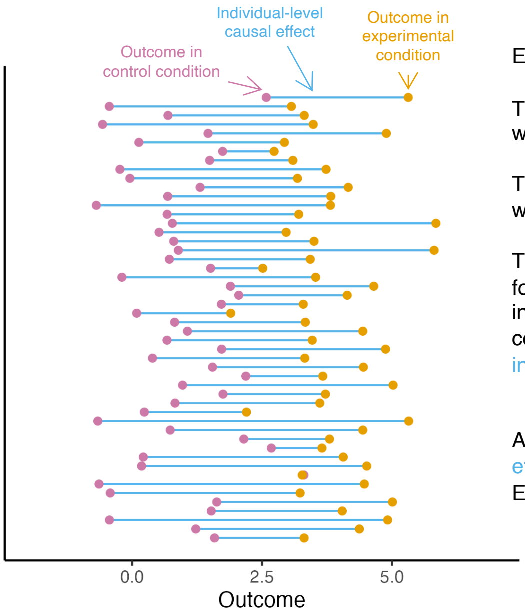
Why we can estimate the average causal effect with the help of randomization



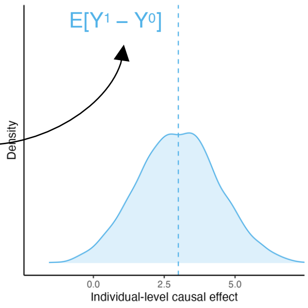
Every line represents an individual.

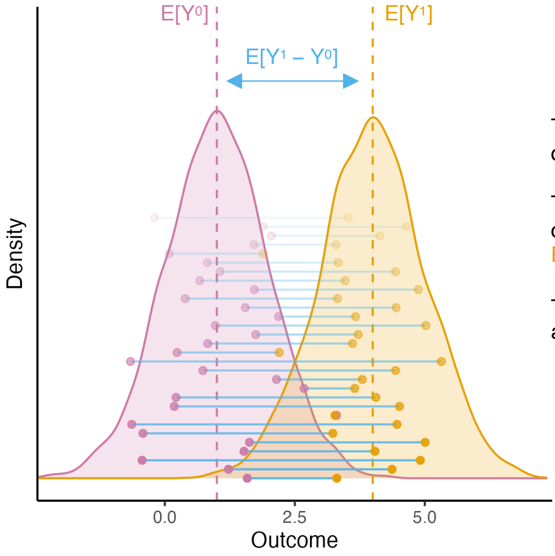
The point on the left marks the individual‘s outcome we would observe if the individual was assigned to the control condition, their potential outcome Y^0.

The point on the right marks the individual‘s outcome we would observe if the individual was assigned to the experimental condition, their potential outcome Y^1.

The difference between these two potential outcomes is the individual-level causal effect for that person, $Y^1 – Y^0$. The fundamental problem of causal inference is that at any point in time, we can only observe a person in either the control condition or in the treatment condition. Thus, we never know both Y^1 and Y^0 which means we never know the individual-level causal effect.

Across people, these (unknowable) individual-level causal effects follow a distribution with the expectation $E[Y^1 – Y^0]$, the average causal effect.

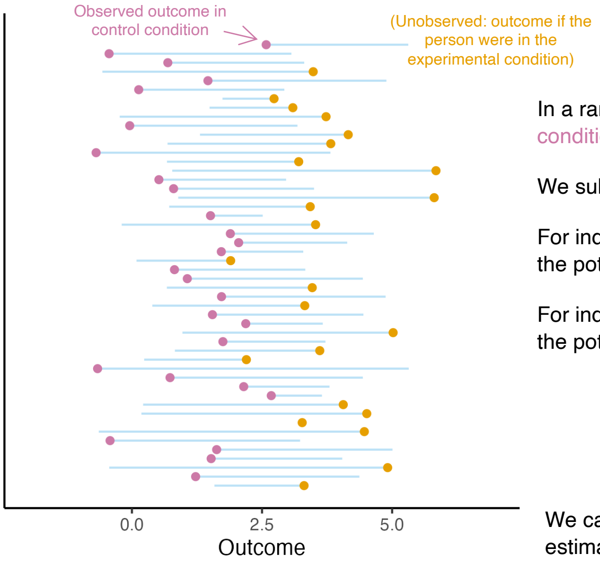




The potential outcomes Y^0 and Y^1 follow their own distributions which have the expectations $E[Y^0]$ and $E[Y^1]$.

Thanks to the linearity of the expectation, the difference between these expectations corresponds to the average causal effect:
$E[Y^1] – E[Y^0] = E[Y^1 – Y^0]$.

This means that if we can estimate both $E[Y^1]$ and $E[Y^0]$, we can also estimate the average causal effect $E[Y^1 – Y^0]$

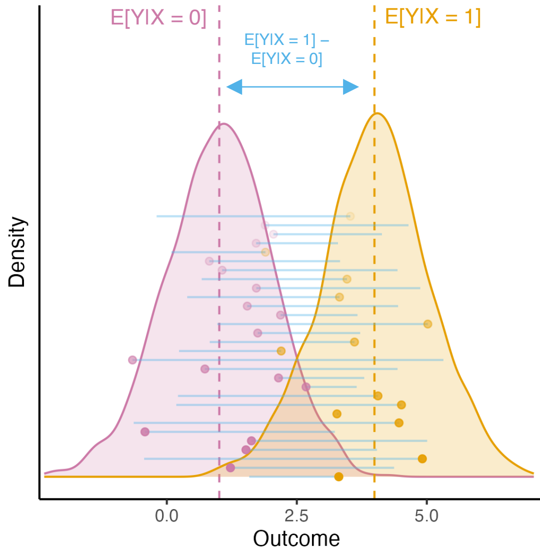


In a randomized experiment, we randomly assigned people either to the control condition $(X = 0)$ or to the experimental condition $(X = 1)$.

We subsequently measure their outcomes.

For individuals in the control condition, their outcome Y will correspond to Y^0, the potential outcome in the control condition. Y^1 remains unobserved.

For individuals in the experimental condition, their outcome Y will correspond to Y^1, the potential outcome in the experimental condition. Y^0 remains unobserved.



We can now simply calculate the average outcome in the control condition to estimate $E[Y|X = 0]$, which as we have seen above equals $E[Y^0|X = 0]$.
We can also calculate the average outcome in the experimental condition to estimate $E[Y|X = 1]$, which equals $E[Y^1|X = 1]$.

Recall that we have randomly assigned people to either condition, so we essentially randomly sample from the distributions of the potential outcomes Y^0 and Y^1. If we randomly sample from a distribution, we can estimate its expectation without any bias – this means that $E[Y^0|X = 0]$ is an unbiased estimate of $E[Y^0]$ and $E[Y^1|X = 1]$ is an unbiased estimate of $E[Y^1]$.

Another way to think about this: Since we have randomly assigned the conditions, there is no reason why the people who actually end up in the experimental condition $(X = 1)$ should have systematically lower or higher potential outcomes Y^1 than the (unobserved) potential outcomes Y^1 of the people who instead ended up in the control condition $(X = 0)$; $E[Y^1|X = 1] = E[Y^1|X = 0] = E[Y^1]$.

The same holds for Y^0.

So, we can estimate the expectations of both potential outcomes Y^0 and Y^1 *across everybody* by simply averaging the observed outcomes in the two experimental groups $X = 0$ and $X = 1$.

Taking the difference of these averages provides an estimate $E[Y^1] – E[Y^0]$ which equals $E[Y^1 – Y^0]$ which is in turn the average causal effect, the expectation of the (unknowable) individual level causal effects.

Figure by Julia M. Rohrer