

Taux d'attaque et le 1 % que personne n'atteint

Le taux d'attaque exige la population ; la létalité exige un dénominateur qui exclut le cas encore hospitalisé. 3,90 % contre une cible de 1 %, 6,30 % dans un district — et l'explication repose sur une statistique de délai que la leçon précédente a montrée cassée.

Pierre Robentz Cassion

Leçon 4 sur 8

Santé publique et épidémiologie appliquée — Une épidémie, un cas à la fois

Ce que couvre cette leçon

- Le taux d'attaque
- Taux d'attaque par âge
- La létalité
- Le seuil
- L'explication, et pourquoi elle est indisponible
- Le bloc de rapportage
- La suite

Le taux d'attaque — En Python

```
import pandas as pd

cases = pd.read_csv("cholera-line-list-2024.v1.csv")
population = pd.read_csv("district-population-2024.v1.csv")

by_district = (
    cases.groupby("district").size().rename("cases")
    .to_frame()
    .join(population.groupby("district")["population"].sum())
)
by_district["per_1000"] = 1000 * by_district["cases"] / by_district["population"]
print(by_district.round(2))
```

Le taux d'attaque — En R

```
library(dplyr)

cases |> count(district, name = "cases") |>
  left_join(summarise(population, population = sum(population), .by = district),
            by = "district") |>
  mutate(per_1000 = 1000 * cases / population)
```

Le taux d'attaque

District	Cas	Population	Pour 1 000
Nord	381	48 000	7,94
Centre	401	62 000	6,47
Sud	193	35 000	5,51

Taux d'attaque par âge — En Python

```
by_band = (  
    cases.groupby(["district", "age_band"]).size().rename("cases")  
    .to_frame()  
    .join(population.set_index(["district", "age_band"])["population"])  
)  
by_band["per_1000"] = 1000 * by_band["cases"] / by_band["population"]  
print(by_band["per_1000"].unstack().round(2))
```

Taux d'attaque par âge — En R

```
cases |> count(district, age_band, name = "cases") |>  
  left_join(population, by = c("district", "age_band")) |>  
  mutate(per_1000 = 1000 * cases / population) |>  
  tidyr::pivot_wider(id_cols = age_band, names_from = district, values_from = per_1000)
```

Taux d'attaque par âge

Tranche d'âge	Nord	Centre	Sud
0-4	14,77	12,69	11,69
5-14	8,26	8,06	6,29
15-44	3,93	4,03	3,51
45+	6,41	4,75	5,71

Taux d'attaque par âge

- Les moins de cinq ans sont trois à quatre fois plus touchés que les adultes en âge de travailler — dans chaque district

```
with_outcome = cases[cases["outcome"].notna()]

print(f"cases: {len(cases)}, with an outcome: {len(with_outcome)}")
print(f"CFR: {(with_outcome['outcome'] == 'died').mean():.2%}")
```

```
cases |> filter(!is.na(outcome)) |>  
  summarise(n = n(), cfr = mean(outcome == "died"))
```

- 3,90 % sur 974 cas à issue enregistrée

```
TARGET = 0.01
cfr = (with_outcome["outcome"] == "died").mean()
print(f"CFR {cfr:.2%} against a target of {TARGET:.0%}: "
      f"{'meets' if cfr < TARGET else 'does not meet'} the standard")

by_district_cfr = with_outcome.groupby("district")["outcome"].apply(
    lambda s: (s == "died").mean()
)
print((by_district_cfr * 100).round(2))
```

```
cases |> filter(!is.na(outcome)) |>  
  summarise(cfr = mean(outcome == "died"), n = n(), .by = district)
```

District	Létalité	n
Nord	6,30 %	381
Sud	3,11 %	193
Centre	2,00 %	400

L'explication, et pourquoi elle est indisponible — En Python

```
delay = (  
    pd.to_datetime(cases["admission_date"], errors="coerce")  
    - pd.to_datetime(cases["onset_date"], errors="coerce")  
) .dt.days  
  
testable = cases.assign(delay=delay).dropna(subset=["delay", "outcome"])  
banded = testable.assign(  
    band=pd.cut(testable["delay"], [-1, 1, 3, 99], labels=["0-1", "2-3", "4+"])  
)  
print(banded.groupby("band")["outcome"].apply(lambda s: (s == "died").mean()).round(3))
```

```
cases |>
  filter(!is.na(onset_date), !is.na(admission_date), !is.na(outcome)) |>
  mutate(delay = as.integer(admission_date - onset_date),
         band = cut(delay, c(-1, 1, 3, 99), labels = c("0-1", "2-3", "4+"))) |>
  summarise(cfr = mean(outcome == "died"), n = n(), .by = band)
```

L'explication, et pourquoi elle est indisponible

- La raison de fond — l'admission elle-même est la protection — Les cas jamais admis portent l'essentiel de la mortalité, . . .
- Celle qui n'en est pas — le délai de Nord est fabriqué par son registre — La leçon précédente a trouvé 80 % de ses cas. . .

L'explication, et pourquoi elle est indisponible — En Python

```
print("Nord CFR 6.30% with a median recorded delay of 0 days.")  
print("The delay is a register artefact; the CFR is not.")
```

L'explication, et pourquoi elle est indisponible — En R

One of these two numbers is real. Say which, in the report.

L'explication, et pourquoi elle est indisponible

- Consignez cela comme une limite, non comme un résultat — « La létalité est la plus élevée à Nord ; le délai entre début. . .

Le bloc de rapportage — Exemple

Cholera outbreak, weeks 1-16

Cases	975	attack rate 6.7 per 1,000 (145,000 population)
Attack rate by district	7.94 / 6.47 / 5.51 per 1,000 (Nord / Centre / Sud)	crude; see standardised rates before comparing
Deaths	38	case fatality 3.90%
Denominator	974	one case still admitted at cut-off, excluded
Against Sphere target	1%	not met in any district

Highest CFR: Nord at 6.30% (n=381). Onset-to-admission delay is not usable in Nord (80% of cases record onset = admission), so the usual explanation cannot be tested there from this register.

- Les taux d'attaque par district ressemblent à une comparaison et n'en sont pas encore une, car les districts ne contiennent pas les mêmes gens.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/public-health-epidemiology/
epi-04-attack-rates-and-cfr](https://data-analysis.cassion.dev/fr/cours/public-health-epidemiology/epi-04-attack-rates-and-cfr)