

# Outils Statistiques pour la data intelligence (Module 1)

---

## Introduction

---

On étudie un caractère (discret ou continu) sur une population. Un paramètre (moy., prop., var.) est inconnu ou la loi de proba de ce caractère est inconnue.

On formule ensuite des hypothèses statistiques sur la valeur ou la loi de distribution de ce paramètre.

On posera ensuite un jugement sur cette hypothèse sur base du résultat(s) obtenus sur un échantillon. Il faut que la différence entre les résultats théoriques et observées ne soient pas trop grands. La méthode qui consiste à vérifier que la différence entre ces 2 résultats n'est pas trop grande est appelée le **test d'hypothèse**

## Hypothèse Nulle, Hypothèse alternative, seuil de signification

---

### Hypothèses

L'**hypothèse nulle** ( $H_0$ ) est l'hypothèse à tester. Elle s'exprime sous la forme d'une **égalité** (=) pour les tests de valeur et d'égalité et sous la forme d'une affirmation positive pour les tests du khi-deux.

L'**hypothèse alternative** ( $H_1$ ) est celle que nous accepterons si nous sommes amenées à rejeter  $H_0$ . Elle s'exprime sous la forme :

- d'une ( $\neq$ ) → **tests bilatéraux**
- d'une ( $<, >$ ) → **tests unilatéraux**
- d'une affirmation négative pour les tests du khi-deux

Il existe des hypothèses nulles qui utilisent  $\leq$  et  $\geq$ . L'hypothèse alternative est alors automatiquement l'inverse de l'hypothèse nulle

### Test Unilatéral Vs Test Bilatéral

Le choix entre un **test unilatéral** ou **bilatéral** dépend souvent de qui demande le test. En effet, passer de l'un à l'autre peut avoir pour effet de changer la conclusion du test

### Seuil de signification

Le **seuil de signification** est la probabilité de rejeter  $H_0$  ( $H_1$  est vrai à tort) alors que  $H_0$  est vrai. On appelle cette erreur **erreur ou risque de 1<sup>ère</sup> espèce**, noté  $\alpha$

note : le risque de 2<sup>e</sup> espèce ou  $\beta$  serait le risque d'accepter  $H_0$  alors qu'il est faux.

### Démarche

1. Fixer les hypothèses  $H_0$  et  $H_1$
2. Préciser le seuil de signification de  $\alpha$

3. Préciser la loi de probabilité utilisée
4. Déterminer le seuil de rejet, la zone de non refus
5. Calculer la grandeur expérimentale
6. Tirer des conclusions

## Tests de valeur

---

Dans un **test de valeur**, on teste l'hypothèse selon laquelle le paramètre étudié de la population est égale à une valeur déterminée. L'hypothèse alternative sera alors que le paramètre étudié est différent( $\neq$ ,  $<$ ,  $>$ ) de la valeur déterminée.

## Tests d'égalité

---

## Tests du khi-carré

---