

Une page, écrite avant l'arrivée des données

Chaque protocole de ce cours a été reconstitué après coup, et chaque reconstitution contenait un choix que le résultat aurait pu influencer. Cette leçon écrit les choix d'abord — le protocole, le résultat, le seuil et l'analyse — sur une page qu'un programme produira réellement.

Pierre Robentz Cassion

Leçon 7 sur 8

Méthodes d'évaluation d'impact — Ce que l'évaluation affirme

Ce que couvre cette leçon

- Six choix que le résultat aurait pu faire à votre place
- La page
- Pourquoi chaque section est là
- S'écarter du plan
- Préspecifier ne veut pas dire ne pas explorer
- Quand il n'y a pas de plan et que les données sont arrivées
- Rapportez-le en entier
- La suite

Six choix que le résultat aurait pu faire à votre place

Leçon	Le choix	Ce qu'il aurait pu être choisi pour produire
1	Avant-après, coupe transversale ou différences de différences	+6,97, +2,12 ou -0,96 point
2	Quelles covariables comptent pour l'équilibre	Un tableau d'apparence acceptable
3	Erreurs-types groupées ou naïves	Un intervalle excluant zéro
4	Apparier ou non	-0,71 ou -0,96, et quelles écoles sont dedans
5	Quelle fenêtre	Celle qui est significative
6	Quelle CCI	Un EMD sous l'effet observé

Six choix que le résultat aurait pu faire à votre place

- Six choix binaires produisent soixante-quatre analyses — Aucun n'est malhonnête et chacun se défend isolément, ce qui...
- C'est le problème des comparaisons multiples du cours de statistiques sous un autre habit — Là c'étaient vingt-quatre...

Evaluation plan: school feeding and learning outcomes

Written 2024-01-15, before any endline data exists.

1. QUESTION

Does the school feeding programme raise literacy scores?

2. DESIGN

Difference-in-differences. 15 programme schools, 9 comparison schools, baseline and endline literacy assessment on the same children.

Assignment was not randomised; the balance table will be reported.

3. OUTCOME

Percent correct on the literacy assessment. Baseline is out of 40 items and endline out of 50, so raw scores are not comparable and percent correct is the primary outcome. Numeracy is secondary.

4. THRESHOLD OF INTEREST

3 percentage points. Below that the programme would not be expanded on these grounds, so an effect smaller than 3 points is reported as "no programme-relevant effect" regardless of its p-value.

5. ANALYSIS

OLS of the individual gain on a programme indicator, standard errors clustered on school. Adjusted for baseline score and district. Reported with a 95% confidence interval, not a p-value alone.

6. POWER

ICC assumed 0.065 from the attendance data. Minimum detectable effect 4.9 points, which is above the 3-point threshold. The evaluation is therefore underpowered for the effect size that matters, and this is stated in the report whatever the result.

7. WHAT WOULD CHANGE THE CONCLUSION

Attrition above 20%, or differential attrition between arms above 5 points, would make the panel unrepresentative and the primary analysis would be reported alongside a bounds analysis.

8. WHAT WE WILL NOT CLAIM

Causality beyond difference-in-differences. Effects on attendance, enrolment or nutrition, which are not measured here.

- Le point 6 est ce qui rend ce document digne d'être écrit — Il est connaissable en janvier, il dit que l'étude ne peut. . .

Pourquoi chaque section est là

- La question, en une phrase — S'il en faut plus d'une, c'est plus d'une évaluation
- Le protocole, nommé — « Nous comparerons les écoles avec et sans programme » n'est pas un protocole ; ce sont trois. . .
- Le résultat, défini à l'item près — L'exercice du cours de statistiques a trouvé une demande comparant une épreuve de. . .
- Le seuil d'intérêt, en unités de programme — C'est la section que personne n'écrit et c'est celle qui empêche un rien. . .
- L'analyse, spécifiée — Y compris les erreurs-types, parce que le choix entre naïves et groupées a déplacé l'intervalle. . .
- La puissance, calculée — Avant, non après

Pourquoi chaque section est là (suite)

- Ce qui changerait la conclusion — Nommer les modes d'échec d'avance signifie que le rapport n'aura pas à débattre de...
- Ce que vous n'affirmerez pas — La section la plus courte et celle qui protège la crédibilité de l'évaluation quand...

- Les écarts sont permis et doivent être listés — Un court tableau dans le rapport — ce qui était prévu, ce qui a été. . .

S'écarter du plan — Exemple

Deviations from the evaluation plan

Planned: adjust for baseline score and district.

Done: as planned.

Planned: primary analysis on all enrolled students.

Done: restricted to the 585 students with both assessment rounds.

Why: 156 students have no endline. Attrition analysis added; the students who left scored 39.3% at baseline against 56.8% for those who stayed, so the panel is not representative and this is reported as a limitation.

- Un écart listé est une force — Un écart non listé, découvert par un relecteur, est la fin de la crédibilité de...

Préspécifier ne veut pas dire ne pas explorer

- Deux sections, clairement étiquetées — L'analyse préspécifiée répond à la question posée

Préspécifier ne veut pas dire ne pas explorer — En Python

```
PRIMARY = "gain ~ feeding_programme + baseline + district"      # pre-specified
EXPLORATORY = [                                                  # labelled as such
    "gain ~ feeding_programme * sex",
    "gain ~ feeding_programme * baseline_quartile",
    "gain ~ feeding_programme + C(school_id)",
]
print(f"pre-specified: 1 model.  exploratory: {len(EXPLORATORY)} models.")
```

Préspécifier ne veut pas dire ne pas explorer — En R

Two objects, two headings in the report. That is the whole discipline.

Préspécifier ne veut pas dire ne pas explorer

- Un constat de sous-groupe est exploratoire à moins que le sous-groupe ait été nommé d'avance — Le cours de statistiques. . .

Quand il n'y a pas de plan et que les données sont arrivées

- Écrivez le plan quand même, daté, avant de lancer l'analyse — Vous avez vu les données ; vous n'avez pas encore vu le. . .
- Préspecifiez l'analyse principale et rapportez tout le reste comme exploratoire — La distinction garde du sens même. . .
- Dites combien d'analyses vous avez lancées — Le bloc de rapport du cours de statistiques se termine par « comparaisons. . .

Rapportez-le en entier — Exemple (suite)

Analysis plan and deviations

The evaluation plan was written on 2024-01-15, before endline data collection, and is reproduced in Annex A.

Primary analysis as pre-specified: difference-in-differences on percent correct, clustered on school, adjusted for baseline and district.

Deviations: one, listed in Annex A. The panel was restricted to students with both rounds; an attrition analysis was added.

Exploratory analyses: 3, reported in Annex B and labelled as exploratory. None is presented as a finding.

Threshold of interest: 3 percentage points, agreed with the programme team in January. The observed estimate is -0.96 points (95% CI -3.5 to +1.6)

Rapportez-le en entier — Exemple (suite)

and is below the threshold in magnitude.

- La date de la première ligne est l'élément porteur de tout le document — Tout le reste est une affirmation d'intention. . .

- Le plan dit ce que l'évaluation affirmera.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/impact-evaluation-methods/
ie-07-decide-the-design-first](https://data-analysis.cassion.dev/fr/cours/impact-evaluation-methods/ie-07-decide-the-design-first)