Tests non paramétriques

Mohamed LEMDANI

MISO
Université de Lille

30 Septembre 2021

Tests de Student

Un seul échantillon : $H_0 : \{\mu_1 = \mu_0\}$ versus $H_1 : \{\mu_1 \neq \mu_0\}$.

Variable de décision : $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$.

Sous H_0, $t \sim St_{n-1}$, à condition d'avoir :

- X normalement distribuée
- ou n "grand" ($\geqslant 30$).

Deux échantillons : $H_0 : \{\mu_1 = \mu_2\}$ versus $H_1 : \{\mu_1 \neq \mu_2\}$.

Variable de décision : $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$ ou $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$.

Sous H_0, $t \sim St$, à condition d'avoir :

- X normalement distribuée (sur les deux populations),
- ou n_1 et n_2 "grands" ($\geqslant 30$).

Problème si petit(s) échantillon(s) et absence de normalité $\Longrightarrow$ alternative d'un test *non paramétrique* Wilcoxon (observé/théorique) ou Mann et Whitney (observé/observé).

Principe et mise en œuvre

À n'utiliser que si l'on observe une variable quantitative X sur 2 échantillons indépendants et que (pour au moins l'un des deux) on ait :

- $n < 30$
- et X n'admettant pas une distribution normale.

Comparaison des médianes et non des moyennes (sauf si distribution symétrique).

Mise en œuvre du test de Mann et Whitney :

Exemple 7 : Comparaison des tailles moyennes (cm) de deux espèces de plantes A et B.

Echantillon de A : 21,5 ; 29,3 ; 72,4 ; 24,2 ; 65,6 ; 63,0 ; 24,1 ; 79,2 ; 75,1 $(n_1 = 9)$.

Echantillon de B : 42,4 ; 37,2 ; 25,2 ; 36,6 ; 45,1 ; 29,5 $(n_1 = 6)$.

Étape 0 : X : taille de la plante.

Petits échantillons et rejet de la normalité de X pour l'espèce B (test fait au préalable).

$H_0 : \{Me_1 = Me_2\}$ versus $H_1 : \{Me_1 \neq Me_2\}$.

Mise en œuvre

Étape 1 : Ranger, par ordre croissant, l'ensemble des observations des 2 échantillons (en les distinguant l'un de l'autre) :

$$21,5_1; \quad 24,1_2; \quad 24,2_3; \quad \underline{25,2}_4; \quad 29,3_5; \quad \underline{29,5}_6; \quad \underline{36,6}_7; \quad \underline{37,2}_8;$$

$$\underline{42,4}_9; \quad \underline{45,1}_{10}; \quad 63,0_{11}; \quad 65,6_{12}; \quad 72,4_{13}; \quad 75,1_{14}; \quad 79,2_{15}$$

Étape 2 : Déterminer les rangs. Si nécessaire, calculer le rang moyen.

Étape 3 : Calcul de la variable de décision $U = \min(U_{12}, U_{21}) = 23$.

$T_1 = \Sigma$ rangs de l'échantillon 1 $= 1 + 2 + 3 + 5 + 11 + 12 + 13 + 14 + 15 = 76$.

$T_2 = \Sigma$ rangs de l'échantillon 2 $= 4 + 6 + 7 + 8 + 9 + 10 = 44$.

$$U_{12} = n_1 n_2 + \frac{n_1(n_1+1)}{2} - T_1 = T_2 - \frac{n_2(n_2+1)}{2}$$

.

$$U_{21} = n_1 n_2 + \frac{n_2(n_2+1)}{2} - T_2 = T_1 - \frac{n_1(n_1+1)}{2} \Longrightarrow U_{12} + U_{21} = n_1 n_2$$

$$U_{12} = T_2 - \frac{n_2(n_2+1)}{2} = 44 - \frac{6 \times 7}{2} = 23 \Longrightarrow U_{21} = 9 \times 6 - 23 = 31.$$

Mise en œuvre (suite)

Interprétation de U_{12} et U_{21} :

U_{12} : nombre de fois où une observation de l'échantillon 1 précède une observation de l'échantillon 2.

$$21,5; \quad 24,1; \quad 24,2; \quad \underline{25,2}; \quad 29,3; \quad \underline{29,5}; \quad \underline{36,6}; \quad \underline{37,2};$$

$$\underline{42,4}; \quad \underline{45,1}; \quad 63,0; \quad 65,6; \quad 72,4; \quad 75,1; \quad 79,2$$

$U_{12} = 6+6+6+5+0+0+0+0+0 = 23$.

Étape 4 : Recherche de la zone de rejet sur la table.

Lecture table $\longrightarrow$ 12 $\Longrightarrow$ Zone de rejet $= [0, 12]$ $\Longrightarrow$ Non-rejet de H_0 au seuil de 5%.

Cas d'un test unilatéral : choisir au préalable entre U_{12} et U_{21} et lecture adaptée sur la table (formules identiques).

Cas de "grandes" valeurs de n_1 et/ou n_2 : Utilisation des formules d'approximation et de la table de la loi normale (1.96 si $\alpha = 5\%$ en bilatéral).

Test de Wilcoxon

Exemple 8 : Comparaison des effets de deux traitements A et B sur le niveau de dosage d'une hormone (séries appariées).

X_A, X_B : dosages (en U) suite à chacun des traitements $\Longrightarrow D = X_B - X_A$ $(n = 10)$.

Taille d'échantillon n "petite" et D peut-être non normalement distribuée.

$H_0 : \{Me_D = 0\}$ versus $H_1 : \{Me_D > 0\}$.

X_A	2,5	4,3	6,6	5,4	3,8	4,2	3,9	4,7	2,9	3,7
X_B	3,2	3,8	6,8	5,5	4,0	4,0	4,2	5,3	3,8	3,6
D	0,7	$-0,5$	0,2	0,1	0,2	$-0,2$	0,3	0,6	0,9	$-0,1$

Étape 1 : Calcul de $D = X_B - X_A \Longrightarrow$ éliminer (s'il y en a) les différences nulles (nouvelle taille d'échantillon).

Étape 2 : Ranger, par ordre croissant, les valeurs absolues de D

$$-0,1_{1.5};\quad 0,1_{1.5};\quad -0,2_4;\quad 0,2_4;\quad 0,2_4;\quad 0,3_6;\quad -0,5_7;\quad 0,6_8;\quad 0,7_9;\quad 0,9_{10}$$

Calculer si nécessaire le rang moyen.

Test de Wilcoxon (suite)

Étape 3 : Calculer $P = \Sigma$ rangs des différences $+$.
$M = \Sigma$ rangs des différences $- = 1.5 + 4 + 7 = 12.5$.
Remarque : $P + M = n(n+1)/2$.
Test bilatéral : $\Longrightarrow T = \min(M, P)$.
Test unilatéral : choisir au préalable entre M et P celui qui sera plus petit sous H_1.
Sous $H_1, Me_D > 0 \Longrightarrow$ les différences sont généralement positives $\Longrightarrow P$ "grand" et M "petit" $\Longrightarrow T = M = 12.5$.

$$-0,1_{1.5}; \quad 0,1_{1.5}; \quad -0,2_4; \quad 0,2_4; \quad 0,2_4; \quad 0,3_6; \quad -0,5_7; \quad 0,6_8; \quad 0,7_9; \quad 0,9_{10}$$

Étape 3 : Recherche de la zone de rejet sur la table.
Lecture table $\longrightarrow 10 \Longrightarrow$ Zone de rejet $= [0, 10] \Longrightarrow$ Non-rejet de H_0 au seuil de 5%.
Cas de $n > 20$: utilisation des formules d'approximation.