Semaine 11
Hypothèses et tests statistiques

Matière aujourd'hui

1. Hypothèse nulle et hypothèse alternative
2. Erreurs de 1ère et de 2e espèce
3. Seuil du test et valeur p
4. Statistique de test

Hypothèse nulle et hypothèse alternative

Hypothèse nulle et hypothèse alternative

H_0 = hypothèse nulle = hypothèse qu'on tente de rejeter
H_1 = hypothèse alternative = hypothèse qu'on veut prouver

Ex. On s'intéresse à la proportion de gauchers et de droitier dans la population.
H_0 = il y a autant de gauchers que de droiters
H_1 = il n'y a pas autant de gauchers que de droitiers

Hypothèse nulle et hypothèse alternative

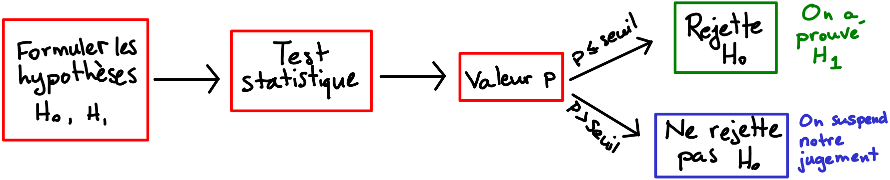
Principe

On ne peut pas montrer que l'hypothèse nulle est vraie. On peut seulement montrer l'hypothèse alternative en rejetant l'hypothèse nulle.

- L'hypothèse nulle n'est donc **pas** l'hypothèse d'intérêt d'une étude.
- On veut formuler l'hypothèse nulle de telle sorte que son rejet implique l'hypothèse d'intérêt (H_1).

Hypothèse nulle et hypothèse alternative

- Un test statistique est associé à une **hypothèse nulle**.
- Le **but** d'un test statistique est de prouver qu'une hypothèse nulle est fausse en la confrontant aux données.
- Si les données sont incompatibles avec l'hypothèse nulle, on *rejette* l'hypothèse nulle.
- On démontre l'**hypothèse alternative** par le rejet de l'hypothèse nulle.



Exemple : pile ou face

On se doute qu'une pièce de monaie est mal équilibrée, c'est-à-dire que pile et face n'ont pas chacun 50% de chance d'être lancer.

On considère $X = \begin{cases} 1, & \text{si on lance pile,} \\ 0, & \text{sinon.} \end{cases}$

$X \sim \text{Bernoulli}(p)$, où p est inconnu.

On émet les hypothèses suivantes

$$H_0 : p = \tfrac{1}{2} \qquad\qquad H_1 : p \neq \tfrac{1}{2}.$$

Si on lance pile 5 fois de suite, est-il raisonnable de penser que H_0 est vraie ?

10 fois ? 50 fois ?

Exemple : pile ou face

Suite

$\mathbb{P}(\text{lancer n piles de suite}|H_0) = \left(\frac{1}{2}\right)^n$

$n = 5$: 0,03125

$n = 10$: 0,0009765625

$n = 50$: $8{,}881784197001252 \times 10^{-16}$

À $n = 5$, il semble raisonnable que H_0 soit vraie.

À $n = 50$, il semble absurde que H_0 soit vraie. Dans ce cas, on rejette H_0 et on accepte H_1, c'est-à-dire $p \neq \frac{1}{2}$.

À partir de quelle valeur de n est-il raisonnable de rejeter H_0 ? Cela dépend de notre seuil de tolérance.

Analogie avec un procès

Imaginez que vous êtes l'accusé·e dans un procès. Le jury **ne sait pas** si vous êtes innocent·e ou coupable, mais il considère que vous êtes **innocent·e jusqu'à preuve du contraire**.

- On peut résumer les deux possibilités par

$$H_0 : \text{vous êtes innocent·e}$$
$$H_1 : \text{vous êtes coupable}$$

- Les preuves servent à montrer que vous êtes **coupable**.
- Si les preuves sont insuffisantes, vous serez jugé **innocent**.

		Jugement	
		Innocent	Coupable
Vérité	Innocent	Bon jugement	Mauvaise condamnation
	Coupable	Libérer un coupable	Bon jugement

Erreurs de premier et de deuxième espèce

H_0 : vous êtes innocent·e

H_1 : vous êtes coupable

		Jugement	
		Innocent	Coupable
Vérité	Innocent	Bon jugement	Mauvaise condamnation
	Coupable	Libérer un coupable	Bon jugement

		Décision	
		Garder H_0	Rejeter H_0
Vérité	H_0 vraie	Bonne décision	Erreur de 1ère espèce
	H_0 faux	Erreur de 2e espèce	Bonne décision

Erreur de 1ère espèce : Rejeter H_0 alors qu'elle était vraie (condamner un innocent)
Erreur de 2e espèce : Ne pas rejeter H_0 alors qu'elle était fausse (libérer un coupable)

Probabilité d'erreur

		Décision	
		Garder H_0	Rejeter H_0
Vérité	H_0 vraie	Bonne décision	Erreur de 1ère espèce
	H_0 faux	Erreur de 2e espèce	Bonne décision

$\alpha =$ probabilité de commettre une erreur de 1er espèce, si H_0 est vraie
$\beta =$ probabilité de commettre une erreur de 2e espèce, si H_0 est fausse

		Probabilité	
		de garder H_0	de rejeter H_0
Vérité	H_0 vraie	$1-\alpha$	α
	H_0 faux	β	$1-\beta$

On appelle $1-\beta$ la *puissance du test*. On y reviendra.

Seuil du test

$\alpha =$ probabilité de commettre une erreur de 1er espèce, si H_0 est vraie
$\beta =$ probabilité de commettre une erreur de 2e espèce, si H_0 est fausse

		Probabilité	
		de garder H_0	de rejeter H_0
Vérité	H_0 vraie	$1-\alpha$	α
	H_0 faux	β	$1-\beta$

- Lorsqu'on fait un test statistique, on peut **choisir** α.
  - On l'appelle le **seuil du test**.
- On ne choisit **pas** β.
  - β dépend de α, de la taille de l'échantillon et du « degré de fausseté » de H_0
- On appelle $1-\beta$ la **puissance du test**.

Pourquoi garder le contôle de α plutôt que β

Exemple

Supposons qu'on teste l'efficacité d'un médicament qui a des **effets secondaires importants**. On pose

$$H_0 : \text{le médicament n'est pas efficace.}$$

Question

Vaut-il mieux faire une erreur de 1ère espèce ou une erreur de 2e espèce ?

Choix d'un seuil

Si deux chercheurs adoptent des **seuils α différents**, il se pourrait que pour les mêmes données, l'un rejette H_0 et l'autre non.

On introduit la **valeur** p pour qu'un lecteur puisse choisir son propre seuil.

valeur $p =$ seuil minimal au-delà duquel on rejette l'hypothèse nulle

Rejet ou non rejet

Si $p \leq \alpha$, on rejette H_0.

Si $p > \alpha$, on ne rejette pas H_0. (On « tolère » H_0.)

Exemple : Retour à pile ou face

On reprend l'exemple de pile ou face, avec H_0 : proba de pile est $\frac{1}{2}$.

Supposons que l'on joue 8 fois et que l'on observe PPPFPPPF. Si $X =$ nb de piles, alors on a observé $X = 6$. On peut calculer une valeur-p comme suit :

$$p = \mathbb{P}(X \geq 6 | H_0) = \binom{8}{6}\frac{1}{2^6}\frac{1}{2^2} + \binom{8}{7}\frac{1}{2^7}\frac{1}{2^1} + \binom{8}{8}\frac{1}{2^8} = 0{,}1445$$

Si une personne choisit un seuil de 15%, alors $p < 15\%$, donc elle rejette H_0.

Si une autre personne choisit un seuil de 10%, alors $p > 10\%$, donc elle garde H_0.

La valeur-p permet de faire le calcul une fois pour décider si on garde ou non H_0 pour différents seuils.

Valeur p et seuil du test

La **valeur** p

est la **probabilité** que le hasard de l'échantillonage puisse produire des données aussi loin (ou plus encore) de l'hypothèse nulle que notre échantillon, si l'hypothèse nulle était vraie.

$$p = \mathbb{P}(\text{observer nos données ou pire}|H_0).$$

Plus la valeur p est petite, plus H_0 est contredite par les données

Le **seuil du test**

est la probabilité de rejeter l'hypothèse nulle, alors qu'elle **était vraie**.

- seuil du test = probabilité de faire une erreur de 1ère espèce
- Le seuil du test est **choisie** par la personne (habituellement 5%).

On rejette H_0 si $p \leq$ seuil du test.

Résultat significatif

Un résultat est statistiquement *significatif*

lorsque la valeur $p < 5\%$ (ou le seuil choisi). Cela veut dire que la différence des résultats observés avec l'hypothèse nulle n'est probablement pas due au hasard.

Attention !

Signification $\neq$ Important

L'**importance** d'un résultat n'est pas mesuré par la valeur p.

Statistique de test

On pense aux données comme à une **preuve contre** H_0.

Une statistique de test, notée T_{stat},

est une mesure de *distance* entre les données et l'hypothèse nulle.

- T_{stat} est une variable aléatoire construite à partir d'un échantillon
- On l'appelle *variable décisionnelle*, car elle permet de formuler une règle de décision pour un test statistique
- t_{stat} est une réalisation (un nombre) de T_{stat} calculé à partir des données

Si $X_1, \ldots, X_n$ sont des v.a. qui représente notre échantillon, alors

$$T_{\text{stat}} = g(X_1, \ldots, X_n) \quad \text{et} \quad t_{\text{stat}} = g(x_1, \ldots, x_n),$$

où g est une fonction réelle déterminée selon le contexte.

Statistique de test et probabilité

Comme on souhaite mesurer une distance entre les données et H_0, on s'intéresse à la loi de probabilité de T_{stat} **en supposant que** H_0 **est vraie**. On calcule la valeur p par

$$p = \quad \mathbb{P}(T_{\text{stat}} \geq t_{\text{stat}} \mid H_0 \text{ vraie}) \quad \text{ou} \quad \mathbb{P}(|T_{\text{stat}}| \geq t_{\text{stat}} \mid H_0 \text{ vraie})$$

Attention !

$p = \mathbb{P}(t_{\text{stat}} \mid H_0) \neq \mathbb{P}(H_0 \mid t_{\text{stat}})$

Autrement dit, la valeur p n'est pas la probabilité que H_0 soit vraie.

La valeur p est la probabilité d'observer une valeur aussi loin de H_0 (ou plus extrême), en supposant que H_0 est vraie.

Test statistique (ou test d'hypothèse)

Il s'agit d'une procédure qui permet de rejeter ou de ne pas rejeter une hypothèse statistique.

1. Émettre une hypothèse nulle et une hypothèse alternative
2. Choisir une statistique de test calculable sur un échantillon
3. Établir mathématique la distribution de probabilité de T_{stat} sous H_0
4. Calculer la réalisation t_{stat} de T_{stat} à partir des données
5. Comparer t_{stat} avec la distribution de T_{stat} sous H_0 (calculer la valeur p)
6. Rejeter ou ne pas rejeter H_0