

Impact Evaluation in Firms and Organizations

With Applications in R and Python

Chapter 2: Basics of Impact Evaluation

2.1 The Fundamental Problem of Impact Evaluation

2.2 Characterizing the Impact

2.3 The Problem of Comparing Apples to Oranges

Goal of impact evaluation

- Assessing how a specific intervention influences an outcome
- Outcomes may be measures such as key performance indicators

Defining relevant outcomes

- Selection of outcomes should be guided by core business objectives and goals
- These considerations include risk tolerance and financial stability

Examples of business objectives

- Growth aims such as market expansion or new markets
- Profit-oriented goals like profitability or cost efficiency
- Metrics such as market share or sales per customer

Core challenge in impact evaluation

- Interest in causes and effects of interventions on outcomes
- Inability to observe the world with and without intervention simultaneously $\Rightarrow$ fundamental problem of causal inference

Example

Evaluating the impact of a marketing campaign

- Evaluating a marketing campaign's impact on store sales
- We would like to compare sales with and without the campaign
- However, a store can only run or not run the campaign but not do both simultaneously

More Examples of the Fundamental Problem

Example

Pricing strategy evaluation

- Discount offered when purchasing products in combination
- Same customer cannot be observed with and without discount

Employee training programs

- Training aimed at improving productivity and performance
- Cannot observe the same employee with and without training

Implications for causal analysis

- Causal effects cannot be directly observed because of mutually exclusive treatment states
- Challenge for evaluating effectiveness of interventions

Assessing impact despite limitations

- Fundamental problem prevents direct observation of causal effects
- Impact assessment is only possible under specific behavioral assumptions

Role of causal methods

- Methods rely on assumptions about human behavior
- Plausibility of assumptions depends on the context
- Implementation of impact evaluation is feasible using statistical software such as R and Python

2.1 The Fundamental Problem of Impact Evaluation

2.2 Characterizing the Impact

2.3 The Problem of Comparing Apples to Oranges

Use of variables

- Use letters to represent variables
- Variables can describe interventions and outcomes

Intervention variable D

- D denotes the intervention of interest
- Takes different values depending on whether intervention is introduced or not
- $D = 1$ indicates intervention and $D = 0$ indicates no intervention

Outcome variable Y

- Y denotes the outcome of interest
- Takes numerical values based on observed outcomes

Causal graphs

- Causal graphs can display relations between variables
- Causal arrow from D to Y represents the effect of interest

Figure 2.1: Impact of intervention (D) on outcome (Y)

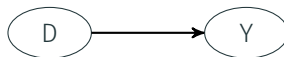


Figure 2.2: Impact of a marketing intervention on sales outcome



Potential outcomes

- $Y(1)$ denotes outcome with the intervention
- $Y(0)$ denotes outcome without the intervention
- Impact is given by $Y(1) - Y(0)$

Fundamental problem of causal inference

- Relationship between observed and potential outcomes:

$$Y = Y(1) \text{ if } D = 1, \quad Y = Y(0) \text{ if } D = 0 \quad (2.2)$$

- $Y(1)$ and $Y(0)$ are never observed simultaneously for the same unit
- Therefore, $Y(1) - Y(0)$ cannot be calculated directly

Potential Outcomes and Impact

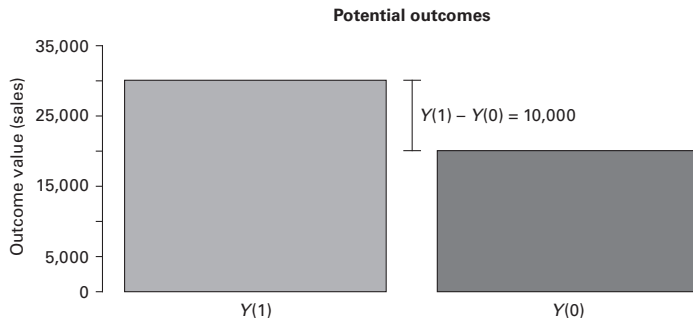


Figure 2.3: Potential outcomes and impact

Overcoming the fundamental problem

- Impact $Y(1) - Y(0)$ cannot be calculated directly because $Y(1)$ and $Y(0)$ are never observed simultaneously
- Overcoming this problem requires specific behavioral assumptions using appropriate causal methods

Stable unit treatment value assumption (SUTVA)

- A store's sales are assumed to depend only on its own intervention and not on interventions in other stores
- Spillover (or interference) effects may arise, for example, when advertising in one store affects sales in another store
- SUTVA rules out such spillover or interference effects

Evaluating effects at an aggregate level

- Individual causal effects cannot be directly observed
- Assumptions such as SUTVA allow evaluation at the group level

Average causal effect

- Focus on average impact in a population
- For example, the average effect of a marketing campaign on sales across stores

Formal definition

- The average causal effect is denoted by Δ
- Defined as the expected difference in potential outcomes

$$\Delta = E[Y(1) - Y(0)] \quad (2.2)$$

where $E[\cdot]$ denotes the average in the population of interest

Average treatment effect (ATE)

- Average causal effect also called average treatment effect
- Measures the impact of an intervention in the whole population

Average treatment effect on the treated (ATET or ATT)

- Impact among units receiving the intervention

$$\Delta_{D=1} = E[Y(1) - Y(0)|D = 1] \quad (2.3)$$

Average treatment effect on the nontreated (ATENT)

- Impact among units not receiving the intervention

$$\Delta_{D=0} = E[Y(1) - Y(0)|D = 0] \quad (2.4)$$

2.1 The Fundamental Problem of Impact Evaluation

2.2 Characterizing the Impact

2.3 The Problem of Comparing Apples to Oranges

The Problem with Naïve Comparison

Naïve comparison

- Compare average outcomes of treatment and control groups
- Treatment group receives the intervention
- Control group does not receive the intervention

Requirement for validity

- Treatment and control groups must be comparable in background characteristics that affect the outcome of interest

Problem in practice

- Treatment and control groups often differ systematically
- For that reason, these outcome comparisons may be contaminated by differing background characteristics
- Simple comparisons may yield biased effects

Selection Bias in Loyalty Card Intervention

Loyalty card intervention

- Shop issues loyalty cards
- Outcome of interest is purchase volume

Issue with naïve comparison

- Customers with loyalty cards may buy more even without the card because card possession is correlated with prior shopping behavior

Differences in buying intent

- Frequent shoppers see higher benefits from a loyalty card
- Less frequent shoppers gain little or no benefit
- Treatment and control groups thus differ in buying intent

Customer churn and discounts

- Customer churn refers to switching service providers
- Discounts may be used as an intervention to increase customer retention

Nonrandom discount assignment

- Discounts targeted at customers with high churn intention
- Leads to treated and untreated customers differing systematically

Differences in churn intention

- Simple comparison mixes discount effects and churn intention
- Treatment and control groups differ in churn intention biasing the comparison

Definition

- Treatment and control groups differ in background characteristics
- These characteristics affect the outcome of interest

Implications

- Difference in average observed outcomes does not equal intervention impact
- Comparing $D = 1$ and $D = 0$ may mix multiple effects
- Naïve comparison of treatment and control outcomes is invalid in the presence of differences in background characteristics

Graphical Representation of Selection Bias

Causal graph

- Introduction of U for background characteristics
- U may represent buying intent or churn intent
- U affects both intervention D and outcome Y

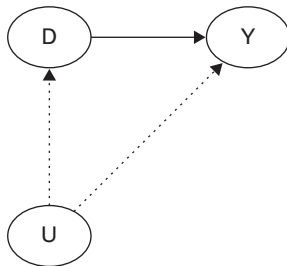
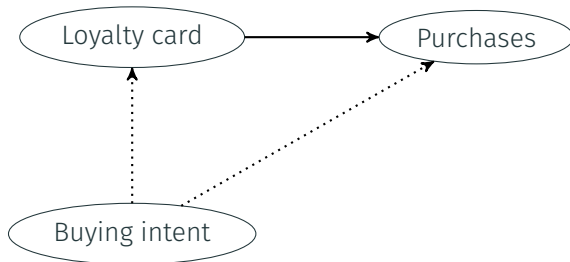


Figure 2.4: Selection bias

Selection Bias When Evaluating a Loyalty Card

Figure 2.5: Selection bias when evaluating a loyalty card



Requirement for valid impact evaluation

- Treated and control groups must be comparable in background characteristics U

Ceteris paribus condition

- Means everything else equal apart from the intervention
- If U affects Y , it must not affect D
- Background characteristics must not differ systematically

Implications for causal effects

- Differences in outcomes attributed to the intervention alone
- No confounding by background characteristics
- Allows identification of the average treatment effect

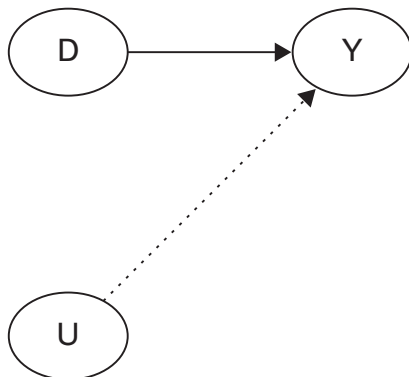


Figure 2.6: No selection bias

Incomparable Treatment and Control Groups

Visual representation of selection bias

- Figures represent subjects in treatment and control groups
- Same colors indicate shared background characteristics
- Treatment and control groups differ in their composition

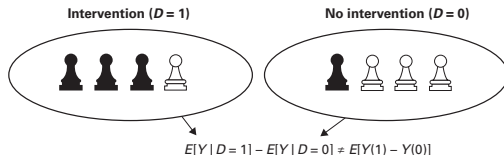


Figure 2.7: Incomparable treatment and control groups

Implication for impact evaluation

- Average outcome differences do not correspond to the ATE

When Simple Differences Identify the ATE

Absence of selection bias

- Treatment and control groups are comparable
- No systematic differences in background characteristics

Implications for impact evaluation

- Difference in average outcomes equals the ATE
- Formally

$$E[Y|D = 1] - E[Y|D = 0] = \underbrace{E[Y(1) - Y(0)]}_{\Delta} \quad (2.6)$$

Interpretation

- Groups are comparable and represent the full population
- Compares apples with apples

Selection bias preventing identification of ATE

- Treatment and control groups differ systematically in background characteristics
- Treatment and control groups are not representative
- ATE cannot be identified and thus comparison yields biased estimates

Why selection bias is common in practice

- Observed data often reflect strategic human behavior
- Personal characteristics influence intervention decisions