

RANDOMIZED EXPERIMENTS

Author

Vinitra Muralikrishnan

October 7, 2024

Contents

1 Randomised Experiments

If causal inference is about reducing or eliminating bias to establish causal links, how do we do that?. The answer is Randomized Experiments. Considering our tablet treatment assignment, we know that there is certain bias introduced due to the schools being very different from one another. In order to truly understand the effect of this treatment both the schools need to be comparable. We cannot eliminate the effects of the superior facilities from the schools but we can however, balance their effect. This is done using randomized experiments, which involves registering treatment on random instead of registering it to everyone in both groups. This would make, average treatment effect comparable. Intuitively, we negate the effect of factors that overshadow some others by only assigning the treatments, randomly. The probability distribution can be anything with only 20% getting the treatment. In doing so we also make Y_0, Y_1 the potential outcomes independent of the treatment but still able to maintain the dependency between true outcome Y and T .

Given below is an example of randomized experiment:

In this example, the effect of online classroom teaching is studied. To do this, online courses are assigned at random to students. This way, we negate effect of factors such as better propensity for studying through on-

	gender	asian	black	hawaiian	hispanic	unknown	white	format_ol	format_blended	falseexam
0	0	0.0	0.0	0.0	0.0	0.0	1.0	0	0.0	63.29997
1	1	0.0	0.0	0.0	0.0	0.0	1.0	0	0.0	79.96000
2	1	0.0	0.0	0.0	0.0	0.0	1.0	0	1.0	83.37000
3	1	0.0	0.0	0.0	0.0	0.0	1.0	0	1.0	90.01994
4	1	0.0	0.0	0.0	0.0	0.0	1.0	1	0.0	83.30000

We can see that we have 323 samples. It's not exactly big data, but something we can work with. To estimate the causal effect, we can simply compute the mean score for each of the treatment groups.

```
(data
 .assign(class_format = np.select(
 [data["format_ol"].astype(bool), data["format_blended"].astype(bool)],
 ["online", "blended"],
 default="face_to_face"
 ))
 .groupby(["class_format"])
 .mean())
```

	gender	asian	black	hawaiian	hispanic	unknown	white	format_ol	format_blended	falseexam
class_format										
blended	0.550459	0.217949	0.102564	0.025641	0.012821	0.012821	0.628205	0.0	1.0	77.093731
face_to_face	0.633333	0.202020	0.070707	0.000000	0.010101	0.000000	0.717172	0.0	0.0	78.547485
online	0.542553	0.228571	0.028571	0.014286	0.028571	0.000000	0.700000	1.0	0.0	73.635263

line courses, more time for study, etc.

In the above example, we can see that by assigning online randomly, we get comparable groups where the outcome before treatment is similar. The falseexam variable gives an estimate of the outcomes of the 3 groups and the ATE for online students ; ATE for offline students.

In real world there can be challenges in performing such randomized experiments or randomised controlled trials. Sometimes it's just unethical to do RCT. For ex - in order to study the effects of smoking during pregnancy, we can't just randomly pick expectant mothers and ask them to smoke. They are also practically mandatory for drug trials but still can be challenging to conduct them. They can be also expensive to conduct, hence we do conditional randomization, to reduce the cost.

The assignment that's done during these randomized experiments while can be random, still needs to be informed by informed mechanism about how the system or world works. In order to understand the direct causal factors, we need to apply the knowledge of system and do the assignment such that it actually makes the groups comparable. For this, understanding of the assignment mechanism is a must and pivotal to answer the causal inference questions.

You can find more information in Chapter 1 of [Causality Handbook](#).