

Estimations de tests non paramétriques par permutations

Jacques VAILLÉ jacques.vaille@free.fr

Permutations :

La simulation de permutations et leur comptage (voir la fiche Comment compter des simulations) permettent d'obtenir des estimations dans le cas de tests non paramétriques. Ces estimations ne se basent que sur les données, mais elles utilisent toutes les données de l'échantillon.

Test sensible à une différence des valeurs centrales :

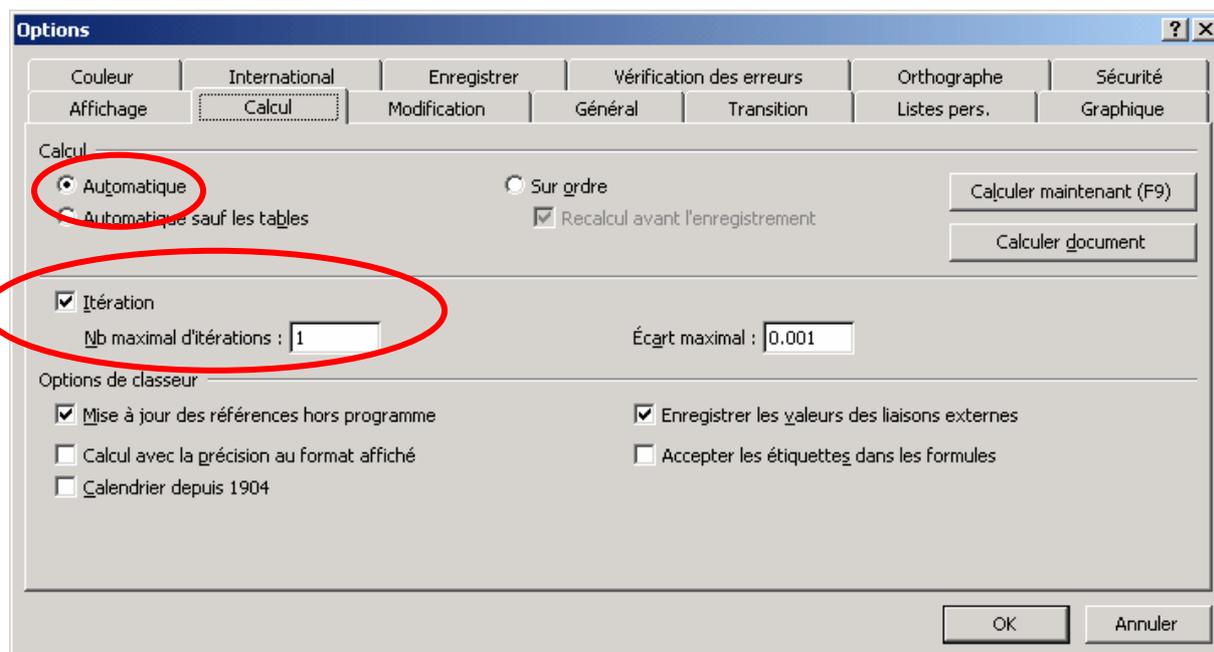
Voir la feuille TestCentralePermut.xls.

L'hypothèse nulle à tester est que deux échantillons de taille n_1 et n_2 proviennent de populations de même distribution. Sous cette hypothèse, N échantillons de taille n_1 , obtenus en tirant sans remise leurs éléments au hasard dans l'ensemble des deux échantillons, pourront servir à estimer la distribution d'un paramètre d'un échantillon de taille n_1 .

On utilise comme paramètre la somme des éléments de l'échantillon. Si k échantillons donnent une somme inférieure à la somme S_1 du premier échantillon, $\frac{k}{N}$ sera une estimation de $P_{H_0}(S < S_0)$.

Pour utiliser la feuille de calcul :

Vérifier d'abord les options de calcul en allant dans le menu Outils > Options. Il faut que celles de calcul soient les suivantes :



Sur la première ligne, on saisit les effectifs des deux échantillons (leur somme ne doit pas dépasser 255, le nombre de colonnes de la feuille).

En ligne 5 apparaissent deux zones de couleur pour saisir les valeurs des deux échantillons.

En ligne 2 apparaissent $n = n_1 + n_2$ valeurs aléatoires et au-dessous leurs rangs qui constituent une permutation aléatoire des n individus.

En ligne 4 sont calculées les différentes plages de l'échantillon et des permutations.

On trouve des compteurs en **F12**, **F13** et dans la plage **E17 :E27**. Ils permettent de construire la courbe des fréquences cumulées et de calculer la probabilité d'obtenir une somme inférieure à celle observée.

En **F15**, on saisit le seuil choisi (0.05). La probabilité estimée en **I15** apparaît sur fond rouge si elle est inférieure à ce seuil.

Enfin la liste déroulante en A10 permet d'initialiser les compteurs ou de passer en mode simulation. Il ne reste plus qu'à maintenir appuyée la touche F9 jusqu'à ce que la courbe se stabilise.

Test d'autres paramètres :

La feuille Test2EchantPermut.xls utilise les mêmes méthodes pour tester divers paramètres dans le cas d'un échantillon ayant reçu un traitement et d'un échantillon témoin.

On peut tester ainsi la moyenne, la médiane, la demi-somme des extrêmes (notée Milieu), la variance, la moyenne des écarts absolus à la moyenne (notée écart moyen), la moyenne des écarts absolus à la médiane (notée écart médian) et l'étendue. On obtient chaque fois une estimation de la fonction de répartition sous l'hypothèse que les deux échantillons proviennent de populations de même distribution.

Comme on ne peut plus utiliser la formule **SOMME.SI()**, il y a ici une ligne supplémentaire où apparaît l'échantillon permuté. On a aussi, pour éviter les rangs égaux au cas où **ALEA()** donnerait deux valeurs identiques, ajouté $0.0001 * \text{numéro de colonne}$ et calculé les rangs des nouvelles valeurs qui sont sûrement distinctes.

Dans les cas où il y a des cases en jaune, on aura besoin de modifier ces valeurs pour couvrir l'étendue de variation du paramètre étudié. (Après modification, il faut refaire la simulation). On sera amené aussi à modifier les bornes sur l'axe horizontal pour adapter le graphique aux séries étudiées.

Test d'un coefficient de corrélation linéaire :

On peut de façon analogue tester la nullité du coefficient de corrélation linéaire entre deux séries. Sous cette hypothèse, on obtient des séries issues de la même population en permutant les valeurs de X et en conservant celles de Y. La simulation permettra donc d'avoir une estimation de la distribution du coefficient de corrélation linéaire r entre les deux séries.

C'est ce que permet de tester la feuille TestCorrelPermut.xls qui s'utilise comme les précédentes :

On retrouve la case Initialiser/Permuter, la case où on saisit l'effectif des deux séries, la case où on saisit le seuil pour le test.

Une fois l'effectif saisi, deux zones colorées indiquent où saisir chacune des deux séries.