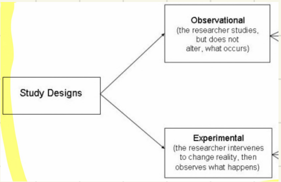
RECRUITMENT TO A PROGRAM

Example 4.4.1 *A government is funding a job training program aimed at getting jobless people back into the workforce. A pilot randomized experiment shows that the program is effective; a higher percentage of people were hired among those who finished the program than among those who did not go through the program. As a result, the program is approved, and a recruitment effort is launched to encourage enrollment among the unemployed, by offering the job training program to any unemployed person who elects to enroll.*

Lo and behold, enrollment is successful, and the hiring rate among the program's graduates turns out even higher than in the randomized pilot study. The program developers are happy with the results and decide to request additional funding.

RANDOMIZED EXPERIMENT: For example, if an experiment compares a new drug against a standard drug, then the patients should be allocated to either the new drug or to the standard drug control using randomization.



CRITICS: There is no proof the program works among those choose to enroll of their own volition.

We must estimate the diff. benefit on those enrolled.

$$\boxed{ETT = E[\underbrace{Y_1 - Y_0}_{\text{causal effect of training } (X) \text{ on hiring } (Y)} \mid X = 1]}$$

those actually choosing the traing program

Serious economic implications.

ETT not estimable from obs nor xp data: clash of worlds,

Y_0 vs. $X = 1$

Expectation a trained person would find a job had he/she not been trained:

$E[Y_0 / X = 1]$

Which can be reduced to estimable expressions when a set Z of covariates satisfies the backdoor criterion regarding X and Y.

Definition 3.3.1 (The Backdoor Criterion) *Given an ordered pair of variables* (X, Y) *in a directed acyclic graph G, a set of variables Z satisfies the* backdoor criterion *relative to* (X, Y) *if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X.*

If a set of variables Z satisfies the backdoor criterion for X and Y, then the causal effect of X on Y is given by the formula

BC ✓ ⇒
$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z)$$

just as when we adjust for $PA(X)$. (Note that $PA(X)$ always satisfies the backdoor criterion.)

A different weighted average: $P(Z = z / X = x')$ for $P(Z = z)$

$$\begin{aligned}
ETT &= E[Y_1 - Y_0|X = 1] \\
&= E[Y_1|X = 1] - E[Y_0|X = 1] \\
&= E[Y|X = 1) - \sum_z E[Y|X = 0, Z = z]P(Z = z|X = 1)
\end{aligned}$$

Identification of ETT for binary X when experimental data is available: $P(Y = y / do(X = x))$

" non-experimental " " " : $P(X = x, Y = y)$

Identification of ETT also when intermediate variables between X and Y satisfy frontdoor criterion:

Definition 2.4.1 (***d*-separation**) *A path p is blocked by a set of nodes Z if and only if*

1. *p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., B is conditioned on), or*
2. *p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.*

Definition 3.4.1 (**Front-Door**) *A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if*

1. *Z intercepts all directed paths from X to Y.*
2. *There is no unblocked path from X to Z.*
3. *All backdoor paths from Z to Y are blocked by X.*

Theorem 3.4.1 (**Front-Door Adjustment**) *If Z satisfies the front-door criterion relative to (X, Y) and if $P(x, z) > 0$, then the causal effect of X on Y is identifiable and is given by the formula*

$$P(y|do(x)) = \sum_z P(z|x) \sum_{x'} P(y|x', z)P(x') \tag{3.16}$$

What is common in these situations is that the causal graph can tell us whether ETT is estimable and, if so, how.

ADDITIVE INTERVENTIONS

Example 4.4.2 *In many experiments, the external manipulation consists of adding (or subtracting) some amount from a variable X without disabling preexisting causes of X, as required by the do(x) operator. For example, we might give 5 mg/l of insulin to a group of patients with varying levels of insulin already in their systems. Here, the preexisting causes of the manipulated variable continue to exert their influences, and a new quantity is added, allowing for differences among units to continue. Can the effect of such interventions be predicted from observational studies, or from experimental studies in which X was set uniformly to some predetermined value x?*

Predict intervention from obs. or exp. studies

Whenever a set Z satisfies BC, ET is estimable using

$$\sum_z P(Y = y|X = x, Z = z)P(Z = z|X = x')$$

.

$$
\begin{aligned}
&E[Y|add(q)] - E[Y] \\
&\quad = \sum_{x'} E[Y_{x'+q}|X = x']P(x') - E[Y] \\
&\quad = \sum_{x'} \sum_z E[Y|X = x' + q, Z = z]P(Z = z|X = x')P(X = x') - E[Y]
\end{aligned}
$$

Why not a randomized trial, add(q) to one group and add(0) to another?

Because we cannot reduce $E[Y/add(q)]$ to do-expressions: we can't infer the desired quantity from a standard experiment where X is set uniformly to all the population.

Scientists would like: $E[Y/do(X=x+q)] - E[Y/do(X=x)]$ but each one has its own x'. So we want the avg effect of adding q to any x'.

The reader may also wonder why $E[Y|add(q)]$ is not equal to the average causal effect

$$\sum_x E[Y|do(X = x + q)] - E[Y|do(X = x)]P(X = x)$$

randomly chosen subjects ≠ policy

After all, if we know that adding q to an individual at level $X = x$ would increase its expected Y by $E[Y|do(X = x + q)] - E[Y|do(X = x)]$, then averaging this increase over X should give us the answer to the policy question $E[Y|add(q)]$. Unfortunately, this average does *not* capture the policy question. This average represents an experiment in which subjects are chosen at random from the population, a fraction $P(X = x)$ are given an additional dose q, and the rest are left alone. But things are different in the policy question at hand, since $P(X = x)$ represents the proportion of subjects who entered level $X = x$ by free choice, and we cannot rule out the possibility that subjects who attain $X = x$ by free choice would react to $add(q)$ differently from subjects who "receive" $X = x$ by experimental decree. For example, it is quite possible that subjects who are highly sensitive to $add(q)$ would attempt to lower their X level, given the choice.

We translate into counterfactual analysis and write the inequality:

$$E[Y|add(q)] - E[Y] = \sum_x E[Y_{x+q}|x]P(X = x) \neq \sum_x E[Y_{x+q}]P(X = x)$$

Equality holds only when Y_x is independent of X, a condition that amounts to nonconfounding (see Theorem 4.3.1). Absent this condition, the estimation of $E[Y|add(q)]$ can be accomplished either by q-specific intervention or through stronger assumptions that enable the translation of ETT to *do*-expressions, as in Eq. (4.21).

PERSONAL DECISION MAKING

Example 4.4.3 *Ms Jones, a cancer patient, is facing a tough decision between two possible treatments: (i) lumpectomy alone or (ii) lumpectomy plus irradiation. In consultation with her oncologist, she decides on (ii). Ten years later, Ms Jones is alive, and the tumor has not recurred. She speculates: Do I owe my life to irradiation?* PN

Mrs Smith, on the other hand, had a lumpectomy alone, and her tumor recurred after a year. And she is regretting: I should have gone through irradiation. PS

Can these speculations ever be substantiated from statistical data? Moreover, what good would it do to confirm Ms Jones's triumph or Mrs Smith's regret?

Probability of Necessity ("I would have died if I hadn't treated myself")

$$PN = P(Y_0 = 0 | X = 1, Y = 1) \tag{4.23}$$

It reads: the probability that remission would *not* have occurred ($Y = 0$) had Ms Jones not gone through irradiation, given that she did in fact go through irradiation ($X = 1$), and remission did occur ($Y = 1$). The label PN stands for "probability of necessity" that measures the degree to which Ms Jones's decision was *necessary* for her positive outcome.

Probability of Sufficiency ("Treating myself was enough to not die.")

Similarly, the probability that Ms Smith's regret is justified is given by

$$PS = P(Y_1 = 1 | X = 0, Y = 0) \tag{4.24}$$

It reads: the probability that remission would have occurred had Mrs Smith gone through irradiation ($Y_1 = 1$), given that she did not in fact go through irradiation ($X = 0$), and remission did not occur ($Y = 0$). PS stands for the "probability of sufficiency," measuring the degree to which the action not taken, $X = 1$, would have been sufficient for her recovery.

- "Probabities of causation."
- Play a major role in questions of "attribution": from legal liability to personal decision making.
- They are estimable under certain conditions with both observational and experimental data.

Formally, Ms Daily's dilemma is to quantify the probability that irradiation is both *necessary and sufficient* for eliminating her tumor, or

$$PNS = P(Y_1 = 1, Y_0 = 0) \tag{4.25}$$

where Y_1 and Y_0 stand for remission under treatment (Y_1) and nontreatment (Y_0), respectively. Knowing this probability would help Ms Daily's assessment of how likely she is to belong to the group of individuals for whom $Y_1 = 1$ and $Y_0 = 0$.

PNS cannot be assesed from experimental studies. (Because we can never tell " " " whether an outcome would have been different had the person been assigned to a different treatment.)

However, assuming monotonicity ("irradiation cannot cause recurrence of a tumor that was about to remit.")

Theorem 4.5.1 *If Y is monotonic relative to X, that is, $Y_1(u) \geq Y_0(u)$ for all u, then PN is identifiable whenever the causal effect $P(y|do(x))$ is identifiable, and*

$$PN = \frac{P(y) - P(y|do(x'))}{P(x, y)} \tag{4.28}$$

or, substituting $P(y) = P(y|x)P(x) + P(y|x')(1 - P(x))$, we obtain

$$PN = \frac{P(y|x) - P(y|x')}{P(y|x)} + \frac{P(y|x') - P(y|do(x'))}{P(x, y)} \tag{4.29}$$

$$PNS = P(Y = 1|do(X = 1)) - P(Y = 1|do(X = 0))$$

If PNS = 25%, cancer will be cured with irradiation but will be back without it.

SEX DISCRIMINATION IN HIRING

Example 4.4.4 *Mary files a law suit against the New York-based XYZ International, alleging discriminatory hiring practices. According to her, she has applied for a job with XYZ International, and she has all the credentials for the job, yet she was not hired, allegedly because she mentioned, during the course of her interview, that she is gay. Moreover, she claims, the hiring record of XYZ International shows consistent preferences for straight employees. Does she have a case? Can hiring records prove whether XYZ International was discriminating when declining her job application?*

- Not population based, but one that appeals to the individual case of the plaintiff.
- We cannot intervene in the employer hiring had Mary been straight variable in an experimental setting.
- So, can observational data prove employer discrimination?

$$PS = P(Y_1 = 1 / \underbrace{X = 0}_{\text{interview perceives Mary as gay}}, \underbrace{Y = 0}_{\text{Mary is hired}})$$

- Although discrimination cannot be proved individually, the PC that " took place can be determined.

MEDIATION AND PATH-DISABLING INTERVENTION

Example 4.4.5 *A policy maker wishes to assess the extent to which gender disparity in hiring can be reduced by making hiring decisions gender-blind, rather than eliminating gender inequality in education or job training. The former concerns the "direct effect" of gender on hiring, whereas the latter concerns the "indirect effect," or the effect* mediated *via job qualification.*

CONTENDING POLICIES: Fighting employers' prejudices vs. New program in which ♀ obtain = education

Knowing that hiring disparities are due to prejudices would render educational reforms superfluos.

Because we are dealing with disabling processes rather than changing levels of variables, there is no way we can express the effect of such interventions using a *do*-operator, as we did in the mediation analysis of Section 3.7. We can express it, however, in a counterfactual language, using the desired end result as an antecedent. For example, if we wish to assess the

For X, Y, Z (Z mediator of X, Y)

the *controlled direct effect* (CDE) on Y of changing the value of X from x to x' is defined as

$$CDE = P(Y = y|do(X = x), do(Z = z)) - P(Y = y|do(X = x'), do(Z = z)) \quad (3.18)$$

The obvious advantage of this definition over the one based on conditioning is its generality; it captures the intent of "keeping Z constant" even in cases where the $Z \rightarrow Y$ relationship is

The analysis proceeds as follows: the hiring status (Y) of a female applicant with qualification $Q = q$, given that the employer treats her as though she is a male is captured by the counterfactual $Y_{X=1,Q=q}$, where $X = 1$ refers to being a male. But since the value q would vary among applicants, we need to average this quantity according to the distribution of female qualification, giving $\sum_q E(Y_{X=1,Q=q})P(Q = q|X = 0)$. Male applicants would have a similar chance at hiring except that the average is governed by the distribution of male qualification, giving

$$\sum_q E[Y_{X=1,Q=q}]P(Q = q|X = 1)$$

If we subtract the two quantities, we get

$$\sum_q E[Y_{X=1,Q=q}][P(Q = q|X = 0) - P(Q = q|X = 1)]$$

which is the indirect effect of gender on hiring, mediated by qualification. We call this effect the natural indirect effect (NIE), because we allow the qualification Q to vary naturally from applicant to applicant, as opposed to the controlled direct effect in Chapter 3, where we held the mediator at a constant level for the entire population. Here we merely disable the capacity of Y to respond to X but leave its response to Q unaltered.

The next question to ask is whether such a counterfactual expression can be identified from data. It can be shown (Pearl 2001) that, in the absence of confounding the NIE can be estimated by conditional probabilities, giving

$$\boxed{NIE = \sum_q E[Y|X = 1, Q = q][P(Q = q|X = 0) - P(Q = q|X = 1)]}$$

mediation formula