

# Les périodes, et l'opérateur d'agrégation que personne ne règle

Un fichier, un indicateur, trois taux de couverture annuels — 6,5 %, 60,8 % et 77,5 % — tous issus de phrases qui sonnent défendables. La différence est la manière dont le dénominateur s'agrège dans le temps.

---

Pierre Robentz Cassion

Leçon 3 sur 8

Données de routine et DHIS2 — Périodes et complétude

## Ce que couvre cette leçon

- Types de période, et l'identifiant qui les porte
- À quelle période appartient une valeur ?
- Les trois chiffres annuels
- Pourquoi ils diffèrent
- L'opérateur d'agrégation
- Échéances, expiration et verrouillage
- La complétude est un enregistrement à part
- La suite

## Types de période, et l'identifiant qui les porte — Exemple

|           |                               |
|-----------|-------------------------------|
| 202408    | monthly                       |
| 2024W32   | weekly                        |
| 2024Q3    | quarterly                     |
| 2024      | yearly                        |
| 2024April | financial year starting April |

## Types de période, et l'identifiant qui les porte — En Python

```
import pandas as pd

vax = pd.read_csv("vaccination-coverage-2024.v1.csv", parse_dates=["period"])
vax["period_id"] = vax["period"].dt.strftime("%Y%m")
print(sorted(vax["period_id"].unique())[:4])
```

## Types de période, et l'identifiant qui les porte — En R

```
library(dplyr)
vax <- vax |> mutate(period_id = format(period, "%Y%m"))
```

## À quelle période appartient une valeur ?

- **Juste** — la période où l'activité a eu lieu. Une dose administrée le 28 août est d'août, quelle que soit la date. . .
- **Faux** — la période où le formulaire a été transmis. Fréquent là où la saisie se fait par lots, et cela décale. . .
- **Faux** — la période où la valeur a été saisie. Même défaillance, un cran plus loin.

## Les trois chiffres annuels — En Python

```
penta3 = vax[vax["antigen"] == "penta3"]
reported = penta3[penta3["report_submitted"] == True]

doses = reported["doses_administered"].sum()

summed_denominator = reported["target_population"].sum()
annual_target = penta3.groupby("facility_id")["target_population"].first().sum()

print(f"doses: {doses:,}")
print(f"a) doses / summed monthly target : {doses / summed_denominator:.1%}")
print(f"b) doses / annual target, all 38 : {doses / annual_target:.1%}")
```

## Les trois chiffres annuels — En R

```
penta3 <- vax |> filter(antigen == "penta3")
reported <- penta3 |> filter(report_submitted)

doses <- sum(reported$doses_administered)
summed <- sum(reported$target_population)
annual <- penta3 |> summarise(t = first(target_population), .by = facility_id) |>
  summarise(sum(t)) |> pull()

c(a = doses / summed, b = doses / annual)
```

## Les trois chiffres annuels

| Phrase                                                                                  | Chiffre       |
|-----------------------------------------------------------------------------------------|---------------|
| « Doses sur population cible, sommée sur l'année »                                      | <b>6,5 %</b>  |
| « Doses annuelles sur population cible annuelle »                                       | <b>60,8 %</b> |
| « Doses annuelles sur cible annuelle, parmi les couples formation-mois ayant rapporté » | <b>77,5 %</b> |

## Pourquoi ils diffèrent

- 6,5 % est un chiffre mensuel portant une étiquette annuelle — `target_population` est une cohorte *annuelle* — les...
- 60,8 % compte les formations muettes comme n'ayant vacciné personne — Le dénominateur est la cohorte annuelle complète...
- 77,5 % est le défendable — et il exige un ajustement explicite du dénominateur

## Pourquoi ils diffèrent — En Python

```
months_reported = reported.groupby("facility_id").size()
targets = penta3.groupby("facility_id")["target_population"].first()
prorated = (targets * months_reported.reindex(targets.index).fillna(0) / 12).sum()

print(f"c) pro-rated denominator: {doses / prorated:.1%}")
```

## Pourquoi ils diffèrent — En R

```
months <- reported |> summarise(m = n(), .by = facility_id)
targets <- penta3 |> summarise(t = first(target_population), .by = facility_id)

prorated <- targets |>
  left_join(months, by = "facility_id") |>
  mutate(m = coalesce(m, 0)) |>
  summarise(sum(t * m / 12)) |> pull()

doses / prorated
```

# L'opérateur d'agrégation

| Opérateur       | Sens                       | Convient à                                                               |
|-----------------|----------------------------|--------------------------------------------------------------------------|
| Somme           | Additionner entre périodes | Décomptes d'événements — doses, admissions, consultations                |
| Moyenne         | Moyenne entre périodes     | Niveaux de stock, effectifs, tout ce qui est un état et non un événement |
| Dernière valeur | Prendre la plus récente    | Populations, cibles, tailles de registre                                 |

- Demandez l'opérateur d'agrégation avec les métadonnées — C'est un champ par élément de données, invisible dans. . .

- **L'échéance** est la date à laquelle l'ensemble de données est dû. Elle alimente le chiffre de promptitude que calcule. . .
- **Les jours d'expiration** verrouillent le formulaire un nombre fixe de jours après la fin de la période. Ensuite, la. . .
- **Une exception de verrouillage** déverrouille un ensemble, une unité, une période. Chacune est une décision, et un. . .

## La complétude est un enregistrement à part

- La complétude dans DHIS2 est un enregistrement, non un calcul — Quand un utilisateur clique « complet » sur un...

## La complétude est un enregistrement à part — En Python

```
grid = (vax["facility_id"].nunique() * vax["period"].nunique()
        * vax["antigen"].nunique())
print(f"{grid} rows expected, {len(vax)} present, "
      f"{(vax['report_submitted'] == True).sum()} flagged reported")
```

## La complétude est un enregistrement à part — En R

```
c(expected = n_distinct(vax$facility_id) * n_distinct(vax$period) * n_distinct(vax$antigen),  
  present = nrow(vax),  
  reported = sum(vax$report_submitted))
```

- L'indicateur de rapportage est la matière première du chiffre de complétude, et la leçon suivante en fait un dénominateur.

Lire la leçon complète, avec le code exécutable

`https://data-analysis.cassion.dev/fr/cours/routine-data-and-dhis2/  
dhis2-03-periods-and-datasets`