

Une différence standardisée de 0,85, quand la limite est 0,1

Quinze écoles ont la cantine et neuf non. Elles diffèrent de 0,85 écart-type sur la littératie initiale et de 0,68 sur la taille, et le Centre compte cinq écoles avec cantine pour une sans. Voilà à quoi ressemble un groupe de comparaison quand personne n'a randomisé.

Pierre Robentz Cassion

Leçon 2 sur 8

Méthodes d'évaluation d'impact — Le contrefactuel

Ce que couvre cette leçon

- Construisez le tableau d'équilibre avant tout le reste
- Ce qu'est la différence standardisée, et pourquoi pas une valeur p
- Ce que la randomisation aurait acheté
- Quand la randomisation est refusée, et que dire
- Les trois questions à poser avant un protocole d'évaluation
- Rapportez-le en entier
- La suite

Construisez le tableau d'équilibre avant tout le reste — En Python (suite)

```
import pandas as pd
import numpy as np

enrolment = pd.read_csv("school-enrolment-2024.v1.csv")
current = enrolment[(enrolment["school_year"] == 2024)
                    & (enrolment["age_years"] <= 20)] # drop the age outliers
roster = pd.read_csv("school-roster-2024.v1.csv").drop_duplicates(
    "student_id", keep="last")
joined = current.merge(roster[["student_id", "school_id", "feeding_programme"]],
                      on="student_id", suffixes=("", "_r"))
joined["over_age"] = joined["age_years"] > joined["grade"] + 5
joined["displaced"] = joined["displacement_status"] != "resident"

by = joined.groupby("school_id")
schools = pd.DataFrame({
    "feeding_programme": by["feeding_programme"].first(),
```

Construisez le tableau d'équilibre avant tout le reste — En Python (suite)

```
"baseline": by["baseline"].mean(),      # from the assessment file
"age_years": by["age_years"].mean(),
"size": by.size(),
"over_age": by["over_age"].mean(),
"disability": by["disability_reported"].mean(),
"displaced": by["displaced"].mean(),
"district": by["admin2"].first(),
}).reset_index()

def std_diff(treated, control):
    pooled = np.sqrt((treated.var(ddof=1) + control.var(ddof=1)) / 2)
    return (treated.mean() - control.mean()) / pooled

for column in ["baseline", "age_years", "size", "over_age",
               "disability", "displaced"]:
    t = schools[schools["feeding_programme"]][column].dropna()
```

Construisez le tableau d'équilibre avant tout le reste — En Python (suite)

```
c = schools[~schools["feeding_programme"]][column].dropna()
print(f"{column:12} {t.mean():8.4f} {c.mean():8.4f}"
      f"   std.diff {std_diff(t, c):+.3f}")
```

Construisez le tableau d'équilibre avant tout le reste — En R

```
library(dplyr)

schools |>
  summarise(across(c(baseline, age_years, size, over_age), mean),
            .by = feeding_programme)
```

Construisez le tableau d'équilibre avant tout le reste

Caractéristique	Avec cantine (15)	Sans (9)	Différence standardisée
Littératie initiale	54,10 %	51,05 %	+0,85
Âge moyen	9,65	9,42	+0,79
Taille de l'école	50,3	44,8	+0,68
Handicap signalé	7,6 %	9,7 %	−0,52
Déplacé ou retourné	23,7 %	26,9 %	−0,51
Retard d'âge	39,2 %	36,8 %	+0,43

Construisez le tableau d'équilibre avant tout le reste

- Le seuil conventionnel vaut 0,10, et chacune des six le dépasse — quatre d'entre elles d'un facteur cinq ou plus

Construisez le tableau d'équilibre avant tout le reste — En Python

```
print(pd.crosstab(schools["district"], schools["feeding_programme"]))
```

Construisez le tableau d'équilibre avant tout le reste — En R

```
table(schools$district, schools$feeding_programme)
```

Construisez le tableau d'équilibre avant tout le reste

District	Sans cantine	Avec cantine
Artibonite	3	3
Centre	1	5
Nord-Ouest	1	5
Sud	4	2

Construisez le tableau d'équilibre avant tout le reste

- Deux districts sont à cinq contre un en faveur du programme et un est à deux contre quatre — Tout ce qui diffère par...

Ce qu'est la différence standardisée, et pourquoi pas une valeur p — En Python

```
t = schools[schools["feeding_programme"]]["baseline"]
c = schools[~schools["feeding_programme"]]["baseline"]
from scipy import stats
print(f"std.diff {std_diff(t, c):+.3f}    p = {stats.ttest_ind(t, c).pvalue:.3f}")
```

Ce qu'est la différence standardisée, et pourquoi pas une valeur p — En R

Both, and only one of them is the right tool.

Ce qu'est la différence standardisée, et pourquoi pas une valeur p

- Différence standardisée $+0,85$, $p = 0,064$ — Un relecteur ne lisant que la valeur p conclurait que les écoles sont. . .
- La valeur p d'un test d'équilibre confond le déséquilibre et la taille d'échantillon — ce qui est l'inverse de ce dont. . .
- Rapportez la différence standardisée, qui ne dépend pas de n — Au-delà de $0,10$ c'est un déséquilibre qu'il faut traiter. . .

Ce que la randomisation aurait acheté

- Cette dernière proposition est toute la valeur — L'ajustement peut corriger un déséquilibre de littérature initiale. . .

Ce que la randomisation aurait acheté — En Python

```
rng = np.random.default_rng(20260729)
draws = []
for _ in range(2000):
    assigned = rng.permutation(schools["feeding_programme"].values)
    t = schools["baseline"][assigned]
    c = schools["baseline"][~assigned]
    draws.append(std_diff(t, c))
print(f"under random assignment, |std.diff| > 0.85 in "
      f"{np.mean(np.abs(draws) > 0.85):.1%} of draws")
```

Ce que la randomisation aurait acheté — En R

Permute the assignment 2,000 times and see how unusual the real split is.

Ce que la randomisation aurait acheté

- Réaffecter au hasard les mêmes 24 écoles produit un déséquilibre de cette ampleur dans 5,3 % de 2 000 tirages — Le...

Quand la randomisation est refusée, et que dire

Objection	La réponse honnête
« On ne peut pas priver des écoles dans le besoin »	Randomisez l' ordre d'un déploiement par vagues ; tout le monde est servi, la séquence est aléatoire
« C'est le ministère qui choisit les écoles »	Randomisez à l'intérieur de la liste éligible du ministère, ou évaluez la règle d'éligibilité comme un protocole à seuil
« C'est déjà en cours »	Dites-le, employez un protocole de comparaison, et publiez le tableau d'équilibre
« Le bailleur veut des résultats cette année »	Un essai randomisé sous-dimensionné et des différences de différences bien argumentées sont tous deux défendables ; un avant-après non

Quand la randomisation est refusée, et que dire

- Le déploiement par vagues est le protocole le plus sous-employé de ce secteur — Tout programme qui s'étend sur trois. . .
- Là où rien de tout cela n'est possible, le tableau d'équilibre est ce que vous devez au lecteur — Ce n'est pas une. . .

Les trois questions à poser avant un protocole d'évaluation

- Qui a décidé qui reçoit le programme, et sur quoi ? — Si la réponse est une règle, vous avez peut-être un protocole à . . .
- Y a-t-il quelque chose de mesuré avant le début du programme ? — Sans point de départ, les différences de différences. . .
- Combien d'unités ont été affectées ? — Non pas combien de personnes ont été mesurées — combien d'écoles, de centres, de . . .

Rapportez-le en entier — Exemple (suite)

Comparability of feeding and non-feeding schools

15 schools with the programme, 9 without. Assignment was not randomised.

Standardised differences at baseline:

Baseline literacy	+0.85	School size	+0.68
Mean age	+0.79	Disability	-0.52
Displacement	-0.51	Over-age	+0.43

All six characteristics exceed the 0.10 conventional threshold and all six exceed 0.25. Programme schools are concentrated in Centre and Nord-Ouest (5 of 6 schools in each) and scarce in Sud (2 of 6).

Balance tests are not reported as p-values: with 24 units they would show balance regardless. The standardised difference does not depend on n.

Rapportez-le en entier — Exemple (suite)

The comparison is adjusted for baseline literacy and district. Adjustment cannot address unmeasured differences in why these schools were selected, and no result in this report should be read as an effect of the programme alone.

Rapportez-le en entier

- Le dernier paragraphe est là où une évaluation gagne ou perd un relecteur — et sa longueur est la bonne : deux phrases,...

- Si les deux groupes différaient au départ, la façon d'employer ce point de départ est de le soustraire.

Lire la leçon complète, avec le code exécutable

`https://data-analysis.cassion.dev/fr/cours/impact-evaluation-methods/
ie-02-what-randomisation-buys`