

Taux, rapport, proportion — et personnes-temps

Trois formes que l'on appelle toutes « taux » et une seule l'est, plus le dénominateur qui donne son sens à l'incidence quand les gens entrent et sortent d'une population à des moments différents.

Pierre Robentz Cassion

Leçon 1 sur 8

Santé publique et épidémiologie appliquée — Le dénominateur, c'est l'épidémiologie

Ce que couvre cette leçon

- Trois formes, un seul mot
- Prévalence et incidence
- Ce qu'apportent les personnes-temps
- Mettez le dénominateur et le multiplicateur dans le nom
- La population à risque n'est pas toujours toute la population
- Le contrôle à faire sur tout taux
- La suite

Trois formes, un seul mot

Forme	Structure	Plage	Exemple
Proportion	Partie sur tout, mêmes unités	0 à 1	Létalité — décès parmi les cas
Rapport	Deux quantités non emboîtées	quelconque	Sex-ratio des cas ; rapport de cotes
Taux	Événements sur population- temps	positif	Incidence pour 1 000 personnes-années

Trois formes, un seul mot

- Seule la troisième est un taux — et la différence est le temps au dénominateur

- **La prévalence** — combien de personnes ont l'affection *maintenant*. Une proportion, issue d'une enquête, un. . .
- **L'incidence** — combien de *nouveaux* cas surviennent par unité de population-temps. Un taux, issu de la. . .

Prévalence et incidence — En Python

```
import pandas as pd

cases = pd.read_csv("cholera-line-list-2024.v1.csv", parse_dates=["onset_date"])
population = pd.read_csv("district-population-2024.v1.csv")

total_population = population["population"].sum()
print(f"{len(cases)} cases in a population of {total_population:,}")
print(f"attack rate over the outbreak: {1000 * len(cases) / total_population:.2f} per 1,000")
```

Prévalence et incidence — En R

```
library(dplyr)

cases <- readr::read_csv("cholera-line-list-2024.v1.csv")
population <- readr::read_csv("district-population-2024.v1.csv")

c(cases = nrow(cases),
  population = sum(population$population),
  per_1000 = 1000 * nrow(cases) / sum(population$population))
```

Ce qu'apportent les personnes-temps — En Python

```
# One person followed for 6 months contributes 0.5 person-years
follow_up = pd.DataFrame({
    "person_id": ["A", "B", "C"],
    "entered": pd.to_datetime(["2024-01-01", "2024-01-01", "2024-07-01"]),
    "exited": pd.to_datetime(["2024-12-31", "2024-06-30", "2024-12-31"]),
})
follow_up["person_years"] = (
    (follow_up["exited"] - follow_up["entered"]).dt.days / 365.25
)

events = 1
print(f"{events / follow_up['person_years'].sum():.3f} events per person-year")
print(f"naive headcount rate: {events / len(follow_up):.3f} per person")
```

Ce qu'apportent les personnes-temps — En R

```
follow_up <- tibble::tibble(  
  person_id = c("A", "B", "C"),  
  entered = as.Date(c("2024-01-01", "2024-01-01", "2024-07-01")),  
  exited  = as.Date(c("2024-12-31", "2024-06-30", "2024-12-31"))  
) |>  
  mutate(person_years = as.numeric(exited - entered) / 365.25)  
  
c(per_person_year = 1 / sum(follow_up$person_years),  
  naive = 1 / nrow(follow_up))
```

Ce qu'apportent les personnes-temps

- **le suivi varie** — une cohorte de traitement où l'on s'inscrit tout au long de l'année ;
- **la population change** — déplacements, arrivées en camp, bassin qui grandit ;
- **vous comparez des périodes de longueurs différentes** — une réponse de neuf mois face à une de douze.

Mettez le dénominateur et le multiplicateur dans le nom — En Python

```
indicators = {  
    "cholera_attack_rate_per_1000_outbreak": 1000 * len(cases) / total_population,  
    "case_fatality_percent_of_cases_with_outcome": None,  
    "incidence_per_1000_person_years": None,  
}
```

Mettez le dénominateur et le multiplicateur dans le nom — En R

The same discipline: the name carries the denominator and the multiplier

Mettez le dénominateur et le multiplicateur dans le nom

- Le multiplicateur — Pour 100, pour 1 000, pour 100 000
- La population du dénominateur — « Pour 1 000 habitants » et « pour 1 000 enfants de moins de cinq ans » sont deux. . .

La population à risque n'est pas toujours toute la population

- **Le ratio de mortalité maternelle** — décès pour 100 000 *naissances vivantes*, non pour la population, car seules les. . .
- **La mortalité néonatale** — décès dans les 28 premiers jours pour 1 000 *naissances vivantes*.
- **Le taux d'attaque du choléra** — cas pour 1 000 habitants, car tout le monde est à risque.
- **Le taux d'attaque de la rougeole** — sans doute pour 1 000 personnes *réceptives*, et la population réceptive est. . .

La population à risque n'est pas toujours toute la population — En Python

```
under_five = population.loc[population["age_band"] == "0-4", "population"].sum()
u5_cases = (cases["age_band"] == "0-4").sum()

print(f"under-five attack rate: {1000 * u5_cases / under_five:.2f} per 1,000")
print(f"all-age attack rate:      {1000 * len(cases) / total_population:.2f} per 1,000")
```

La population à risque n'est pas toujours toute la population — En R

```
under5 <- population |> filter(age_band == "0-4") |> summarise(sum(population)) |> pull()
u5_cases <- sum(cases$age_band == "0-4")

c(under_five = 1000 * u5_cases / under5,
  all_age = 1000 * nrow(cases) / sum(population$population))
```

Le contrôle à faire sur tout taux

- Que compte le numérateur — des événements, des personnes, des épisodes ?
- Qui est au dénominateur, et tous étaient-ils à risque ?
- Sur quelle période, et tous y étaient-ils présents ?
- Quel est le multiplicateur, et est-ce la convention de cet indicateur ?

Le contrôle à faire sur tout taux — En Python

```
def describe_rate(numerator, denominator, period, multiplier, at_risk_note):  
    return {  
        "value": multiplier * numerator / denominator,  
        "numerator": numerator, "denominator": denominator,  
        "period": period, "multiplier": multiplier, "at_risk": at_risk_note,  
    }
```

Le contrôle à faire sur tout taux — En R

```
describe_rate <- function(numerator, denominator, period, multiplier, at_risk) {  
  list(value = multiplier * numerator / denominator, numerator = numerator,  
        denominator = denominator, period = period, at_risk = at_risk)  
}
```

- Un dénominateur est simple.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/public-health-epidemiology/
epi-01-rates-and-denominators](https://data-analysis.cassion.dev/fr/cours/public-health-epidemiology/epi-01-rates-and-denominators)