Tests d'inférence (rappels et grands échantillons)

Yohann.Foucher@univ-nantes.fr

Equipe d'Accueil 4275 "Biostatistique, recherche clinique et mesures subjectives en santé", Université de Nantes

Master 2 - Cours #3

UNIVERSITÉ DE NANTES
CENTRE HOSPITALIER UNIVERSITAIRE DE NANTES
itun institut transplantation urologie néphrologie INSERM - UMR 643

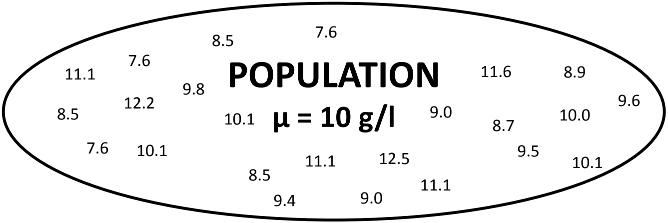
Plan

1. Principe des tests

2. Test de Student

3. Test du Chi-deux

Rappels : fluctuations d'échantillonnage



Rappels : fluctuation d'échantillonnage

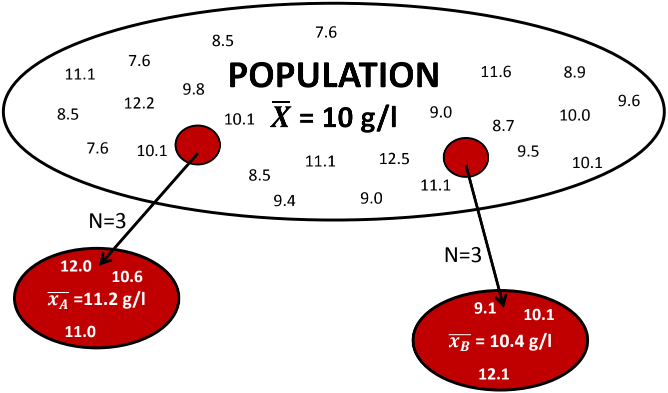
- Soit $\mathcal{P}$ la population cible.

  Ex : $\mathcal{P}$ = *L'ensemble des patients anémiques*.
- Soit μ le taux moyen d'hémoglobine

  Ex : $\bar{X} = 10$ g/l.
- Il n'est pas possible de mesurer $\bar{X}$ à partir de tous les patients de la population.
- On réalise un échantillon de N patients à partir desquels on observe une moyenne $\bar{x}$.

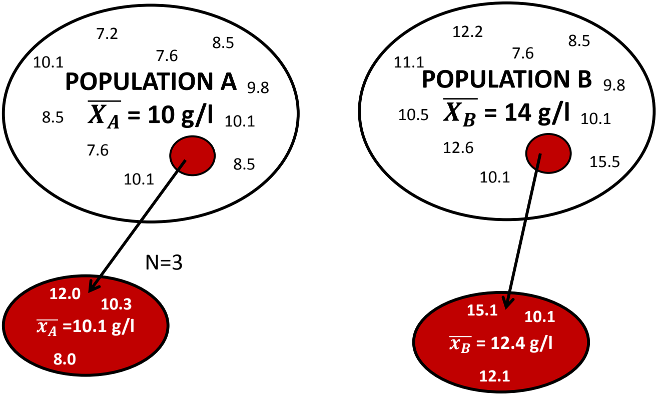
  Ex : $N = 3$.
- Problème : si plusieurs échantillons sont réalisés, on observera autant de moyennes.

Rappels : fluctuation d'échantillonnage



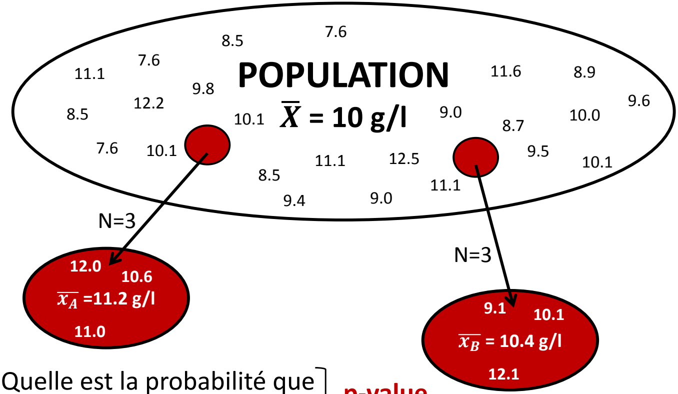
Deux hypothèses bien contrastées

Hypothèse alternative (H_1): le traitement a un effet sur le taux d'hémoglobine, c'est à dire $\overline{X_A} \neq \overline{X_B}$



Deux hypothèses bien contrastées

Hypothèse nulle (H_0): le traitement n'a pas d'effet sur le taux d'hémoglobine, c'est à dire $\overline{X_A} = \overline{X_B} = \overline{X}$



Quelle est la probabilité que cette hypothèse soit vraie ? } **p-value**

Les tests d'hypothèse

- J'observe une moyenne $\bar{x}_A$ à partir d'un échantillon A et $\bar{x}_B$ à partir d'un échantillon B (et leur variance).
- Est ce que la différence que j'observe est due à la fluctuation d'échantillonnage ?
- Deux hypothèses :

  1. La différence entre les moyennes observées est minime. Elle est due au fait que trop peu de sujets de la population ont été inclus. Si tous les sujets des deux populations avaient été inclus, on aurait observé une égalité.
     $\rightarrow$ Hypothèse nulle : H_0 $(\bar{X}_A = \bar{X}_B)$.

  2. La différence observée est importante. Elle ne peut pas être due au fait que tous les sujets n'ont pas été inclus. Il est évident que si on avait inclus tout le monde, les moyennes observées auraient été différentes.
     $\rightarrow$ Hypothèse alternative : H_1 $(\bar{X}_A \neq \bar{X}_B)$.

Les erreurs des tests d'hypothèse

H_0 $(\mu_A = \mu_B)$.
H_1 $(\mu_A \neq \mu_B)$.

- Probabilité de se tromper si on rejette H_0 :
    - Risque de première espèce.
    - On l'estime à partir des tests statistiques : probabilité critique.
    - En anglais : *p-value ou type I error*
- Probabilité de se tromper si on rejette H_1 :
    - Risque de seconde espèce (*type II error*).
    - 1 moins la *puissance* (*power*).

La mécanique à respecter

1. Définition a priori d'un seuil de décision :
   $\rightarrow$ Risque de première espèce maximal accepté.
   Ex : $\alpha = 0.05$
2. Calcul de la p-value.
   Ex : comparaison de deux moyennes et grands échantillons : t-test.
3. Conclusions :
   - Si la p-value est trop importante, c.-à-d. supérieure à α : tendance au non-rejet de H_0.
   - Si la p-value est faible, c.-à-d. inférieure à α : rejet de H_0.

Les erreurs d’interprétation répandues

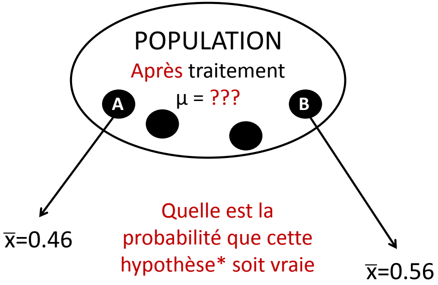
- Si la p-value est supérieure à 5%, une conclusion du type "il semble qu’il n’ait pas d’effet du traitement" est fausse !
  - Le risque d’erreur lié au rejet de H_1 est le risque de 2nd espèce.
  - Le test statistique ne donne pas cette probabilité.
  - Ex :
    - $\rightarrow$ Quand les effectifs des deux échantillons $\searrow$ alors la p-value $\nearrow$.
    - $\rightarrow$ On rejette rarement H_0 pour des effectifs très faibles.
    - $\rightarrow$ Cela ne veut pas dire qu’il n’y a pas de différence en réalité.
    - $\rightarrow$ Manque de puissance.
  - Tests d’équivalence ou de non-infériorité.
- Toujours inclure une notion d’incertitude car les populations ne sont jamais observées.

Puissance et effectifs nécessaire

- Pour tous les tests d'inférence, il est nécessaire que les effectifs soient d'autant plus importants que :
  - la différence attendue et intéressante cliniquement est petite.
  - le critère de jugement est variable.
  - les risques de première et seconde espèces sont faibles.
- Les effectifs nécessaires dépendent aussi des tests statistiques utilisés.
- Les effectifs nécessaires sont calculés a priori.

Rappels : tests pour v.a continue et grands échantillons

- On considère deux populations $\mathcal{P}_A$ et $\mathcal{P}_B$ desquelles sont extraits deux échantillons de tailles N_A et N_B. A partir de ces observations, on cherche à savoir si les caractéristiques des deux populations peuvent être considérées comme égales, ou bien si elles semblent être différentes.
- Exemple :



* Pas de différence entre les deux traitements

Rappels : tests pour v.a continue et grands échantillons

- Définition des populations et des v.a. :
  - X_A : v.a. continue dans la population $\mathcal{P}_A$ de moyenne μ_A.
    $\rightarrow$ On observe un échantillon de taille N_A $\{x_{A,1}, ..., x_{A,N_A}\}$.
  - X_B : v.a. continue dans la population $\mathcal{P}_B$ de moyenne μ_B.
    $\rightarrow$ On observe un échantillon de taille N_B $\{x_{B,1}, ..., x_{B,N_B}\}$.
- Choix des hypothèses :
  - H_0 : $\mu_A = \mu_B$ $(= \mu)$
  - H_1 : $\mu_A \neq \mu_B$ (bilatéral mais peut-être unilatéral))

Test de Student ou t-test (N_A et $N_B > 30$)

- Si $p < \alpha$.
  - Il semble que l'échantillon *A* soit issu d'une population où la moyenne μ_A est supérieur à la moyenne μ_B.
- Si $p > \alpha$.
  - On ne peut pas montrer que la moyenne de la population *A* soit supérieure à celle de *B*.

Comparaison de deux proportions

- On considère deux populations $\mathcal{P}_A$ et $\mathcal{P}_B$ desquelles sont extraits deux échantillons de tailles N_A et N_B. On suppose les proportions π_A et π_B d'une caractéristique dans chacune des deux populations. A partir des proportions observés p_A et p_B, on cherche à savoir si les proportions des deux populations peuvent être considérées comme égales, ou bien si elles semblent être différentes.

Comparaison de deux proportions

- Définition des populations et des v.a. :
  - X_A : v.a. continue binaire dans la population $\mathcal{P}_A$ de moyenne π_A.
    $\rightarrow$ On observe un échantillon de taille N_A $\{x_{A,1}, ..., x_{A,N_A}\}$.
  - X_B : v.a. continue binaire dans la population $\mathcal{P}_B$ de moyenne π_B.
    $\rightarrow$ On observe un échantillon de taille N_B $\{x_{B,1}, ..., x_{B,N_B}\}$.
- Choix des hypothèses :
  - H_0 : $\pi_A = \pi_B$ $(= \pi)$
  - H_1 : $\pi_A \neq \pi_B$ (bilatéral)

Test du Chi-deux

- Si $u \in RC \rightarrow p_c < \alpha$.
  - Rejet de H_0 car moins de α % de chance qu'elle soit vraie.
  - Il semble que l'écart entre les proportions des deux populations soit différent.
- Si $u \notin RC \rightarrow p_c > \alpha$.
  - Non rejet de H_0 car plus de α % de chance qu'elle soit vraie.
  - On ne peut pas montrer une différence significative entre les proportions des deux populations.