Randomization tests and bootstrap

This assignment must be submitted before **February 3th at 5pm** on Moodle.

Data

This lab uses the Portal database, which contains long-term monitoring data for several rodent species at a study site in 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.

The portal_surveys.csv dataset contains one row per captured individual. Variables include the capture date (day, month, year), plot number, species code, sex, hindfoot length and weight of individuals.

```
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 ...
```

The portal_plots.csv dataset indicates the type of treatment applied to each plot. The treatments are designed to exclude different types of rodents: "Control" = no fence, no exclusion; "Rodent Exclusion" = fence, all rodents excluded; "Krat Exclusion" = fence with a gate for small rodents, but not for kangaroo rats. These treatments were randomly assigned after setting up the plots.

```
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. Randomization tests

a) First, we must prepare the data for analysis:

- In the `surveys` table, keep only the observations from *Néotoma albigula* (NL) where the weight is not missing. *Reminder*: The function `is.na(x)` checks if `x` is a missing value.
- Finally, join the `surveys` and `plots` data frames and only keep plots of type "Long-term Krat Exclosure","Short-term Krat Exclosure", and "Control". to find out which plot treatment is related to each observation. You can use the `merge` function in R or the `inner_join` function, which requires the *dplyr* package. Name the resulting data frame `surveys_plots`.

Next, view the distribution of the weight (in grams) of the individuals according to the year.

b) We will use a randomization test based on linear regression to determine if the weight of captured individuals changes with year. Why do you think a permutation approach is appropriate? To do this, we will write a function that randomizes year, before running the lm.

c) Create the function described in (b), which performs a randomization of `year`, performs an lm of the weight of individuals as a function of year, and then returns the value t. Determine the distribution of this statistic for the null hypothesis with 4999 permutations. What is the p value for the observed t value if time has no effect on the mass of individuals captured?

d) Is the difference significant with a threshold $\alpha = 0.01$?

e) Perform a new randomization test to check if the decline in mass differs between treatments.(ie. if there is an interaction between year and plot_type)

2. Bootstrap

a) Calculate the 99% confidence interval for the change in mass of the different treatments.

b) Is the confidence interval obtained in a) consistent with the test result in 1.e)? Does the bootstrap accurately represent the sampling process for this problem?

c) Use the bootstrap method with 10,000 replicates to calculate the difference in weight of individuals between the start and the end of the study for the "Long-term Krat Exclosure" and "Control" treatments. Perform bias correction and report the corrected difference with its standard error.)