Tests non-paramétriques - Variables aléatoires continues

Petits échantillons

Yohann.Foucher@univ-nantes.fr

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

Ecole Doctorale - Module Biostatistique I, Mardi 27 Mars 2012

UNIVERSITÉ DE NANTES

CENTRE HOSPITALIER UNIVERSITAIRE DE NANTES

itun institut transplantation urologie néphrologie INSERM - UMR 643

Plan

1. Rappels

2. Test de Mann-Withney

3. Test de Wilcoxon

Plan

Rappels

- Tests paramétriques de comparaisons de moyennes utilisables quand :
  - Les effectifs sont supérieurs à 30 sujets (utilisation du théorème central limite)
  - Les effectifs sont plus petits mais on suppose que :
    - Les variables aléatoires étudiées suivent une loi normale
    - Les variances entre les groupes sont égales (homoscédasticité).
- Problèmes quand les effectifs sont petits :
  1. Si les variables ne semblent pas suivre une loi normale.
  2. Si les variances ne semblent pas être égales.
  3. Manque de puissance pour montrer (1) et (2). A ne pas négliger même si ça nous arrange !

Rappels

- Les tests non-paramétriques ne font aucune hypothèse sur la distribution des variables aléatoires.
- Ils sont souvent basés sur la notion de rangs.
- Si les distributions entre deux groupes sont différentes, les rangs seront différents.
- 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.
- Si les deux distributions sont équivalentes, on attend des sommes des rangs proches entre les deux groupes.

Plan

Test de Mann-Withney

- Permet de tester si deux échantillons indépendants (A et B) sont issus de la même population.
- 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
- Somme des rangs :
  - R_A : somme des rangs occupés par les valeurs de X_A.
  - R_B : somme des rangs occupés par les valeurs de X_B.
- Si des ex-aequos entre des valeurs, on affecte la moyenne des rangs (si les ex-aequos sont du même groupe, rien ne change).

Test de Mann-Withney

- Statistique de test :

$$M = min(M_A, M_B)$$

avec

$$M_A = n_A n_B + n_A(n_A + 1)/2 - R_A$$

et

$$M_B = n_A n_B - U_A = n_A n_B + n_B(n_B + 1)/2 - R_B$$

- Choix du seuil α (5%) et définition de la région critique. 2 cas :
  - $n_A \leq 10$
  - $n_A > 10$

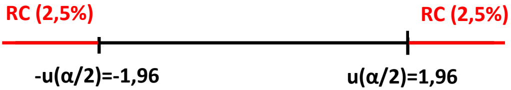
Test de Mann-Withney

- Si $n_A > 10$, alors sous H_0 :

$$M \sim N(n_A n_B/2, \sqrt{n_A n_B(n_A + n_B + 1)/12})$$

$$U = \frac{M - n_A n_B/2}{\sqrt{n_A n_B(n_A + n_B + 1)/12}} \sim N(0,1)$$

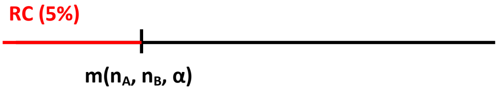
- test bilatéral, loi normale



- 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-Withney

- Si $n_A \leq 10$, alors :



- 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.

Plan

Test de Wilcoxon

- Permet de tester si deux échantillons appariés (A et B) sont issus de la même population.
- Définition des variables aléatoires (n sujets) :
  - (X_A, X_B) : variables aléatoires observées pour chaque sujet.
  - $X = X_A - X_B$: différence pour chaque sujet.
- 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
- On classe les valeurs absolues $|X|$ en excluant les valeurs nulles et en notant le signe de la différence.
- Soit n le nombre de différences non nulles.
- Somme des rangs :
  - $R(-)$: somme des rangs occupés par les différences négatives.
  - $R(+)$: somme des rangs occupés par les différences positives
- Si des ex-aequos entre des valeurs, on affecte la moyenne des rangs (si les ex-aequos sont du même groupe, le calcul ne change pas).

Test de Wilcoxon

- Statistique de test :

$$W = min(R(-), R(+))$$

- Choix du seuil α (5%) et définition de la région critique. 2 cas :
  - $n \leq 20$
  - $n > 20$
- Si $n \leq 20$, alors :

- Mêmes conclusions que pour le test de Mann-Withney.

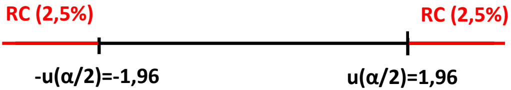
Test de Wilcoxon

- Si $n > 20$, alors sous H_0 :

$$W \sim N(n(n+1)/4, \sqrt{n(n+1)(2n+1)/24})$$

$$U = \frac{W - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \sim N(0,1)$$

- test bilatéral, loi normale



- Mêmes conclusions que pour le test de Mann-Withney.