Les tests statistiques servent à vérifier si les données obtenues sont compatibles avec une propriété (par exemple, une moyenne théorique) et les différences observées dû au hasard de l'échantillonnage.

* H0 la différence vient de l'échantillonnage.
* H1 différence entre l'échantillon et la population.

|  |  |  |
| --- | --- | --- |
| Choix Réalité | H0 | H1 |
| H0 |  |  |
| H1 |  | (puissance du test) |

Interprétation : Au risque est accepté H1 alors que c’est faux. Cela correspond au rejet à tort de l’hypothèse nulle.

Rmq : le risque est généralement fixé à 5%.

Règle de décision :

* L’appartenance à intervalle de confiance.
* Par des statistiques du test soit car le seuil soit par la p-value (par comparaison du risque).

Deux types de tests :

|  |  |
| --- | --- |
| Unilatéral (à gauche) | Bilatéral |
| H0 : µ = µ0  Ha : µ < µ0 | H0 : µ = µ0  Ha : µ ≠ µ0 |

## Intervalle de confiance

Exemple de comparaison d’une moyenne observée avec une moyenne de référence :

avec la moyenne théorique.

Intervalle de confiance

Pour comparer la valeur de deux échantillons, il suffit de vérifier si les intervalles se chevauchent.

## Les tests

|  |  |  |
| --- | --- | --- |
| Hypothèse | Seuil | P-value |
| On ne peut pas rejeter H0 au risque  inconnu |  |  |
| On accepte H1 au risque |  |  |

Par exemple, au risque alpha de 5%,

p-valeur est la probabilité d’obtenir une valeur aussi extrême sous l’hypothèse H0. Énoncer de façon différente, elle indique dans quelle mesure les données sont conformes à l’H0.

Exemple : on effectue des tirages de pile ou face. H0 la pièce n’est pas biaisée.

* 4 tirages et 4 piles : % on ne rejette pas H0.
* 5 tirages et 5 piles : % on rejette H0.

Rmq : une valeur est significativement différente si l’hypothèse H0 (les valeurs sont égales) est rejetée.

Les types de tests principaux :

* Indépendance, exemple : la couleur des cheveux est-elle indépendante du sexe ?
* Homogénéité : deux séries de données sont-elles identiquement distribuées.
* Adéquation à une loi ou une famille de lois définies à priori, par exemple : la taille d’une population suit-elle une loi normale ?

# Test d’homogénéité

|  |  |
| --- | --- |
| Test | Formule |
| Moyenne observée et moyenne théorique |  |
| Distribution de deux échantillons (test Student) |  |
| Un échantillon et une loi de probabilité | Test du Xhi2 |

# Corrélation entre deux variables

* H0 : les variables sont indépendantes
* H1 : les variables sont liées.

## Une variable quantitative et une qualitative (ANOVA)

L’ANalysis Of Variance (ANOVA) sert à savoir si une variable qualitative à une influence sur une variable quantitative.

* H0 : La moyenne des groupes est issue d’une même population.
* H1 : Les moyennes possèdent des différences significatives entre les groupes.

Degré de liberté :

MANOVA : dans le cas où il existe plusieurs variables quantitatives à expliquer on utilise MANOVA (Multivariate analysis of variance).

Avec :

* une constante
* l’erreur qui suit une loi normale

Rmq : Lorsqu’il n’y a que deux modalités, il est possible d’utiliser un test de student.

## Deux variables qualitatives

Degré de liberté :

Il faut faire :

1. Tableau de contingence (appelé aussi effectif)
2. Tableau des effectifs théoriques : ou directement
3. Calcul du khi2 théorique pour chaque croissement de modalité : .
4. Somme des valeurs pour chaque modalité et calcul de la statistique :
   1. LOI.KHIDEUX.INVERSE.DROITE(proba; degré de liberté)
   2. P-value : LOI.KHIDEUX.DROITE(valeur; degré de liberté)

## Deux variables quantitatives

Degré de liberté :

1. Calcul du coefficient de corrélation de Pearson PEARSON()
2. Calcul de la statistique
   1. :
   2. P-value : LOI.STUDENT.BILATERALE()

|  |  |
| --- | --- |
| ECARTYPE.P | Écart type |
| ECARTYPE | Écart type corrigé |