

## TP5 : Test statistique

*Objectifs : Comprendre la notion de test, d'hypothèses nulle et alternative, de p-valeur.*

Un test statistique est une décision prise à partir de l'information disponible dans les données. Plusieurs étapes sont nécessaires pour mettre en place un test :

1. Modéliser la variable d'intérêt avec une variable aléatoire appropriée.
2. Définir l'hypothèse nulle et alternative, notées  $H_0$  et  $H_1$ .
3. Construire une règle de décision pour rejeter  $H_0$  ou non. La règle de décision est basée sur un estimateur bien choisi et sur sa distribution sous  $H_0$ . Aucune observation n'est nécessaire pour construire la règle.
4. Appliquer la règle aux données et décider si  $H_0$  est rejetée ou non.

Lorsqu'on applique un test statistique, on peut réaliser deux types d'erreur :

Vérité	décision du test	
	$H_0$	$H_1$
$H_0$	OK	erreur de type I Proba = $\alpha$
$H_1$	erreur de type II Proba = $\beta$	OK Proba = $1 - \beta$ = puissance

Un test n'est pas symétrique. L'erreur de type I est la plus grave, c'est celle qu'on cherche à contrôler (le risque  $\alpha$ ).

### Exercice 1

Pour un adulte, le logarithme de la concentration de D-dimer est indicateur du risque de thrombose. Il est modélisé par une variable normale de loi  $\mathcal{N}(\mu, \sigma^2)$ . L'écart-type est supposé connu et égal à 0.3. Pour un individu sain (sans thrombose),  $\mu = -1$ . Pour un individu à risque,  $\mu = 0$ .

1. Dr House ne veut pas inquiéter ses patients pour rien. Quelles hypothèses  $H_0$  et  $H_1$  choisit-il pour son test ?
2. Donner la règle de décision du test qui décide, avec un risque de 5% entre  $H_0$  et  $H_1$ . Calculer le seuil avec R. On utilisera la fonction `qnorm` pour trouver la zone de rejet exprimée soit sur l'estimateur soit sur la statistique de test (estimateur centré et réduit sous  $\mathcal{H}_0$ ) :  

```
mu <- -1; sig <- 0.3  
qnorm(0.95)-> u
```
3. Calculer la zone de rejet pour un risque de 1%. Comparer avec celle obtenue au risque de 5%.
4. Un patient à un logarithme de concentration de D-dimer égal à  $-0.46$ . Donner la décision du test en utilisant la fonction `pnorm` (on cherchera la plus petite valeur de risque de première espèce pour laquelle on conclut que le patient est un "patient à risque de thrombose"). Est ce que le patient doit être considéré à risque ?

5. Dr Cuddy a un autre point de vue : elle s'inquiète qu'un patient à risque ne soit pas prévenu de son risque (elle préfère limiter le risque de déclarer un patient sain à tort quand House préfère à l'inverse limiter le risque de déclarer un patient "à risque pour la thrombose" à tort). Quelles hypothèses  $H_0$  et  $H_1$  choisit-elle pour son test ?
6. Donner la règle de décision du test qui décide, avec un risque de 5% et de 1 % entre  $H_0$  et  $H_1$ . On utilisera la fonction `qnorm` pour trouver la zone de rejet.
7. Selon le niveau du risque, indiquer pour quelle valeur Drs House et Cuddy sont d'accord.

## Exercice 2

Un taux de paracétamol plus grand que 150 mg par kilogramme est considéré comme dangereux. La mesure du taux de paracétamol par prélèvement sanguin est modélisée par une variable aléatoire gaussienne de loi  $\mathcal{N}(\mu, \sigma^2)$ . L'écart-type est supposé connu et égal à 5 mg.

1. Donner les hypothèses et la règle de décision du test qui décide, avec un seuil de 5% si un patient est à risque à partir de son prélèvement sanguin (vous êtes un médecin prudent).
2. Un patient arrive à l'hôpital avec une suspicion d'empoisonnement au paracétamol. Un prélèvement sanguin est réalisé et une concentration de 140 mg est mesurée. Donner la décision du test. Est ce que le patient doit être considéré à risque ?

## Exercice 3

Une ligne de production produit des paquets de 1kg de farine. Le poids réel d'un paquet est modélisé par une loi normale d'écart type 20g.

1. Le responsable de production ne veut pas produire des paquets dont le poids est trop éloigné de la valeur de 1kg. Quelles hypothèses  $H_0$  et  $H_1$  choisit-il ? Définir la zone de rejet et la règle de décision au seuil de 5% et 1%.
2. Le responsable de l'usine pense que les paquets produits sont trop lourds, et qu'il perd de l'argent. Quelles hypothèses  $H_0$  et  $H_1$  choisit le responsable de production pour lui répondre ? Définir la zone de rejet et la règle de décision au seuil de 5% et 1%.
3. Un paquet pèse 1018 grammes. Quelle décision prenez vous pour les deux tests ?
4. Une association de consommateur veut poursuivre la compagnie pour vente de paquets trop légers. Quelles hypothèses  $H_0$  et  $H_1$  choisit le responsable de production pour leur répondre ? Définir la zone de rejet et la règle de décision au seuil de 5% et 1%.
5. Un paquet pèse 982 grammes. Quelle décision prenez vous avec le test précédent ?

## Exercice 4

On mesure un indice de pollution près d'une usine chimique. On suppose que cet indice est modélisé par une variable aléatoire gaussienne de loi  $\mathcal{N}(\mu, \sigma^2)$ . L'écart-type est supposé connu et égal à 4. L'état fixe un seuil maximal de 30 pour cet indice de pollution.

1. Le responsable de l'usine veut montrer que l'usine respecte la loi et le seuil maximal. Quelles hypothèses  $H_0$  et  $H_1$  proposez vous ? Donner la règle de décision du test qui décide, avec un seuil de 5% et 1 %.
2. Une association écologiste veut montrer que la pollution dépasse le seuil. Quelles hypothèses  $H_0$  et  $H_1$  proposez vous ? Donner la règle de décision du test qui décide, avec un seuil de 5% et 1 %.