

En retard, et de plus en plus

Le retard d'âge vaut 17,0 % en première année et 50,8 % en sixième. Il monte parce que le redoublement le crée — 94,5 % des enfants ayant redoublé en 2023 sont en retard d'âge en 2024 — et parce qu'être en retard quadruple le risque de partir.

Pierre Robentz Cassion

Leçon 2 sur 8

Analyse de programmes d'éducation — L'inscription et son dénominateur

Ce que couvre cette leçon

- La définition, et le contrôle d'étendue qui doit venir avant
- Il se compose avec l'année d'études
- Pourquoi cela compte : le lien avec l'abandon
- La désagrégation qui ne donne rien
- Rapportez le profil, non le chiffre principal
- La suite

La définition, et le contrôle d'étendue qui doit venir avant — En Python

```
import pandas as pd

enrolment = pd.read_csv("school-enrolment-2024.v1.csv")
print(enrolment["age_years"].describe().round(1))
print(f"ages above 20: {(enrolment['age_years'] > 20).sum()}")
```

La définition, et le contrôle d'étendue qui doit venir avant — En R

```
library(dplyr)
enrolment |> summarise(min = min(age_years), max = max(age_years),
                      impossible = sum(age_years > 20))
```

La définition, et le contrôle d'étendue qui doit venir avant

- Vingt et une lignes portent un âge supérieur à 20 dans un registre du primaire — SCH04 enregistre une partie de ses. . .

La définition, et le contrôle d'étendue qui doit venir avant — En Python

```
clean = enrolment[enrolment["age_years"] <= 20].copy()
clean["official_age"] = clean["grade"] + 5
clean["over_age"] = clean["age_years"] > clean["official_age"]
```

La définition, et le contrôle d'étendue qui doit venir avant — En R

```
enrolment |> filter(age_years <= 20) |>  
  mutate(official_age = grade + 5, over_age = age_years > official_age)
```

La définition, et le contrôle d'étendue qui doit venir avant

- Faites le contrôle d'étendue avant l'indicateur, non après — Un âge de 112 est en retard pour toutes les années, si...

Il se compose avec l'année d'études — En Python

```
primary = clean[(clean["school_year"] == 2024) & clean["grade"].between(1, 6)]
by_grade = primary.groupby("grade").agg(
    students=("student_id", "size"),
    over_age=("over_age", "mean"),
)
print((by_grade * [1, 100]).round(1))
```

Il se compose avec l'année d'études — En R

```
primary |> summarise(n = n(), over_age = mean(over_age), .by = grade)
```

Il se compose avec l'année d'études

Année	Âge officiel	Élèves	En retard d'âge
1	6	53	17,0 %
2	7	153	25,5 %
3	8	247	29,6 %
4	9	258	36,8 %
5	10	209	47,8 %
6	11	132	50,8 %

Il se compose avec l'année d'études

- Le retard d'âge triple entre la première et la sixième année — Deux mécanismes produisent cette forme et ce ne sont pas. . .
- L'entrée tardive — place un enfant au-dessus de l'âge dès la première année et l'y maintient
- Le redoublement — crée le retard d'âge *pendant* la scolarité

Il se compose avec l'année d'études — En Python

```
previous = clean[clean["school_year"] == 2023].set_index("student_id")
current = clean[clean["school_year"] == 2024].copy()
current["last_year"] = current["student_id"].map(previous["end_of_year_status"])

print(current.groupby("last_year")["over_age"].agg(["mean", "size"]).round(3))
```

Il se compose avec l'année d'études — En R

```
enrolment |>  
  filter(age_years <= 20) |>  
  select(student_id, school_year, grade, age_years, end_of_year_status) |>  
  tidyr::pivot_wider(names_from = school_year,  
                     values_from = c(grade, age_years, end_of_year_status))
```

Il se compose avec l'année d'études

- 94,5 % des élèves ayant redoublé en 2023 sont en retard d'âge en 2024, contre 30,6 % de ceux qui ont été promus — Le...

Pourquoi cela compte : le lien avec l'abandon — En Python

```
year23 = clean[clean["school_year"] == 2023]
dropout = year23.groupby("over_age")["end_of_year_status"].apply(
    lambda s: (s == "dropped-out").mean()
)
counts = year23.groupby("over_age").size()
print(pd.DataFrame({"dropout": (dropout * 100).round(1), "n": counts}))
```

Pourquoi cela compte : le lien avec l'abandon — En R

```
enrolment |> filter(school_year == 2023) |>  
  summarise(dropout = mean(end_of_year_status == "dropped-out"),  
            n = n(), .by = over_age)
```

Pourquoi cela compte : le lien avec l'abandon

	Élèves	Ont abandonné
En retard d'âge	478	19,5 %
À l'âge attendu	814	4,8 %

Pourquoi cela compte : le lien avec l'abandon

- Être en retard quadruple le risque de partir — et c'est ce qui boucle la chaîne : un enfant redouble, prend du retard. . .
- Le redoublement n'est donc pas une intervention neutre — Il est présenté comme une seconde chance et il augmente de. . .

La désagrégation qui ne donne rien — En Python

```
for column in ("sex", "disability_reported", "displacement_status"):
    print(primary.groupby(column, dropna=False)["over_age"].agg(
        ["mean", "size"]).round(3), "\n")
```

La désagrégation qui ne donne rien — En R

```
primary |> summarise(over_age = mean(over_age), n = n(), .by = sex)
```

La désagrégation qui ne donne rien

Ventilation	En retard d'âge	n
Garçons	39,2 %	510
Filles	33,8 %	542
Handicap signalé	33,7 %	89
Sans handicap signalé	36,7 %	963

La désagrégation qui ne donne rien

- Cinq points entre garçons et filles, sur 510 et 542 élèves — C'est à la limite de ce que produit la variation. . .
- Le handicap montre trois points en sens inverse sur 89 élèves — ce qui est du bruit sur un dénominateur aussi petit
- Une désagrégation qui ne montre presque rien est un résultat — Le retard d'âge est ici un phénomène de système porté. . .

Rapportez le profil, non le chiffre principal — Exemple

Over-age enrolment, primary, 2024

Overall	36.4%	383 of 1,052 primary enrollees
Grade 1	17.0%	the late-entry floor
Grade 6	50.8%	after five years of repetition
Repeated in 2023	94.5%	over-age in 2024, against 30.6% promoted
Dropout, over-age	19.5%	against 4.8% in-age (2023 cohort)

21 ages recorded in months at SCH04, corrected by division.

Sex difference is 39.2% (boys) against 33.8% (girls); report with an interval or not at all.

Rapportez le profil, non le chiffre principal

- Le profil par année est le livrable — car 36,4 % au total pourrait signifier une entrée tardive uniforme ou un...

- Tous les enfants de cette leçon sont inscrits.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/education-programme-analysis/
edu-02-the-over-age-student](https://data-analysis.cassion.dev/fr/cours/education-programme-analysis/edu-02-the-over-age-student)