

## 88 % de présence, 62 % des élèves

L'assiduité moyenne quotidienne vaut 88,4 %. La part des élèves présents au moins 90 % de leurs jours pointés vaut 62,1 %. Les deux viennent des mêmes 70 245 lignes, et les vingt-six points d'écart sont les enfants qu'une moyenne ne voit pas.

---

Pierre Robentz Cassion

Leçon 3 sur 8

Analyse de programmes d'éducation — Inscrit n'est pas présent

## Ce que couvre cette leçon

- Un registre, deux questions
- L'assiduité moyenne quotidienne
- La proportion d'élèves assidus
- Sur lequel un programme est jugé
- Énoncez le seuil, et d'où il vient
- La décision de dénominateur que personne ne prend explicitement
- Rapportez la paire
- La suite

# Un registre, deux questions — En Python

```
import pandas as pd

attendance = pd.read_csv("school-attendance-2024.v1.csv")
print(f"{len(attendance):,} rows, {attendance['student_id'].nunique():,} students")
print(attendance["present"].value_counts(dropna=False))
```

## Un registre, deux questions — En R

```
library(dplyr)
attendance |> count(present)
```

## Un registre, deux questions

- La colonne `present` a cinq valeurs, non deux — SCH09 a saisi Y et N au lieu de `true` et `false` sur 873 de ses...

## Un registre, deux questions — En Python

```
MARKS = {"true": True, "Y": True, "false": False, "N": False}
marked = attendance[attendance["present"].isin(MARKS)].copy()
marked["attended"] = marked["present"].map(MARKS)
print(f"usable marks: {len(marked):,} of {len(attendance):,}")
```

## Un registre, deux questions — En R

```
attendance |>
  mutate(attended = case_when(present %in% c("true", "Y") ~ TRUE,
                              present %in% c("false", "N") ~ FALSE)) |>
  filter(!is.na(attended))
```

## L'assiduité moyenne quotidienne — En Python

```
ada = marked["attended"].mean()  
print(f"average daily attendance: {ada:.1%}")
```

## L'assiduité moyenne quotidienne — En R

```
marked |> summarise(ada = mean(attended))
```

## L'assiduité moyenne quotidienne

- 88,4 % — Chaque pointage compte une fois, si bien qu'un élève présent 60 jours sur 60 et un élève présent 30 sur 60...

## La proportion d'élèves assidus — En Python

```
by_student = marked.groupby("student_id")["attended"].agg(["mean", "size"])
eligible = by_student[by_student["size"] >= 20]

for threshold in (0.80, 0.85, 0.90):
    share = (eligible["mean"] >= threshold).mean()
    print(f"attending at least {threshold:.0%} of marked days: {share:.1%}")
```

## La proportion d'élèves assidus — En R

```
marked |>  
  summarise(rate = mean(attended), days = n(), .by = student_id) |>  
  filter(days >= 20) |>  
  summarise(across(everything(), ~ mean(rate >= 0.9)))
```

## La proportion d'élèves assidus

| Seuil                   | Élèves l'atteignant |
|-------------------------|---------------------|
| Au moins 80 % des jours | 87,3 %              |
| Au moins 85 %           | 78,9 %              |
| <b>Au moins 90 %</b>    | <b>62,1 %</b>       |

## La proportion d'élèves assidus

- 62,1 % au seuil de 90 %, contre une assiduité moyenne quotidienne de 88,4 % — Les deux nombres sont écartés de. . .
- Une moyenne sur des pointages décrit le système ; une proportion sur des élèves décrit des enfants — Une école où. . .

## Sur lequel un programme est jugé

- Rapportez les deux, et mettez en tête celui qui correspond à la décision

## Sur lequel un programme est jugé

| La décision                                    | Le nombre                                        |
|------------------------------------------------|--------------------------------------------------|
| Combien de repas, de pupitres, de manuels      | Assiduité moyenne quotidienne                    |
| Quels enfants ont besoin d'une visite de suivi | Proportion sous le seuil                         |
| Si une intervention a marché                   | Les deux, car ils peuvent bouger en sens opposés |

## Sur lequel un programme est jugé — En Python

```
struggling = eligible[eligible["mean"] < 0.75]
print(f"students below 75%: {len(struggling)} "
      f"({len(struggling) / len(eligible):.1%})")
print(f"their marks as a share of all marks: "
      f"{struggling['size'].sum() / eligible['size'].sum():.1%}")
```

## Sur lequel un programme est jugé — En R

```
# The students furthest behind are a small share of the rows and the whole  
# of the problem.
```

## Énoncez le seuil, et d'où il vient — En Python

```
sensitivity = {f"{t:.0%}": f"{(eligible['mean'] >= t).mean():.1%}"  
               for t in (0.75, 0.80, 0.85, 0.90, 0.95)}  
print(sensitivity)
```

## Énoncez le seuil, et d'où il vient — En R

*# Print the curve, not the point.*

## La décision de dénominateur que personne ne prend explicitement

- Sans lui — un élève inscrit la dernière semaine du trimestre et présent trois jours sur trois apparaît comme assidu à . . .
- Avec lui — vous avez exclu précisément les élèves arrivés tard et partis tôt, qui sont ceux qu'un programme d'assiduité. . .

## La décision de dénominateur que personne ne prend explicitement — En Python

```
print(f"students with any marks:      {len(by_student)}")  
print(f"students with 20 or more:    {len(eligible)}")
```

# La décision de dénominateur que personne ne prend explicitement — En R

*# Two denominators, both defensible, and the report says which.*

## Rapportez la paire — Exemple

Attendance, February to April 2024

|                                 |       |                                  |
|---------------------------------|-------|----------------------------------|
| Average daily attendance        | 88.4% | 69,974 usable marks              |
| Students attending 90%+ of days | 62.1% | 1,200 students with 20+ marks    |
| Students attending 85%+         | 78.9% |                                  |
| Students below 75%              | 9.5%  | 114 students, the follow-up list |

SCH09 recorded 873 marks as Y/N rather than true/false; recoded, not dropped. 271 days were never marked and are excluded from both figures.

- Les deux nombres de cette leçon divisent par des jours qui ont été pointés.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/education-programme-analysis/  
edu-03-two-attendance-numbers](https://data-analysis.cassion.dev/fr/cours/education-programme-analysis/edu-03-two-attendance-numbers)