

TP no 4 - Tests par permutation

Master parcours SSD - UE Statistique Computationnelle

1 Exercice 1 - tests de permutation et différence de moyennes

Dans cet exercice nous allons mettre en oeuvre une procédure de permutation pour faire un test de différence de moyennes. Pour cela nous travaillerons sur le jeu "chicken weights" qui vise à comparer l'efficacité de différents aliments sur la prise de poids de poulets.

1. Charger le jeu de données via la commande `attach(chickwts)`, ce qui a pour effet de charger deux variables : `weight` et `feed`. De combien d'observations dispose t'on pour chaque type de nourriture ? Tracer la distribution du poids en fonction du type de nourriture.
2. Nous voulons tester s'il y a une différence de poids significative entre des poulets nourris avec `soybean` et `linseed`. Au vu du graphique précédent, pensez-vous que cette différence soit significative ? Quelle est la statistique du test de Student et la p-valeur associée ?
3. Combien de permutations faudrait-il considérer pour faire un test par permutation exact ?
4. Implémenter la procédure de permutation "randomisée" décrite en cours pour un nombre $B = 1000$ de tirages. Quelle p-valeur obtenez-vous ? Représenter la distribution de la statistique de test obtenue lors des permutations et positionner la valeur observée sur l'échantillon.
5. Représenter l'évolution de la p-valeur estimée quand on passe de $B = 100$ à $B = 2000$ tirages par pas de 100. La procédure converge t'elle rapidement ?
6. Faire la même analyse pour comparer l'alimentation par `linseed` et `sunflower` en considérant $B = 1000$.

2 Exercice 2 - tests de permutation et dé pipé

A l'issue de multiples lancers d'un même dé, on a obtenu les résultats suivants :

face	1	2	3	4	5	6
tirages	8	9	19	6	8	10

On souhaite tester si ce dé est pipé, i.e., s'il conduit à obtenir certaine(s) face(s) plus fréquemment qu'attendu.

1. Quel test paramétrique peut-on utiliser et avec quelle hypothèse nulle ? Donner une définition intuitive de la statistique de test.
2. Quelle p-valeur obtient-on en appliquant la fonction R correspondante ?
3. Proposer et appliquer une procédure par permutation (randomisée) pour $B = 1000$ tirages. Les conclusions changent-elles ?

3 Exercice 3 - tests de permutation et corrélation

Reprendre les données **LSAT** et **GPA** du TP précédent¹. On s'intéressera de nouveau à la corrélation entre ces deux variables.

1. Quelle corrélation observe t'on sur ce jeu de données entre ces deux variables ?
2. Comment mettre en oeuvre une stratégie de permutation pour tester l'hypothèse (nulle) que cette corrélation est égale à zéro ?
3. Peut-on considérer qu'elle est significativement différente de zéro selon ce test par permutation ?

4 Exercice 4 - hypothèse d'échangeabilité

1. Reproduire l'expérience illustrant l'impact de l'hypothèse d'échangeabilité donnée dans le cours.
2. Tracer l'évolution de la puissance estimée en fonction des valeurs de σ .
3. Comparer ces résultats avec le t -test. Ce test paramétrique s'en sort-il mieux ?

1. On les retrouve en chargeant le package **bootstrap** et en venant chercher les variables du même nom dans la data.frame **law**.