Better balance, six schools lighter

Matching cuts the baseline imbalance from 3.35 points to 1.41 and gives an estimate of -0.71 instead of -0.96. It does it by discarding six of the fifteen programme schools — the six with the highest baseline scores — so the answer now applies to a different set of schools than the question did.

Pierre Robentz Cassion

What this lesson covers

- Matching, in one paragraph
- Match the schools
- Which six schools left, and why it matters
- What "looks like it" should mean
- Check overlap before you match, not after
- Four things to report, always
- What matching cannot do
- Report it whole
- What comes next

Matching, in one paragraph

- Find, for each treated unit, an untreated unit that looks like it; compare only the pairs — Everything else in the...

Match the schools — In Python (cont.)

```
import pandas as pd
import numpy as np

treated = schools[schools["feeding_programme"]].sort_values("baseline")
control = schools[~schools["feeding_programme"]].sort_values("baseline")

used, pairs = set(), []
for _, c in control.iterrows():
    available = treated[~treated["school_id"].isin(used)]
    nearest = (available["baseline"] - c["baseline"]).abs().idxmin()
    used.add(treated.loc[nearest, "school_id"])
    pairs.append({"control": c["school_id"],
                  "treated": treated.loc[nearest, "school_id"],
                  "base_c": c["baseline"], "base_t": treated.loc[nearest, "baseline"],
                  "gain_c": c["gain"], "gain_t": treated.loc[nearest, "gain"]})
```

Match the schools — In Python (cont.)

```
matched = pd.DataFrame(pairs)
print(f"{len(matched)} pairs from {len(treated)} treated and {len(control)} control")
```

Match the schools — In R

```
library(MatchIt)

m <- matchit(feeding_programme ~ baseline, data = schools,
             method = "nearest", ratio = 1)
summary(m)
```

Match the schools

- Nine pairs from fifteen treated and nine control schools — One-to-one matching without replacement can produce at most. . .

Match the schools

	Before matching	After matching
Treated schools	15	**9**
Control schools	9	9
Baseline gap	+3.35 pts	**+1.41 pts**
Estimate	−0.96 pts	**−0.71 pts**
Standard error	1.29	1.22

Match the schools

- The balance improved and the estimate barely moved — which is the outcome to hope for: it says the...

Which six schools left, and why it matters — In Python

```
dropped = treated[~treated["school_id"].isin(used)]
print(dropped[["school_id", "baseline", "gain"]].round(4))
print(f"dropped mean baseline: {dropped['baseline'].mean():.4f}")
print(f"kept mean baseline:    {matched['base_t'].mean():.4f}")
```

Which six schools left, and why it matters — In R

```
# Which units matching threw away is the first table to print, not the last.
```

Which six schools left, and why it matters

- The six discarded schools are the six with the highest baseline literacy — 53.9% to 65.5%, against a matched-treated...
- That is not a flaw in the matching; it is the matching telling you something true — The nine control schools cannot...
- But it changes the claim — The matched estimate answers "among schools with baseline literacy below about 60%, what did...

What "looks like it" should mean

Approach	Match on	Use when
Exact	Identical values of a few categorical variables	Few covariates, large samples
Nearest neighbour on a propensity score	The predicted probability of treatment	Many covariates
Coarsened exact	Values binned into ranges	You want a guaranteed balance level

What "looks like it" should mean — In Python

```
import statsmodels.formula.api as smf

ps = smf.logit("feeding_programme ~ baseline + size + district",
               data=schools).fit(disp=0)
schools["pscore"] = ps.predict(schools)
print(schools.groupby("feeding_programme")["pscore"].describe()[
    ["min", "mean", "max"]].round(3))
```

What "looks like it" should mean — In R

```
glm(feeding_programme ~ baseline + size + district, data = schools,
    family = binomial())
```

What "looks like it" should mean

- A propensity score is one number summarising many covariates — and matching on it is equivalent to matching on all of...
- It is not a way to get more data — The score reduces dimensionality; it does not create comparators where none exist

Check overlap before you match, not after — In Python

```
t = schools[schools["feeding_programme"]]["pscore"]
c = schools[~schools["feeding_programme"]]["pscore"]
print(f"treated range {t.min():.3f}-{t.max():.3f}")
print(f"control range {c.min():.3f}-{c.max():.3f}")
print(f"treated units above the control maximum: {(t > c.max()).sum()}")
```

Check overlap before you match, not after — In R

```
# Plot the two propensity score distributions. The tails are the whole story.
```

Check overlap before you match, not after

- Units outside the other group's range have no counterfactual in the data — Matching drops them, which is honest;...

Four things to report, always

- How many units were discarded, and which — The first table, not a footnote
- Balance before and after — Standardised differences on every covariate, both columns, so a reader can see what matching…
- What population the estimate now describes — One sentence, in the words a programme manager uses — "schools with…
- The unmatched estimate too — If matching changed the answer materially, that is itself a finding about how much the…

What matching cannot do

- It cannot fix an unmeasured confounder — Two schools with identical baseline scores can still differ in why one was...
- It cannot create a control group — If the untreated units are systematically different, matching will report excellent...
- It does not remove the need for a design — Matching plus a baseline is difference-in-differences on a subset; matching...

Report it whole — Example (cont.)

School feeding and literacy, matched difference-in-differences

Nearest-neighbour matching on baseline literacy, 1:1 without replacement.
9 matched pairs from 15 treated and 9 control schools.

6 of 15 programme schools were discarded: SCH05, SCH12, SCH14, SCH15, SCH19, SCH21. Their mean baseline literacy is 60.7% against 55.8% for the matched treated schools; no control school scored high enough to match them.

Baseline imbalance	+3.35 points before matching, +1.41 after.
Estimate	-0.71 points, SE 1.22, on 9 pairs.
Unmatched estimate	-0.96 points, 95% CI -3.48 to +1.56.

The matched estimate describes schools with baseline literacy below about 60%. It does not describe the six highest-performing programme schools,

Report it whole — Example (cont.)

```
which have no comparator in this data.

Matching balances the covariates it was given. It does not address why
these schools were selected for the programme.
```

Report it whole

- The named list of dropped schools is what makes this reportable — A reader can check whether the six that left are the...

What comes next

- Matching and difference-in-differences both build a comparison group out of units that were not assigned by a rule.

Read the full lesson, with runnable code

https://data-analysis.cassion.dev/courses/impact-evaluation-methods/ie-04-what-matching-discards