

Tests de randomisation et bootstrap

Ce travail doit être remis avant le **3 février à 17h** sur Moodle.

Données

Ce laboratoire utilise la base de données Portal, qui contient des données de suivi à long terme de plusieurs espèces de rongeurs sur un site d'étude en Arizona.

Ernest, M., Brown, J., Valone, T. and White, E.P. (2018) *Portal Project Teaching Database*.
https://figshare.com/articles/Portal_Project_Teaching_Database/1314459.

Le tableau de données `portal_surveys.csv` contient une rangée par individu capturé. Les variables incluent la date (jour, mois, année), le numéro de parcelle, le code d'espèce, le sexe, la longueur de patte arrière et le poids des individus.

```
surveys <- read.csv("../donnees/portal_surveysB.csv")
str(surveys)
```

```
## 'data.frame':   36701 obs. of  10 variables:
## $ X             : int  1 2 3 4 5 6 7 8 9 10 ...
## $ record_id     : int  1 2 3 4 5 6 7 8 9 10 ...
## $ month         : int  7 7 7 7 7 7 7 7 7 7 ...
## $ day          : int  16 16 16 16 16 16 16 16 16 16 ...
## $ year         : int  1977 1977 1977 1977 1977 1977 1977 1977 1977 1977 ...
## $ plot_id      : int  2 3 2 7 3 1 2 1 1 6 ...
## $ species_id   : chr  "NL" "NL" "DM" "DM" ...
## $ sex          : chr  "M" "M" "F" "M" ...
## $ hindfoot_length: int  32 33 37 36 35 14 NA 37 34 20 ...
## $ weight       : num  NA NA NA NA NA NA NA NA NA NA ...
```

Le tableau de données `portal_plots.csv` indique le type de traitement appliqué à chaque parcelle. Les traitements visent à exclure différents types de rongeurs: “Control” = aucune clôture, pas d'exclusion; “Rodent Exclosure” = clôture, tous les rongeurs exclus; “Krat Exclosure” = clôture avec porte laissant passer les petits rongeurs, mais pas les rats-kangourous. Ces traitements ont été assignés aléatoirement après délimitation des parcelles.

```
plots <- read.csv("../donnees/portal_plots.csv")
str(plots)
```

```
## 'data.frame':   24 obs. of  2 variables:
## $ plot_id : int  1 2 3 4 5 6 7 8 9 10 ...
## $ plot_type: chr  "Spectab exclosure" "Control" "Long-term Krat Exclosure" "Control" ...
```

1. Tests de randomisation

a) Tout d'abord, nous devons préparer les données pour l'analyse:

- Dans le tableau `surveys` ne gardez que les captures de *Néotoma albigula* (NL) où le poids n'est pas manquant. *Rappel*: La fonction `is.na(x)` vérifie si `x` est une valeur manquante.
- Finalement, joignez les tableaux `surveys` et `plots` pour connaître les traitements des parcelles liés à chaque observation. Vous pouvez utiliser la fonction `merge` dans R ou la fonction `inner_join`,

qui requiert le package *dplyr*. Nommez le tableau résultant `surveys_plots` et n'y conservez que les observations des sites de type "Long-term Krat Exclosure", "Short-term Krat Exclosure", et "Control".

Ensuite, visualisez la distribution du poids (`weight`, en grammes) des individus selon l'année.

- b) Nous utiliserons un test de randomisation basé sur la régression linéaire pour déterminer si la masse des individus a diminuer avec le temps. Pourquoi penses-tu qu'une approche par permutation soit appropriée ? Pour ce faire, nous écrirons une fonction qui randomise les années, avant d'exécuter la régression.
- c) Créez la fonction décrite en (b), qui effectue une randomisation de `year`, exécute une régression du poids des individus en fonction de l'année, puis retourne la valeur t . Déterminez la distribution de cette statistique pour l'hypothèse nulle avec 4999 permutations. Quelle est la valeur p pour la valeur t observée si il n'y a eu aucun changement en masse des individus capturés au court du temps?
- d) La différence est-elle significative avec un seuil $\alpha = 0.01$?
- e) Effectuez un test de permutation pour tester si les changements en masses diffèrent selon les traitements (ie. si il y a une interaction entre `year` et `plot_type`).

2. Bootstrap

- a) Calculez l'intervalle de confiance à 99% pour le changement en masse au court du temps pour chaque traitement.
- b) L'intervalle de confiance obtenu en a) est-il cohérent avec le résultat du test en 1.d)? Est-ce que le bootstrap représente bien le processus d'échantillonnage pour ce problème?
- c) Utilisez la méthode du bootstrap avec 10 000 réplcats pour calculer la différence de poids des individus entre le début et la fin de l'étude pour les traitements "Long-term Krat Exclosure" et "Control". Effectuez une correction du biais et rapportez la différence corrigée avec son erreur-type.)