations descriptives faites plus tôt o peut en déduire le rapport

En utilisant l’hypothèse de normalité du test nous posons que

* suis une loi du khi-deux à degré(s) de liberté.
* suis une loi du khi-deux à degré(s) de liberté.

On considère alors le rapport

Suivant la loi de Fisher avec et degrés de liberté. (Rappel : sous H0 )

3ème Etape :

Nous recherchons l’écart autorisé par rapport à la moyenne de notre fonction discriminante, en se basant sur le risque α, tel que :

Remarque :

4ème Etape :

On mesure . En se basant sur les étapes précédentes et cette mesure, on peut volider ou invalider  :

* : est acceptée
* : est refusée ( est acceptée)

### Comparaison de deux distributions d’échantillons

Pour cela nous utiliserons les deux tests précédemment rappelés et nous prendront comme hypothèse s’appliquant à tous les échantillons l’union des hypothèses nécessaires aux « tests T et F ».

* Cas 1 : nous sommes en situation d’alerte avec une moyenne .
* Cas 2 : Nous ne sommes pas en situation d’alerte avec une moyenne .
* Cardinalités : et
* Hypothèse implicite du test : loi bi-normale

1ère Etape :

Nous définirons l’hypothèse que deux distributions sont comparables par l’hypothèse et sont opposée par :

2ème Etape :

Nous effectuerons donc les « tests T et F » pour les couples de donnés continues (quantitatif) catégorisé par une propriété discrète (qualitatif).

3ème Etape :

L’hypothèse sera acceptée uniquement si les deux tests sont positifs et rejetée dans tout autre cas.

## Tests non-paramétriques

Remarque : Utilisation du rang comme donnée fondamentale pour faire des tests non-paramétriques.

### Comparaison de deux médianes d’échantillons : « [test Kruskal-Wallis](http://www.itl.nist.gov/div898/handbook/prc/section4/prc41.htm)»

Objectif : La procédure teste l’hypothèse nulle que k échantillons possiblement de différentes populations proviennent en fait d’une seule et même population. Ce test est basé sur une tendance générale représentée par la médiane.

1ère Etape :

Nous allons tester l’hypothèse nulle , les k échantillons proviennent de la même population contre l’hypothèse . Nous avons donc les définitions suivantes des hypothèses :

2ème Etape :

Pour chaque valeur de chaque échantillon on crée un triplet de coordonnées représentant :

* La valeur numérique
* Le numéro d’échantillon d’appartenance
* Le classement de la valeur dans l’union des k échantillons entre 1 et N.

Il est possible ensuite pour chaque échantillon parmi les k, d’obtenir une somme des rangs de l’ensemble constitué par l’union et de sommer pour chaque échantillon les rangs des valeurs. On obtient une somme de rangs, .

La fonction discriminante, H, propre au test de Kruskal suis une loi du Khi-deux comme défini précédemment.

3ème Etape :

On détermine la borne supérieure autorisé à un risque α pour l’écart mesuré par le Khi-deux, la borne inférieure étant 0. Avec le degré de liberté et le risque on détermine :

4ème Etape :

* : est acceptée
* : est refusée ( est acceptée)

### Comparaison de moyennes et variances de deux échantillons : « [test Mann-Whitney](https://jonathanlenoir.files.wordpress.com/2013/12/tests-de-comparaison-de-moyennes-non-param.pdf) »

Objectif : L’objectif est de comparer la distribution de deux échantillons non appariés en prenant en hypothèse l’indépendance des données dans chaque échantillon et des deux échantillons. Ce test prend en compte une variance homogène entre les deux et un décalage des moyennes. La procédure teste l’hypothèse nulle que k échantillons possiblement de différentes populations proviennent en fait d’une seule et même population.

Remarque : Ce test s’effectue seulement sur deux populations, n=2n à comparer et non k comme le test de Kruskal-Wallis. Quand n>2 on utilise le test d’hypothèse suivant :

1ère Etape :

Nous allons tester l’hypothèse nulle , les 2 échantillons proviennent de la même population contre l’hypothèse . Nous avons donc les définitions suivantes des hypothèses :

2ème Etape :

La fonction discriminante est basée sur les rangs des valeurs après union des deux échantillons. Une comparaison 2 à 2 est faits entre les xi et xj:

D’où :

On peut montrer que , d’où :

3ème Etape :

Après correction de la variable, l’analyse porte sur une loi normale centrée réduite. On peut définir la borne d’un risque de niveau alpha par :

4ème Etape :

* : est acceptée
* : est refusée ( est acceptée)

### Test de normalité :

Cette classification est la plus fine que l’on peut atteindre avec les données choisies. On utilise le test de normalité de scipy, décrit sur le [papier](http://webspace.ship.edu/pgmarr/geo441/readings/d'agostino%201971%20-%20an%20omnibus%20test%20of%20normality%20for%20moderate%20and%20large%20size%20samples.pdf) d’Agostino, « An Omnibus Test of Normality for Moderate and Large Size Samples ». Il fait l’hypothèse d’échantillon de 50 observations et se base sur l’étude de la déviation de la différence entre la distribution observée et une distribution normal déduite. Ce test est basé sur celui de Shapiro et Wilk, le « test W ».