

**R exercises** | *Exercices R*  
Learning Days practice sessions | *Sessions pratiques Learning Days*

Vincent + Alyssa

2026-06-12

# Section 1

## How to work

## Step 1 — create your exercise folder

### Étape 1 — créer votre dossier

**Before Activity 1**, set up a dedicated folder on your computer, for example:

learningdays\_exercises/

Inside it, create **one empty R script**, for example:

my\_exercises.R

You will build this file as you go — **these slides are your only workbook** (no separate handout).

**Avant l'activité 1**, créez un dossier dédié sur votre ordinateur, par exemple :

learningdays\_exercises/

À l'intérieur, créez **un script R vide**, par exemple :

mes\_exercices.R

Vous construirez ce fichier au fur et à mesure — **ces slides sont votre seul support** (pas de fichier séparé).

## Step 2 — download the data

### Étape 2 — télécharger les données

Download `data_for_analysis.csv` and save it **inside the same folder** as your `.R` file (not in a subfolder):

Download `data_for_analysis.csv`

Your folder should look like:

```
learningdays_exercises/  
  my_exercises.R  
  data_for_analysis.csv
```

In RStudio: **File** → **New Project** → **Existing Directory** and select this folder so `getwd()` points here.

Téléchargez `data_for_analysis.csv` et enregistrez-le **dans le même dossier** que votre fichier `.R` (pas dans un sous-dossier) :

Télécharger `data_for_analysis.csv`

Votre dossier doit ressembler à :

```
learningdays_exercises/  
  mes_exercices.R  
  data_for_analysis.csv
```

Dans RStudio : **File** → **New Project** → **Existing Directory** et sélectionnez ce dossier pour que `getwd()` pointe ici.

## Step 3 — copy, paste, edit

### Étape 3 — copier, coller, modifier

For **each slide** with R code:

- 1 **Copy** the code block from the slide into your `.R` file
- 2 **Edit** placeholders (variable names, seeds, answers as comments)
- 3 **Run** the line or selection with **Ctrl + Enter** (Windows/Linux) or **Cmd + Enter** (Mac)

Add **comments** in your script for answers to questions (`# Answer: ...`).

Do **not** skip ahead — work through Activities 1 → 5 in order.

Pour **chaque slide** avec du code R :

- 1 **Copiez** le bloc de code de la slide dans votre fichier `.R`
- 2 **Modifiez** les espaces réservés (noms de variables, graine, réponses en commentaires)
- 3 **Exécutez** la ligne ou la sélection avec **Ctrl + Entrée** (Windows/Linux) ou **Cmd + Entrée** (Mac)

Ajoutez des **commentaires** dans votre script pour les réponses (`# Réponse : ...`).

Ne **sautez pas** d'étapes — faites les activités 1 → 5 dans l'ordre.

## Step 4 — check your setup

### Étape 4 — vérifier la configuration

Before starting Activity 1, confirm:

- Exercise folder exists
- Empty .R file created and open in RStudio
- `data_for_analysis.csv` is **in that folder**
- RStudio project points to that folder

Then run the check below — you should see the CSV in the list.

```
getwd()  
dir()
```

Avant de commencer l'activité 1, vérifiez :

- Le dossier d'exercices existe
- Un fichier .R vide est créé et ouvert dans RStudio
- `data_for_analysis.csv` est **dans ce dossier**
- Le projet RStudio pointe vers ce dossier

Puis exécutez le contrôle ci-dessous — vous devez voir le CSV dans la liste.

## Section 2

### **Overview**

Five activities — all instructions and code are **on these slides**:

- 1 R basics — load and inspect data
- 2 Randomization — simple, complete, blocked
- 3 Difference in means — estimate the ATE by hand
- 4 Power — EGAP online calculator (no R)
- 5 Estimation — `estimatr` and **analyze as you randomize**

Cinq activités — toutes les instructions et le code sont **sur ces slides** :

- 1 Bases de R — charger et inspecter les données
- 2 Randomisation — simple, complète, par blocs
- 3 Différence des moyennes — estimer l'ATE à la main
- 4 Puissance — calculateur EGAP en ligne (sans R)
- 5 Estimation — `estimatr` et **analysez comme vous randomisez**

## The data

### *Les données*

Learning Days participant data from the in-class experiment. Variables include participant id, **Gender**, country, assigned **treatment**, and outcomes.

**Your job:** inspect the data and identify which columns are treatment and outcome before you analyze.

```
head(read.csv("data_for_analysis.csv"))
```

Données des participants aux Learning Days (expérience en classe). Variables : identifiant, **Gender**, pays, **treatment** assigné, et résultats. **Votre tâche :** inspectez les données et identifiez les colonnes de traitement et de résultat avant d'analyser.

## Section 3

### **Activity 1**

**Goals:** load data, inspect variables, get comfortable with RStudio.

Copy each code block below into your `.R` file as you go.

**Objectifs :** charger les données, inspecter les variables, se familiariser avec RStudio.

Copiez chaque bloc de code ci-dessous dans votre fichier `.R` au fur et à mesure.

## Activity 1 — task 1

### *Activité 1 — tâche 1*

Confirm the CSV is in your project folder. Run `dir()` — do you see `data_for_analysis.csv`?  
If not, ask an instructor.

```
dir()
```

Vérifiez que le CSV est dans votre dossier de projet. Exécutez `dir()` — voyez-vous `data_for_analysis.csv` ?  
Sinon, demandez de l'aide à un instructeur.

## Activity 1 — task 2

### Activité 1 — tâche 2

Read the data into object `experiment_data`.

Check the **Environment** pane — does it appear?

Lisez les données dans l'objet `experiment_data`.

Regardez l'onglet **Environment** — apparaît-il ?

```
experiment_data <- read.csv("data_for_analysis.csv")
```

## Activity 1 — task 3

### Activité 1 — tâche 3

Open the **data viewer** (click `experiment_data` in Environment).

**Question:** What are the variable names? Which might be treatment? Which might be outcomes?

Ouvrez le **data viewer** (cliquez sur `experiment_data` dans Environment).

**Question :** Quels sont les noms de variables ? Lesquels pourraient être le traitement ? Les résultats ?

Section 4

**Activity 2**

## Randomization — goals

### *Randomisation — objectifs*

**Context:** Instructors randomized treatment in the class experiment. Now **you** randomize — simple, complete, and blocked assignment — using the same participant data.

**Contexte :** Les instructeurs ont randomisé le traitement lors de l'expérience en classe. Maintenant **vous** randomisez — assignation simple, complète et par blocs — avec les mêmes données.

## Activity 2 — load data

### *Activité 2 — charger les données*

```
experiment_data <- read.csv("data_for_analysis.csv")
```

## Activity 2 — simple randomization

### Activité 2 — randomisation simple

Set a **seed** so your randomization is reproducible.  
Use `rbinom()` to assign 0/1 with probability 0.5.

Fixez une **graine** (`set.seed`) pour reproduire la randomisation. Utilisez `rbinom()` pour assigner 0/1 avec probabilité 0,5.

```
set.seed(123) # choose your own seed

experiment_data$treatment_simple <-
  rbinom(n = nrow(experiment_data), size = 1, prob = 0.5)
```

## Activity 2 — check simple assignment

### *Activité 2 — vérifier l'assignation simple*

Inspect `treatment_simple`. Count treated vs control with `table()`. Expect **about** half treated (not necessarily exact).

Inspectez `treatment_simple`. Comptez traitement vs contrôle avec `table()`. Attendez-vous à **environ** la moitié en traitement (pas forcément exact).

```
table(experiment_data$treatment_simple)
```

## randomizr Commands Refresher

### *Rappel des commandes randomizr*

```
library(randomizr)  
set.seed(100)
```

```
# Simple random assignment (N = 4; arms not necessarily equal)  
# Assignation aléatoire simple (N = 4 ; bras pas forcément égaux)*  
simple_ra(4)
```

```
[1] 0 0 1 0
```

```
# Complete random assignment (N = 4; exactly half treated)  
# *Assignation complète (N = 4 ; exactement la moitié en traitement)*  
complete_ra(4)
```

```
[1] 0 0 1 1
```

```
# Example: blocked assignment within 4 blocks of 3 units  
# Exemple : assignation par blocs dans 4 blocs de 3 unités*  
block_ra(blocks = c(1,1,1,2,2,2,3,3,3,4,4,4))
```

```
[1] 1 1 0 1 0 0 1 0 1 0 1 0
```

## Activity 2 — complete randomization

### Activité 2 — randomisation complète

```
# Example: shuffle a vector with sample()
# Exemple : mélanger un vecteur avec sample()*
sample(c(1,2,3))
```

```
[1] 3 2 1
```

Build a vector with **exactly**  $m$  treated and  $N - m$  control, then shuffle with `sample()`.

Construisez un vecteur avec **exactement**  $m$  traités et  $N - m$  contrôles, puis mélangez avec `sample()`.

```
N <- nrow(experiment_data)
m <- floor(N / 2)
vec <- c(rep(1, m), rep(0, N - m))

experiment_data$treatment_complete <- sample(vec)
```

## Activity 2 — check complete assignment

### *Activité 2 — vérifier l'assignation complète*

Use `table()`. Treated and control counts should match **exactly** what you put in `vec`.

Utilisez `table()`. Les effectifs doivent correspondre **exactement** à ce que vous avez mis dans `vec`.

```
table(experiment_data$treatment_complete)
```

## Activity 2 — randomizr

### Activité 2 — package randomizr

Load **randomizr** and try `complete_ra()`. Then check with `table()` as before.

Chargez **randomizr** et essayez `complete_ra()`. Vérifiez avec `table()` comme avant.

```
experiment_data$treatment_complete2 <-  
  complete_ra(N = nrow(experiment_data), prob = 0.5)
```

## Activity 2 — blocked randomization

### Activité 2 — randomisation par blocs

Use **Gender** as the block variable. `block_ra()` randomizes within each block.

Utilisez **Gender** comme variable de bloc. `block_ra()` randomise à l'intérieur de chaque bloc.

```
experiment_data$treatment_blocked <-  
  block_ra(blocks = experiment_data$Gender, prob = 0.5)  
  
table(experiment_data$Gender, experiment_data$treatment_blocked)
```

## Activity 2 — cluster randomization

### Activité 2 — randomisation par cluster

Use **HomeCountry** as the cluster variable.  
`cluster_ra()` randomizes by cluster.

Utilisez **HomeCountry** comme variable de cluster. `block_ra()` randomise par cluster (grappe).

```
experiment_data$treatment_clustered <-  
  cluster_ra(clusters = experiment_data$HomeCountry, prob = 0.5)  
  
table(experiment_data$HomeCountry, experiment_data$treatment_clustered)
```

## Section 5

### **Activity 3**

## Difference in means — goals

### *Différence des moyennes — objectifs*

**Context:** Treatment was randomized; outcomes were measured. Estimate the **average treatment effect** by hand (difference in means) — a first step toward hypothesis testing.

**Contexte :** Le traitement a été randomisé ; les résultats ont été mesurés. Estimez l'**effet moyen du traitement** à la main (différence de moyennes) — première étape vers les tests d'hypothèse.

## Activity 3 — identify variables

### *Activité 3 — identifier les variables*

From the data viewer, write down:

- 1 Outcome variable name
- 2 Treatment variable name and which value means treated

D'après le data viewer, notez :

- 1 Nom de la variable de résultat
- 2 Nom de la variable de traitement et quelle valeur = traité

## Activity 3 — overall mean

### Activité 3 — moyenne globale

Use `mean(..., na.rm = TRUE)` to drop missing values (NA).

Replace the placeholder below with your outcome variable.

```
mean(experiment_data$OUTCOME_VARIABLE, na.rm = TRUE)
```

Utilisez `mean(..., na.rm = TRUE)` pour exclure les valeurs manquantes (NA).

Remplacez l'espace réservé par votre variable de résultat.

## Activity 3 — group means

### Activité 3 — moyennes par groupe

Compute the mean outcome in **treatment** and **control** using subsets in [ ].

Then subtract:  $\bar{Y}_{\text{treat}} - \bar{Y}_{\text{control}} = \text{estimated ATE}$ .

Calculez la moyenne du résultat en **traitement** et en **contrôle** avec des sous-ensembles [ ].

Puis soustrayez :  $\bar{Y}_{\text{traitement}} - \bar{Y}_{\text{contrôle}} = \text{ATE estimé}$ .

```
# treatment group - replace OUTCOME and TREATMENT / value
mean(experiment_data$OUTCOME_VARIABLE[
  experiment_data$TREATMENT_VARIABLE == 1],
na.rm = TRUE)

# control group
mean(experiment_data$OUTCOME_VARIABLE[
  experiment_data$TREATMENT_VARIABLE == 0],
na.rm = TRUE)
```

## Activity 3 — interpret

### Activité 3 — interpréter

Compare your ATE to the **scale** of the outcome variable. Is the difference large or small?

Record your answer as a comment in your .R file (# Answer: ...).

Comparez votre ATE à l'**échelle** de la variable de résultat. La différence est-elle grande ou petite ?

Notez votre réponse en commentaire dans votre fichier .R (# Réponse : ...).

## Section 6

### **Activity 4**

## Power analysis — goals

### *Analyse de puissance — objectifs*

**Context:** Sample size is a key design choice.

**Power** = probability of detecting an effect if one exists.

This activity uses the **EGAP Power Calculator** (web app — no R code to copy). Record answers as comments in your `.R` file.

**Contexte :** La taille d'échantillon est un choix de conception clé. **Puissance** = probabilité de détecter un effet s'il existe.

Cette activité utilise le **calculateur de puissance EGAP** (application web — pas de code R à copier). Notez les réponses en commentaires dans votre fichier `.R`.

## Activity 4 — calculator

### *Activité 4 — calculateur*

Open the calculator:

[egap.shinyapps.io/power-app](http://egap.shinyapps.io/power-app)

Key inputs: significance level (0.05), expected effect size, SD of outcome, power target (80%), max  $N$  for the graph.

Ouvrez le calculateur :

[egap.shinyapps.io/power-app](http://egap.shinyapps.io/power-app)

Entrées clés : niveau de significativité (0,05), taille d'effet attendue, écart-type du résultat, cible de puissance (80 %),  $N$  max pour le graphique.

## Activity 4 — tasks

### Activité 4 — tâches

- 1 Effect size = 1, SD = 1: power at  $N = 100$  and  $N = 500$ ?
  - 2 **Double** the effect size — power up or down?
  - 3 **Double** the SD — power up or down?
  - 4 With  $N = 500$  fixed, what is the **smallest detectable** effect size at 80% power?
- 1 Taille d'effet = 1, ET = 1 : puissance pour  $N = 100$  et  $N = 500$  ?
  - 2 **Doublez** la taille d'effet — la puissance augmente ou diminue ?
  - 3 **Doublez** l'écart-type — la puissance augmente ou diminue ?
  - 4 Avec  $N = 500$  fixé, quelle est la **plus petite taille d'effet détectable** à 80 % de puissance ?

## Section 7

### **Activity 5**

**Context:** Different **answer strategies** give different answers. The right estimator depends on how you **randomized** (complete vs blocked).

**Contexte :** Différentes **stratégies d'analyse** donnent des réponses différentes. Le bon estimateur dépend de la façon dont vous avez **randomisé** (complète vs par blocs).

## Activity 5 — load data

### *Activité 5 — charger les données*

```
experiment_data <- read.csv("data_for_analysis.csv")  
library(estimatr)
```

## Activity 5 — complete randomization

### Activité 5 — randomisation complète

If you used **complete** randomization, analyze with `difference_in_means()` **without** blocks. Replace outcome and treatment variable names. Compare to your Activity 3 hand calculation.

Si vous avez utilisé une randomisation **complète**, analysez avec `difference_in_means()` **sans** blocs.

Remplacez les noms de variables. Comparez avec votre calcul manuel de l'activité 3.

```
difference_in_means(  
  OUTCOME_VARIABLE ~ TREATMENT_VARIABLE,  
  data = experiment_data  
)
```

## Activity 5 — blocked randomization

### Activité 5 — randomisation par blocs

If assignment was **blocked** on Gender, add `blocks = Gender`. Compare **Estimate** and **Std. Error** to the complete-randomization model. They should differ — which matches your design?

```
difference_in_means(  
  OUTCOME_VARIABLE ~ TREATMENT_VARIABLE,  
  blocks = Gender,  
  data = experiment_data  
)
```

Si l'assignation était **stratifiée** par Gender, ajoutez `blocks = Gender`. Comparez **Estimate** et **Std. Error** au modèle sans blocs. Ils diffèrent — lequel correspond à votre design ?

## Activity 5 — takeaway

### *Activité 5 — message clé*

**Analyze as you randomize.** Choose the estimator that matches your data strategy.

**Analysez comme vous randomisez.** Choisissez l'estimateur qui correspond à votre stratégie de données.

Section 8

**Wrap-up**

## Next steps

### *Prochaines étapes*

- Save your .R file — it is your complete record of the session.
  - Revisit lecture slides on randomization, estimation, and power if anything is unclear.
  - All activities use the same `data_for_analysis.csv` in your exercise folder.
- Sauvegardez votre fichier .R — c'est votre trace complète de la session.
  - Revenez aux slides de cours sur randomisation, estimation et puissance si besoin.
  - Toutes les activités utilisent le même `data_for_analysis.csv` dans votre dossier d'exercices.