

Moins 0,96 point, et l'hypothèse qu'on ne peut pas tester

Soustraire à chaque groupe son propre point de départ donne des différences de différences de $-0,96$ point. Cela repose sur le fait que les deux groupes étaient sur le point de changer d'autant, hypothèse intenable avec deux passages — la leçon porte donc sur la façon de l'argumenter, non de la vérifier.

Pierre Robentz Cassion

Leçon 3 sur 8

Méthodes d'évaluation d'impact — Quand personne n'a randomisé

Ce que couvre cette leçon

- Deux différences, soustraites
- Ajustez-le comme une régression, parce qu'il vous faut l'intervalle
- L'hypothèse, énoncée exactement
- Pourquoi deux passages ne peuvent pas la tester
- Quatre arguments à faire valoir quand on ne peut pas la tester
- Là où les différences de différences échouent
- Rapportez-le en entier
- La suite

Deux différences, soustraites — En Python

```
import pandas as pd
import numpy as np

cells = d.groupby("feeding_programme")[["baseline", "endline"]].mean()
print(cells.round(4))

did = ((cells.loc[True, "endline"] - cells.loc[True, "baseline"])
       - (cells.loc[False, "endline"] - cells.loc[False, "baseline"]))
print(f"difference-in-differences: {did:+.4f}")
```

Deux différences, soustraites — En R

```
library(dplyr)

d |> summarise(baseline = mean(baseline), endpoint = mean(endpoint),
               .by = feeding_programme) |>
  mutate(gain = endpoint - baseline)
```

Deux différences, soustraites

	Initial	Final	Évolution
Avec cantine	57,75 %	64,41 %	+6,66
Sans cantine	54,67 %	62,29 %	+7,62
Différence	+3,09	+2,12	− 0,96

Deux différences, soustraites

- L'estimation se lit dans la dernière colonne ou dans la dernière ligne et c'est le même nombre — Cette symétrie est ce. . .
- L'écart initial de 3,09 points est ce que le protocole retire — Une comparaison transversale en fin d'année aurait. . .

Ajustez-le comme une régression, parce qu'il vous faut l'intervalle — En Python

```
import statsmodels.formula.api as smf

d = d.assign(gain=d["endline"] - d["baseline"])

naive = smf.ols("gain ~ feeding_programme", data=d).fit()
clustered = naive.get_robustcov_results(cov_type="cluster",
                                       groups=d["school_id"])

by_school = d.groupby(["school_id", "feeding_programme"])["gain"].mean().reset_index()
aggregated = smf.ols("gain ~ feeding_programme", data=by_school).fit()
```

Ajustez-le comme une régression, parce qu'il vous faut l'intervalle — En R

```
lm(gain ~ feeding_programme, data = d)           # naive  
lm(gain ~ feeding_programme, data = by_school)    # school level
```

Ajustez-le comme une régression, parce qu'il vous faut l'intervalle

Approche	Estimation	ET	IC à 95 %
Niveau élève, naïf	−0,96 pt	0,98	−2,89 à +0,96
Niveau élève, groupé par école	−0,96 pt	1,29	−3,48 à +1,56
Niveau école, 15 contre 9	−0,68 pt	1,18	−2,99 à +1,63

Ajustez-le comme une régression, parce qu'il vous faut l'intervalle

- L'intervalle groupé est celui à rapporter — pour la raison qu'a établie le cours de régression : le programme a été. . .
- Chaque version traverse zéro largement — Il n'y a aucun effet à rapporter, et l'intervalle dit que l'étude n'a pas pu. . .

L'hypothèse, énoncée exactement

- Tendances parallèles : en l'absence du programme, les résultats des deux groupes auraient évolué d'autant
- Elle n'exige pas que les groupes se ressemblent — Ils partent avec 3,09 points d'écart et c'est sans importance — le. . .
- Elle n'exige pas que les tendances soient plates — Les deux groupes peuvent progresser ; l'hypothèse est qu'ils. . .
- Elle n'exige ni même variance, ni même effectif, ni même composition — Cela agit sur l'intervalle, non sur. . .
- Ce qu'elle exige est inobservable — parce que c'est une affirmation sur un monde où le programme n'a pas eu lieu

Pourquoi deux passages ne peuvent pas la tester — En Python

```
rounds = assessment["assessment_round"].unique()  
print(rounds)
```

Pourquoi deux passages ne peuvent pas la tester — En R

```
unique(assessment$assessment_round)
```

Pourquoi deux passages ne peuvent pas la tester

- Il y en a deux : initial et final — Avec deux points, vous ne pouvez tracer qu'une droite par groupe, si bien que toute. . .
- Avec trois passages ou plus avant le programme, vous pouvez regarder — Tracez le résultat de chaque groupe sur la. . .
- L'implication pour le protocole est une décision de collecte prise des années plus tôt — Si vous comptez évaluer par. . .

Quatre arguments à faire valoir quand on ne peut pas la tester

- Nommez pourquoi les écoles du programme ont été choisies — Si la sélection a porté sur quelque chose d'invariant dans...
- Montrez que d'autres résultats ont bougé en parallèle — La numératie est mesurée sur les mêmes enfants et n'est pas ce...

Quatre arguments à faire valoir quand on ne peut pas la tester — En Python

```
def panel(domain):
    wide = (assessment[assessment["domain"] == domain]
            .pivot_table(index="student_id", columns="assessment_round",
                          values="pct").dropna())
    out = wide.join(roster.set_index("student_id", how="inner").reset_index())
    return out.assign(gain=out["endline"] - out["baseline"])

for domain in ("literacy", "numeracy"):
    frame = panel(domain)
    fit = smf.ols("gain ~ feeding_programme", data=frame).fit(
        cov_type="cluster", cov_kwds={"groups": frame["school_id"]})
    print(f"{domain}: {fit.params['feeding_programme[T.True]']:.4f}")
```

Quatre arguments à faire valoir quand on ne peut pas la tester — En R

A placebo outcome: it should show nothing, and here it does.

Quatre arguments à faire valoir quand on ne peut pas la tester

- La numératie donne +0,39 point, IC à 95 % $-1,16$ à $+1,94$ — le même protocole appliqué à un résultat que le programme. . .
- Testez une période placebo s'il en existe une — Deux passages antérieurs au programme devraient donner des différences. . .
- Montrez que le résultat n'est pas porté par une seule unité — Retirez chaque école à tour de rôle et réajustez ; si. . .

Quatre arguments à faire valoir quand on ne peut pas la tester — En Python

```
for school in sorted(d["school_id"].unique()):  
    subset = d[d["school_id"] != school]  
    est = smf.ols("gain ~ feeding_programme", data=subset).fit()  
    print(f"without {school}: {est.params['feeding_programme[T.True]':+.4f}")
```

Quatre arguments à faire valoir quand on ne peut pas la tester — En R

24 refits, one line each. Cheap, and a reviewer will ask.

Quatre arguments à faire valoir quand on ne peut pas la tester

- Retirer n'importe quelle école déplace l'estimation entre $-1,61$ et $-0,37$ point —
Aucune école ne la porte à elle seule,...

Là où les différences de différences échouent

- Un calendrier différent — Si le programme a démarré à des mois différents selon les écoles, une seule césure. . .
- Un changement de composition — Les 585 enfants sont ceux qui ont les deux passages
- L'anticipation — Si les écoles ont changé de comportement en apprenant que le programme arrivait, le point de départ. . .
- Un groupe témoin qui est affecté — Des enfants qui changent d'école, des enseignants qui sont mutés, ou une école. . .

Rapportez-le en entier — Exemple (suite)

School feeding and literacy, difference-in-differences

585 students with both assessment rounds, in 24 schools.

	Baseline	Endline	Change
Feeding (15 sch)	57.75%	64.41%	+6.66
No programme (9)	54.67%	62.29%	+7.62
Difference	+3.09	+2.12	-0.96

Estimate -0.96 points, 95% CI -3.48 to +1.56, clustered on 24 schools.

No effect on literacy is detected.

The estimate assumes the two groups would have changed by the same amount without the programme. Only two assessment rounds exist, so this cannot be examined and is argued rather than tested: assignment appears to follow district and school size rather than a trend in results, and the same

Rapportez-le en entier — Exemple (suite)

design applied to numeracy also shows nothing.

The interval excludes effects larger than 1.6 points in either direction but not smaller ones. The design's minimum detectable effect was 4.9 points, so this is a null result about large effects only.

- Le dernier paragraphe convertit un résultat nul en une affirmation bornée — ce que le cours de statistiques demandait. . .

- Les différences de différences emploient le point de départ en le soustrayant.

Lire la leçon complète, avec le code exécutable

```
https://data-analysis.cassion.dev/fr/cours/impact-evaluation-methods/  
ie-03-difference-in-differences
```