Tests de comparaisons de moyennes

Yohann.Foucher@univ-nantes.fr

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

Odontologie - Cours #5

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

Plan

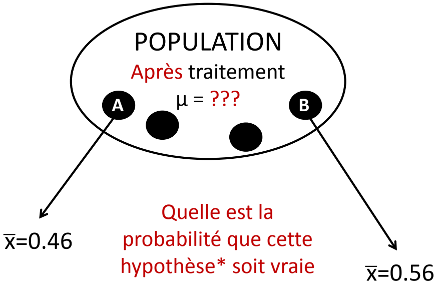
Les tests abordés

- 2 échantillons indépendants avec 2 v.a. continues :
  - grand/petits échantillons.
  - échantillons indépendants/appariés
  - tests unilatéraux/bilatéraux

Plan

Contexte

- 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

Deux moyennes théoriques (bilatéral)

- 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 $\{x_{A,1}, \ldots, x_{A,N_A}\}$.
  - X_B : v.a. continue dans la population $\mathcal{P}_B$ de moyenne μ_B.
    $\rightarrow$ On observe un échantillon de taille N_B $\{x_{B,1}, \ldots, x_{B,N_B}\}$.
- Choix des hypothèses :
  - H_0 : $\mu_A = \mu_B$ $(= \mu)$
  - H_1 : $\mu_A \neq \mu_B$

Deux moyennes théoriques (bilatéral)

- Définition de la statistique de test. Sous H_0, N_A et $N_B > 30$ et les échantillons sont indépendants, on a :

$$\bar{X}_A \sim \mathcal{N}(\mu_A, \sigma_A/\sqrt{N_A}) \text{ et } \bar{X}_B \sim \mathcal{N}(\mu_B, \sigma_B/\sqrt{N_B})$$

$$\Downarrow$$

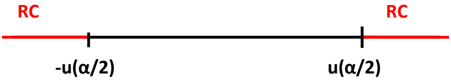
$$(\bar{X}_A - \bar{X}_B) \sim \mathcal{N}(0, \sqrt{\sigma_A^2/N_A + \sigma_B^2/N_B})$$

$$\Downarrow$$

$$U = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\sigma_A^2/N_A + \sigma_B^2/N_B}} \sim \mathcal{N}(0, 1)$$

Deux moyennes théoriques (bilatéral)

- Définition de la région critique (α, test bilatéral)



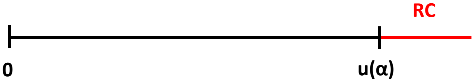
- Application numérique :

$$u = (\bar{x}_A - \bar{x}_B) \Big/ \sqrt{s_A^2/n_A + s_B^2/n_B}$$

- 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 moyennes 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 qu'une différence significative entre les moyennes des deux populations.

Deux moyennes théoriques (unilatéral)

- Identique au cas bilatéral, mais....
- H_1 : $\mu_A > \mu_B$ (l'hypothèse peut aussi être posée en infériorité)
- Loi normale, α, test unilatéral



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

Deux moyennes théoriques (exemple)

Le marqueur FoxP3 (forkhead box P3) est une protéine impliquée dans la réponse immunitaire, en particulier en régulant les lymphocytes T régulateurs. * Elle a déjà été montrée comme intéressante en transplantation rénale. Cette protéine est mesurée dans le sang de 84 patients greffés rénaux. 34 patients ont une biopsie avec des signes d'inflammation (réaction immunitaire) contre 50 patients sans inflammation. La moyenne d'expression dans le premier groupe est égale 2.1 u ($\pm$ 1.4) contre 3.5 u ($\pm$ 0.5) dans le second groupe. Peut-on considérer un niveau de FoxP3 différent entre les deux types de patients ?

*. Ashton-Chess J, et al. Regulatory, effector, and cytotoxic T cell profiles in long term kidney transplant patients. J Am Soc Nephrol. 2009 May ;20(5) :1113-22.

Deux moyennes théoriques (exemple)

- X_I : v.a. continue représentant le niveau de FoxP3 chez les patients avec inflammation. μ_I la moyenne d'expression de FoxP3 dans cette population.
- $X_{\bar{I}}$: v.a. continue représentant le niveau de FoxP3 chez les patients sans inflammation. $\mu_{\bar{I}}$ la moyenne d'expression de FoxP3 dans cette population.
- On observe un échantillon de X_I de taille N_I.
- On observe un échantillon de $X_{\bar{I}}$ de taille $N_{\bar{I}}$.
- Hypothèses à tester :
  - H_0 : L'expression moyenne de FoxP3 est identique dans les deux populations.
  - H_1 : L'expression moyenne de FoxP3 est différente dans les deux populations.

Deux moyennes théoriques

- Sous H_0, puisque N_I et $N_{\bar{I}}$ sont supérieurs à 30 :

$$U = \frac{\bar{X}_I - \bar{X}_{\bar{I}}}{\sqrt{\sigma_I^2/N_I + \sigma_{\bar{I}}^2/N_{\bar{I}}}} \sim \mathcal{N}(0,1)$$

- Région critique ($\alpha = 0.05$, test bilatéral) :

$$RC : U \notin [-1.96; 1.96]$$

- Application numérique :

$$u = (2.1 - 3.5) \Big/ \sqrt{1.4^2/34 + 0.5^2/50} = -5.59$$

- $u \in RC \rightarrow$ Il semble que l'expression moyenne de FOxP3 varie de manière significative entre les deux populations de patient (p<0.05).

Plan

Echantillons appariés

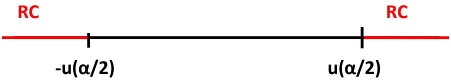
- La statistique de test précédente n'est pas valable si les deux échantillons sont appariés
- Ex : Mesure d'un biomarqueur avant et après sur chaque individu
  - X_A v.a. représentant l'expression du biomarqueur avant.
  - X_B v.a. représentant l'expression du biomarqueur après.
- On souhaite tester :
  - $H_0 : \bar{X}_A = \bar{X}_B$
  - $H_1 : \bar{X}_A \neq \bar{X}_B$
- Solution : On travaille sur la différence des deux mesures.
  - $X = X_A - X_B$
- Les hypothèse s'écrivent alors :
  - $H_0 : \bar{X} = 0$
  - $H_1 : \bar{X} \neq 0$

Echantillons appariés

- Statistique de test sous H_0. Comme $N > 30$:

$$U = \frac{\bar{X}}{\sigma/\sqrt{N}} \sim \mathcal{N}(0,1)$$

- Définir le risque de 1ère espèce maximum α.
- Région critique (RC).



- Application numérique.

Echantillons appariés

- Exemple

BAFF (B-cell activating factor) est une protéine qui est encodée par le gène TNFLSF13B. [†] Elle joue un rôle important dans la réponse immunitaire du receveur en allogreffe de rein. On souhaite tester si un traitement anti-BAFF chez des receveurs ayant développé des anticorps spécifique anti-donneur (DSA) permet de diminuer l'expression de cette protéine. 32 patients sont inclus dans l'étude. On mesure les DSA au moment du traitement et 3 mois après. La différence moyenne (avant-après) est égale à 3.1 u ($\pm$ 12.4). Conclure sur l'efficacité du traitement.

†. Shu HB, Hu WH, Johnson H (1999). TALL-1 is a novel member of the TNF family that is down-regulated by mitogens. J. Leukoc. Biol. 65 (5) : 680-3.

Echantillons appariés

- X : v.a. continue représentant la différence entre le niveau de BAFF avant et après le traitement chez les patients avec DSA (avant-après).
- $\bar{X}$: v.a. représentant la moyenne de X.
- Taille échantillon $N = 32$.
- Hypothèses à tester :
    - H_0 : $\bar{X} = 0$, pas d'effet du traitement.
    - H_1 : $\bar{X} > 0$, diminution de l'expression.

Echantillons appariés

- Sous H_0, puisque N est supérieur à 30 :

$$U = \frac{\bar{X}}{\sigma/\sqrt{N}} \sim \mathcal{N}(0,1)$$

- Région critique ($\alpha = 0.05$, test unilatéral) :

$$RC : U \notin [-\infty; 1.64]$$

- Application numérique :

$$u = 3.1 \Big/ (12.4/\sqrt{32}) = 1.41$$

- $u \notin RC \rightarrow$ L'étude ne permet pas de montrer une diminution significative du niveau de BAFF grâce à ce traitement (p>0.05).

Plan

Quelques précisions...

Tests paramétriques de comparaisons de moyennes (*t-test*) utilisables quand :

- Les effectifs sont supérieurs à 30 sujets (TCL, loi normale)
- Toujours utilisable quand les effectifs sont plus petits si :
  - les v.a. étudiées suivent une loi normale,
  - les variances sont égales (homoscédasticité).
  - $\rightarrow$ Utilisation de la loi de Student.
- Problèmes quand les effectifs sont petits :
  - Si les variables ne semblent pas suivre une loi normale.
  - Si les variances ne semblent pas être égales.
  - Manque de puissance pour montrer (1) et (2).

$\Rightarrow$ Dès que $N \leq 30$: tests non-paramétriques

Principe des tests non paramétriques

- "Aucune" hypothèse sur la distribution des variables aléatoires.
- Tests souvent basés sur la notion de rangs.
  - Si les distributions entre groupes sont $\neq$, les rangs sont $\neq$.
- Exemple :
  - Groupe A ($n = 3$) : 1, 5, 3.
  - Groupe B ($n = 3$) : 7, 6, 10.
- Rangs :
  - Groupe A : 1, 3, 2.
  - Groupe B : 5, 4, 6.
- Somme des rangs :
  - Groupe A : 6.
  - Groupe B : 15.

Principe des tests non paramétriques

- Les tests paramétriques :
  - exigent que l'on spécifie la forme de la distribution.
- Les tests non paramétriques :
  - pas de référence à une répartition particulière.
  - peuvent donc s'appliquer à des petits échantillons.
- Avantages/inconvénients :
  - Les tests non paramétriques sont théoriquement moins puissants que les tests paramétriques.
  - Des études ont cependant prouvé que l'exactitude des tests non-paramétriques sur des grands échantillons n'est que légèrement inférieure à celle des tests paramétriques.
  - Les tests non-paramétriques sont beaucoup plus exacts sur des petits échantillons.

Test de Mann-Whitney

- Permet de comparer la distribution de deux v.a. observées à partir de deux échantillons indépendants (A et B).
- Définition des variables aléatoires :
  - X_A : variable aléatoire continue dans le groupe A de taille N_A
  - X_B : variable aléatoire continue dans le groupe B de taille N_B
- Par convention, on assigne que le groupe A pour l'échantillon le plus petit ($N_A \leq N_B$).
- Choix des hypothèses

  H_0 : X_A et X_B ont la même distribution

  H_1 : X_A et X_B n'ont pas la même distribution

Test de Mann-Whitney

- Si $u \in RC \rightarrow p < \alpha$.
  - Rejet de H_0 car moins de 5% de chance qu'elle soit vraie.
  - Il semble que les deux distributions soient différentes.
- Si $u \notin RC \rightarrow p > \alpha$.
  - Non rejet de H_0 car plus de 5% de chance qu'elle soit vraie.
  - On ne peut pas montrer que les deux distributions soient différentes.

Test de Mann-Whitney

- Remarques de vocabulaire :
  - Ce test est aussi appelé Mann-Whitney/Wilcoxon.
  - On peut aussi voir "test de Wilcoxon pour échantillons indépendants".
  - A éviter mais rencontré dans la littérature : "non-parametric t-test".
- Problèmes des faibles effectifs :
  - Du point de vue statistique :
    - Test possible à partir de 3 sujets par groupe.
  - Du point de vue méthodologique :
    - Résultats très peu robustes : un sujet supplémentaire peut tout changer.
    - Résultats très peu puissants : attention à l'interprétation du non-rejet de H_0.

Plan

Généralisations

- Si plus de deux groupes ($N > 30$) : Analyse de Variance (ANOVA à 1 facteur).
- Si nécessité d'ajuster sur des facteurs de confusion :
  - ANOVA à plusieurs facteurs.
  - Régression linéaire multiple (voir cours suivants).
- Si plus de deux groupes ($N < 30$) : Test de Kruskal-Wallis.
- Si répétition des tests : correction du risque de première espèce (Bonferoni, etc.).

Test de Wilcoxon

- Permet de comparer la distribution de deux v.a. observées à partir de deux échantillons appariés (A et B).
- Définition des variables aléatoires (N sujets par groupe) :
  - (X_A, X_B) : variables aléatoires observées pour chaque paire.
  - $X = X_A - X_B$: différence pour chaque paire.
- Choix des hypothèses
  H_0 : X_A et X_B ont la même distribution
  H_1 : X_A et X_B n'ont pas la même distribution
- Principe des rangs : classement des valeurs absolues $|X|$ en excluant les valeurs nulles et en notant le signe de la différence.

Test de Wilcoxon

- Permet de comparer la distribution de deux v.a. observées à partir de deux échantillons appariés (A et B).
- Définition des variables aléatoires (N sujets par groupe) :
  - (X_A, X_B) : variables aléatoires observées pour chaque paire.
  - $X = X_A - X_B$: différence pour chaque paire.
- Choix des hypothèses
  H_0 : X_A et X_B ont la même distribution
  H_1 : X_A et X_B n'ont pas la même distribution
- Principe des rangs : classement des valeurs absolues $|X|$ en excluant les valeurs nulles et en notant le signe de la différence.