

Les dossiers non clos sont les longs

Le délai médian de clôture vaut quatre mois sur les 636 dossiers clos. 472 autres sont encore ouverts, dont 140 dépassent déjà quatre mois, et la protection de l'enfance est censurée à 58 % contre 32 % pour la protection générale — la même statistique signifie donc des choses différentes selon la colonne.

Pierre Robentz Cassion

Leçon 6 sur 8

Données de protection et VBG — La charge de dossiers

Ce que couvre cette leçon

- Le calcul évident est biaisé
- Corrigez d'abord les durées impossibles
- Trois façons honnêtes de le rapporter
- La censure n'est pas répartie uniformément
- Le motif de clôture est un jugement
- Rapportez-le en entier
- La suite

Le calcul évident est biaisé — En Python

```
import pandas as pd

cases = pd.read_csv("protection-case-management-2024.v1.csv")

opened = cases["opened_month"].str[5:].astype(int)
closed = cases["closed_month"].str[5:].astype("Int64")
duration = closed - opened

print(f"closed cases: {duration.notna().sum()}")
print(f"median months to closure: {duration.median()}")
print(f"still open: {duration.isna().sum()} ({duration.isna().mean():.1%})")
```

Le calcul évident est biaisé — En R

```
library(dplyr)

cases |>
  mutate(opened = as.integer(substr(opened_month, 6, 7)),
         closed = as.integer(substr(closed_month, 6, 7)),
         duration = closed - opened) |>
  summarise(closed = sum(!is.na(duration)),
            median = median(duration, na.rm = TRUE),
            open = sum(is.na(duration)))
```

Le calcul évident est biaisé

- Quatre mois, sur 636 dossiers sur 1 108 — Les 472 autres sont encore ouverts à la clôture de décembre, et ils ne sont. . .

Le calcul évident est biaisé — En Python

```
still_open = 12 - opened[duration.isna()] + 1  
print(f"open cases already past 4 months: {(still_open > 4).sum()}")  
print(f"longest open: {still_open.max()} months and counting")
```

Le calcul évident est biaisé — En R

```
cases |> filter(is.na(closed_month)) |>  
  mutate(so_far = 12 - as.integer(substr(opened_month, 6, 7)) + 1) |>  
  summarise(past_four = sum(so_far > 4), longest = max(so_far))
```

Le calcul évident est biaisé

- 140 des dossiers ouverts le sont déjà depuis plus longtemps que le dossier clos médian — et l'un l'est depuis onze mois

Corrigez d'abord les durées impossibles — En Python

```
impossible = duration.notna() & (duration <= 0)
print(f"cases closing on or before the month they opened: {impossible.sum()}")
print(cases.loc[impossible, ["opened_month", "closed_month"]])
```

Corrigez d'abord les durées impossibles — En R

```
cases |> filter(!is.na(closed_month),  
               closed_month <= opened_month) |> nrow()
```

Corrigez d'abord les durées impossibles

- Sept dossiers se clôturent un mois antérieur à leur ouverture — La contradiction est invisible dans chaque colonne. . .

Trois façons honnêtes de le rapporter

- Rapporter la médiane avec la censure énoncée — L'option correcte la moins coûteuse

Trois façons honnêtes de le rapporter — Exemple

Median time to closure: 4 months (636 closed cases).

472 cases (42.6%) were still open at the cut-off and are excluded; 140 of them have already been open longer than 4 months, so the true median is higher.

Trois façons honnêtes de le rapporter

- Rapporter une courbe d'aboutissement par mois — Quelle part des dossiers ouverts au mois m était close dans les k ...

Trois façons honnêtes de le rapporter — En Python

```
def closed_within(k):
    eligible = cases[opened <= 12 - k]
    dur = (eligible["closed_month"].str[5:].astype("Int64")
           - eligible["opened_month"].str[5:].astype(int))
    return (dur <= k).sum() / len(eligible), len(eligible)

for k in (3, 6, 9):
    share, n = closed_within(k)
    print(f"closed within {k} months: {share:.1%} (n={n})")
```

Trois façons honnêtes de le rapporter — En R

For each k, restrict to cases with k months of possible follow-up.

Trois façons honnêtes de le rapporter

- Rapporter le délai de clôture pour une seule cohorte à suivi complet — Les dossiers ouverts en janvier ont onze mois. . .
- Les trois sont défendables et la première ne suffit pas seule — Une médiane sans énoncé de censure à côté est la. . .

La censure n'est pas répartie uniformément — En Python

```
by_category = pd.DataFrame({
    "closed": duration.notna().groupby(cases["case_category"]).sum(),
    "open": duration.isna().groupby(cases["case_category"]).sum(),
})
by_category["censored"] = by_category["open"] / by_category.sum(axis=1)
print(by_category.round(3))
```

La censure n'est pas répartie uniformément — En R

```
cases |> summarise(closed = sum(!is.na(closed_month)),  
                  open = sum(is.na(closed_month)), .by = case_category) |>  
  mutate(censored = open / (open + closed))
```

La censure n'est pas répartie uniformément

Catégorie	Clos	Ouverts	Censure	Médiane des clos
Protection de l'enfance	131	179	58 %	5 mois
VBG	255	173	40 %	4 mois
Protection générale	250	120	32 %	4 mois

La censure n'est pas répartie uniformément

- La protection de l'enfance est censurée à 58 % et la protection générale à 32 % — Les deux médianes de ce tableau sont. . .
- Une censure inégale rend une comparaison de médianes trompeuse dans un sens précis — le groupe le plus censuré est. . .

Le motif de clôture est un jugement — En Python

```
print(cases["closure_reason"].value_counts(normalize=True).round(3))
```

Le motif de clôture est un jugement — En R

```
cases |> filter(!is.na(closure_reason)) |> count(closure_reason) |>  
  mutate(share = n / sum(n))
```

Le motif de clôture est un jugement

Motif	Part des clôtures
Objectifs du plan atteints	29,1 %
Perte de contact	27,2 %
Retrait de la survivante	13,1 %
Clôture administrative	11,3 %
Déménagement	9,7 %
Transfert à une autre agence	9,6 %

Le motif de clôture est un jugement

- Seules 29,1 % des clôtures correspondent à un plan mené à terme — C'est le chiffre principal, et c'est celui par lequel. . .

Le motif de clôture est un jugement — En Python

```
by_area = pd.crosstab(cases["admin2"], cases["closure_reason"], normalize="index")  
print((by_area[["closed-administratively", "lost-contact"]] * 100).round(1))
```

Le motif de clôture est un jugement — En R

```
cases |> filter(!is.na(closure_reason)) |>  
  count(admin2, closure_reason) |> mutate(share = n / sum(n), .by = admin2)
```

Le motif de clôture est un jugement

- Hince classe 36,2 % de ses clôtures en closed-administratively contre 3,8 % à 9,8 % partout ailleurs — et son taux. . .
- Un motif de clôture est le jugement d'un travailleur social, non une observation — Une distribution qui diffère. . .

Rapportez-le en entier — Exemple (suite)

Case closure, 1,108 cases

Closed by the cut-off	636	57.4%	
Still open	472	42.6%	censored, not missing
Median months to closure	4		of closed cases only
open cases already past 4 months	140		so the true median is higher

Censoring by category: child protection 58%, GBV 40%, general 32%.

Medians are not comparable across those columns.

Closure reasons (of 636 closures)

Case plan objectives met	29.1%	
Lost contact	27.2%	
Survivor withdrew	13.1%	
Closed administratively	11.3%	36.2% in Hinche, 3.8-9.8% elsewhere
Relocated	9.7%	

Rapportez-le en entier — Exemple (suite)

Transferred

9.6%

7 cases record a closing month before their opening month and are excluded.

- Tout dans cette unité porte sur des dossiers déjà dans le système.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/protection-and-gbv-data/
prot-06-closure-and-censoring](https://data-analysis.cassion.dev/fr/cours/protection-and-gbv-data/prot-06-closure-and-censoring)