Introduction to Causal Inference: Propensity Score

Quan Zhang, Ph.D.

Introduction

- AB test is a randomized controlled trial. The random assignment of units gurantees that the only difference in the outcome between the test and control groups is due to the treatment

Introduction

- AB test is a randomized controlled trial. The random assignment of units gurantees that the only difference in the outcome between the test and control groups is due to the treatment
- Randomized controlled trials are generally expensive

Introduction

- AB test is a randomized controlled trial. The random assignment of units gurantees that the only difference in the outcome between the test and control groups is due to the treatment
- Randomized controlled trials are generally expensive
- There is a growing interest in using observational (or nonrandomized) studies to estimate the effects of treatments on outcomes.

Introduction

- AB test is a randomized controlled trial. The random assignment of units gurantees that the only difference in the outcome between the test and control groups is due to the treatment
- Randomized controlled trials are generally expensive
- There is a growing interest in using observational (or nonrandomized) studies to estimate the effects of treatments on outcomes.
- However, in observational studies treatment selection is often influenced by subject characteristics. As a result, baseline characteristics of treated subjects often differ systematically from those of untreated subjects.

Introduction

- AB test is a randomized controlled trial. The random assignment of units gurantees that the only difference in the outcome between the test and control groups is due to the treatment
- Randomized controlled trials are generally expensive
- There is a growing interest in using observational (or nonrandomized) studies to estimate the effects of treatments on outcomes.
- However, in observational studies treatment selection is often influenced by subject characteristics. As a result, baseline characteristics of treated subjects often differ systematically from those of untreated subjects.
- E.g., students with a bachelor's degree in maths earn more money at Microsoft. Major in maths $\stackrel{?}{\rightarrow}$ more money earned

Introduction

- AB test is a randomized controlled trial. The random assignment of units gurantees that the only difference in the outcome between the test and control groups is due to the treatment
- Randomized controlled trials are generally expensive
- There is a growing interest in using observational (or nonrandomized) studies to estimate the effects of treatments on outcomes.
- However, in observational studies treatment selection is often influenced by subject characteristics. As a result, baseline characteristics of treated subjects often differ systematically from those of untreated subjects.
- E.g., students with a bachelor's degree in maths earn more money at Microsoft. Major in maths $\stackrel{?}{\rightarrow}$ more money earned
- E.g., a master's degree $\stackrel{?}{\rightarrow}$ more money earned

Introduction

- Therefore, one must account for systematic differences in baseline characteristics between treated and untreated subjects when estimating the effect of treatment on outcomes.

Introduction

- Therefore, one must account for systematic differences in baseline characteristics between treated and untreated subjects when estimating the effect of treatment on outcomes.
- Historically, applied researchers have relied on the use of regression adjustment to account for differences in measured baseline characteristics between treated and untreated subjects.

Introduction

- Therefore, one must account for systematic differences in baseline characteristics between treated and untreated subjects when estimating the effect of treatment on outcomes.
- Historically, applied researchers have relied on the use of regression adjustment to account for differences in measured baseline characteristics between treated and untreated subjects.
- Recently, there has been increasing interest in methods based on the propensity score to reduce or eliminate the effects of confounding when using observational data.

Introduction

- Therefore, one must account for systematic differences in baseline characteristics between treated and untreated subjects when estimating the effect of treatment on outcomes.
- Historically, applied researchers have relied on the use of regression adjustment to account for differences in measured baseline characteristics between treated and untreated subjects.
- Recently, there has been increasing interest in methods based on the propensity score to reduce or eliminate the effects of confounding when using observational data.
- The Nobel Prize in Economics 2021 is awarded with one half to Card, and the other half jointly to Angrist and Imbens "for their methodological contributions to the analysis of causal relationships"

Confounding

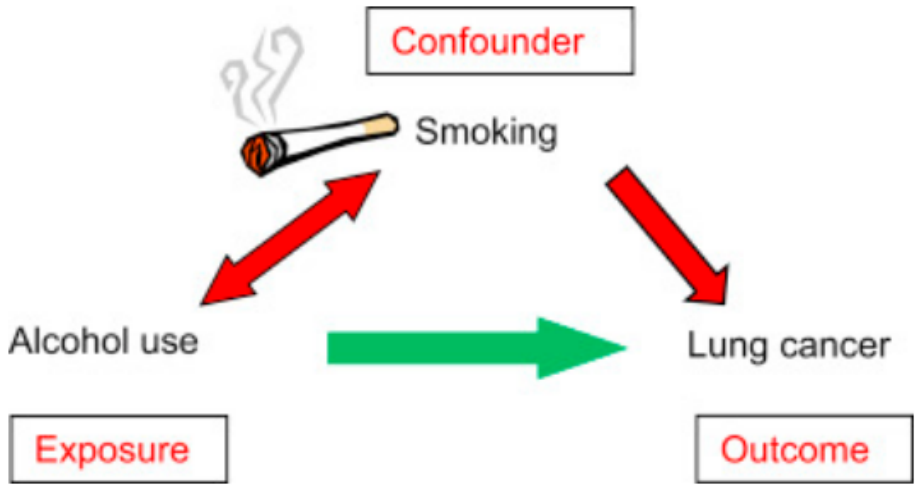
- Definition
  Confounding is a situation in which the effect or association between an exposure (a predictor or risk factor) and outcome is distorted by the presence of a variable;

Confounding

- Definition
  Confounding is a situation in which the effect or association between an exposure (a predictor or risk factor) and outcome is distorted by the presence of a variable;
  this variable is not under investigation. The variable is called a confounding variable, or a confounder

Confounding

- Definition
  Confounding is a situation in which the effect or association between an exposure (a predictor or risk factor) and outcome is distorted by the presence of a variable;
  this variable is not under investigation. The variable is called a confounding variable, or a confounder



Confounding

- A confounder
  - is a risk factor for the outcome
  - is associated with the exposure
  - is not in the causal pathway between the exposure and the outcome

Example

Dental health and Heart disease

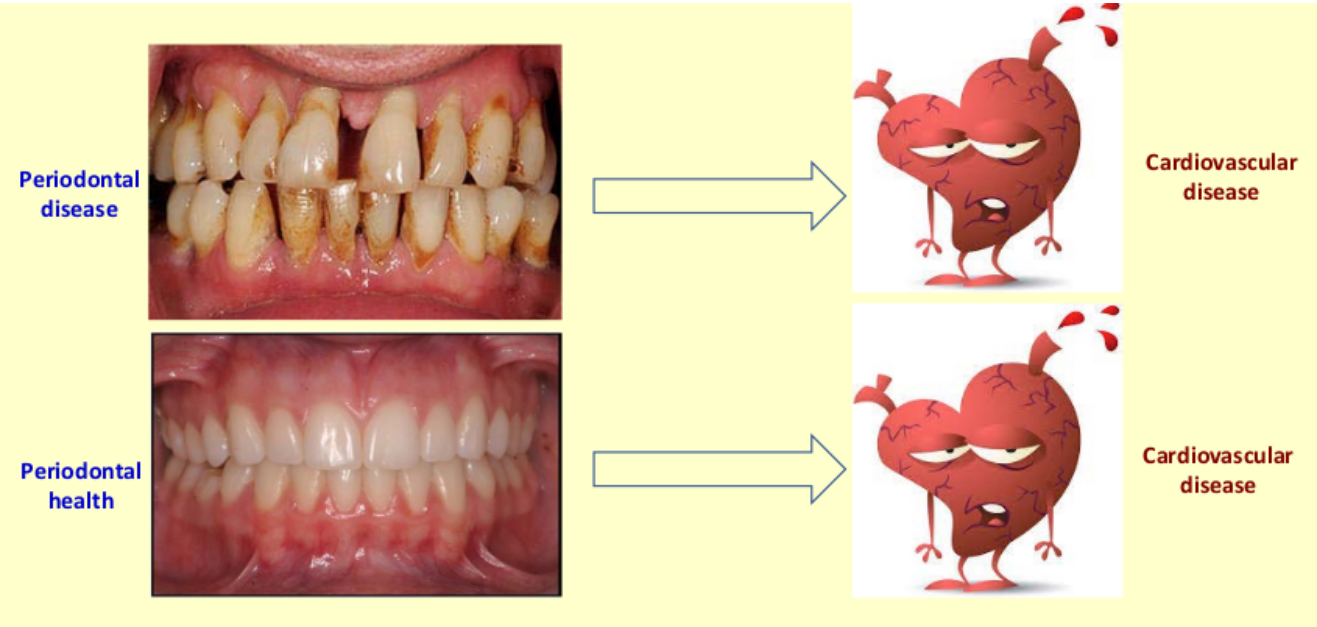


Figure: Example of confounding

Example

Dental health and Heart disease: We don't know whether the increase in the incidence is due to periodontal disease or smoking

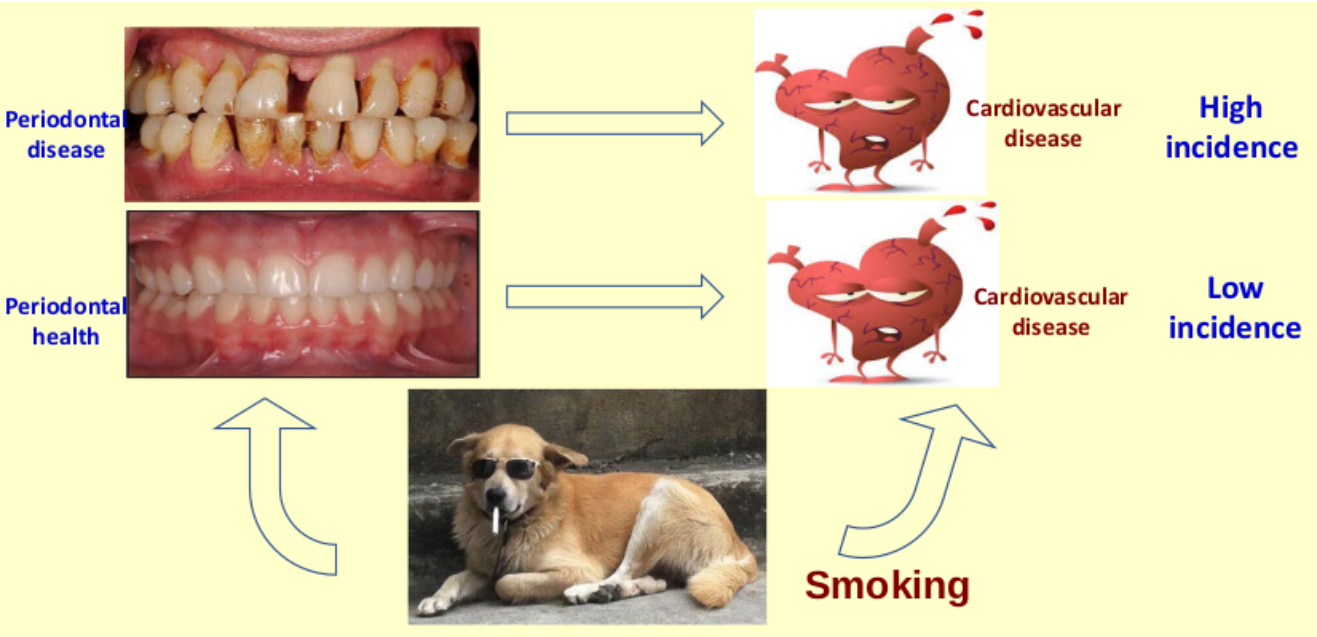


Figure: Example of confounding

Confounding effects

- Positive confounding: the observed association is biased away from the null
  $\rightarrow$ overestimation of the true association

Confounding effects

- Positive confounding: the observed association is biased away from the null
  $\rightarrow$ overestimation of the true association

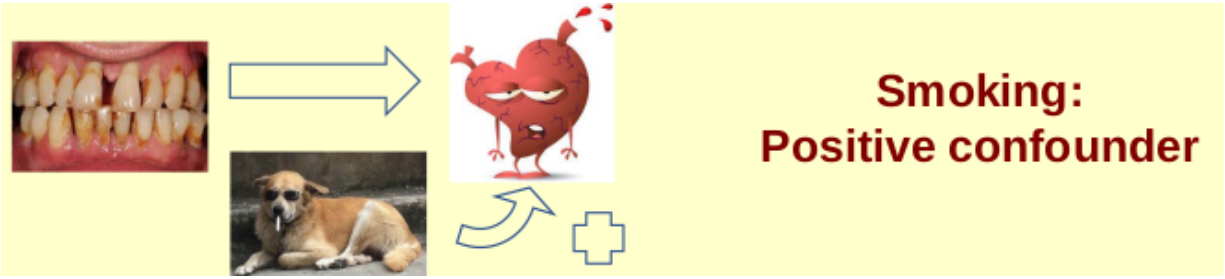


Figure: Positive confounding

Confounding effects

- Positive confounding: the observed association is biased away from the null
  $\rightarrow$ overestimation of the true association

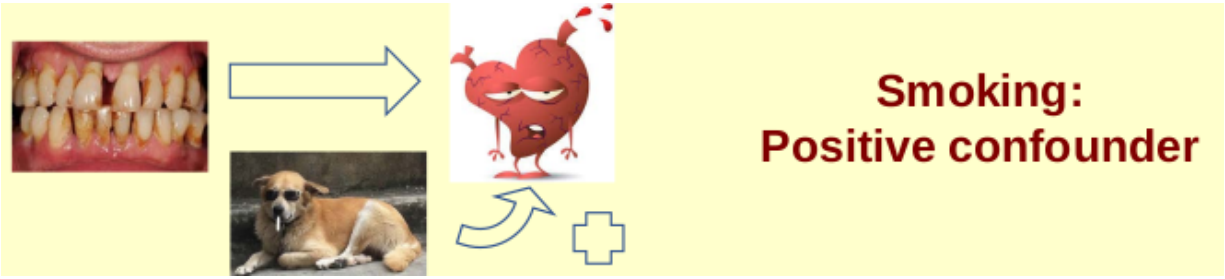


Figure: Positive confounding

- Negative confounding: underestimation of the true association

Confounding effects

- Positive confounding: the observed association is biased away from the null
  $\rightarrow$ overestimation of the true association

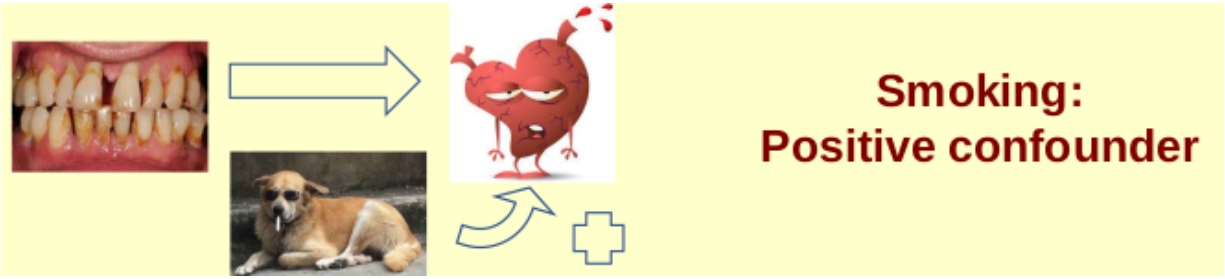


Figure: Positive confounding

- Negative confounding: underestimation of the true association

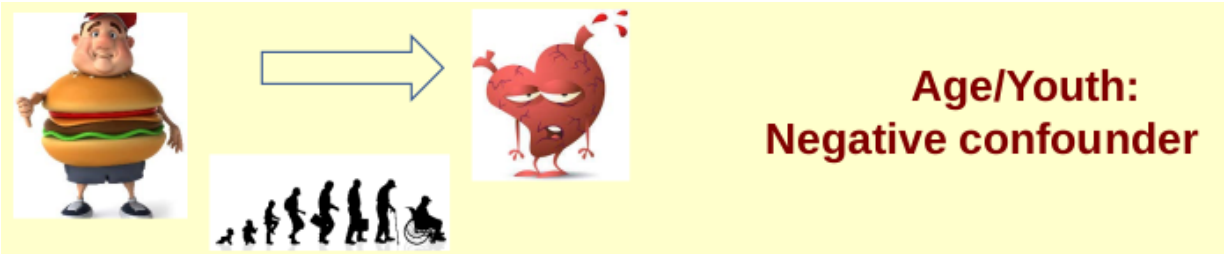


Figure: Negative confounding

Methods to address confounding

- Controlled in the design phase $\rightarrow$ experiment designs
  - Randomization
  - Restriction
  - Matching
  - Stratification

Methods to address confounding

- Controlled in the design phase $\rightarrow$ experiment designs
  - Randomization
  - Restriction
  - Matching
  - Stratification
- Controlled in the analysis phase
  - Stratified analysis
  - Propensity scores

It is more desirable to handle it in the Design Phase; but some ideal methods might not be possible.

Design: randomization

Imagine an AB test

- Randomly assign subjects to the treatment or control group
- It removes bias in the treatment assignment
- It controls both known and unknown confounders
- It guarantees that statistical tests will have valid significance levels
- In short, it is the Gold Standard for experiment designs (in biomedical and social research)

Design: restriction

- Exclusion of individuals with confounding factors or restriction to specific groups.
  - Example 1: Exclusion of smokers from a heart disease study
  - Example 2: Inclusion only males between 40-45 years in a study of relationship between heart disease and overweight

Design: restriction

- Exclusion of individuals with confounding factors or restriction to specific groups.
  - Example 1: Exclusion of smokers from a heart disease study
  - Example 2: Inclusion only males between 40-45 years in a study of relationship between heart disease and overweight
- Limitations
  - Reduces the number of eligible individuals
  - Restriction limits generalizability
  - Unable to evaluate the effects of factors that are restricted for

Design: matching

- Each pair of persons enrolled in a study are similar for one or more characteristics
  - Example: When study the causal relationship between periodontal disease and heart disease, if a 60 year-old Caucasian smoker with periodontal disease is entered then a 60 year old Caucasian smoker without periodontal disease will also be included

Design: matching

- Each pair of persons enrolled in a study are similar for one or more characteristics
  - Example: When study the causal relationship between periodontal disease and heart disease, if a 60 year-old Caucasian smoker with periodontal disease is entered then a 60 year old Caucasian smoker without periodontal disease will also be included
- Limitations
  - Time-consuming and expensive
  - Limits sample size
  - Only for a small number of confounding factors
  - Unable to evaluate the effect of the factors that have been matched

It is very hard, even not possible, to match when there are many confounders or potential confounders; substantially reduced sample size.

Design/Analysis: stratification

- Control for confounding by creating two or more categories or subgroups (strata) in which the confounding variable does not vary.
  - Example 1: Divide subjects with and without periodontal disease into groups based on smoking status: smokers and non-smokers.
  - Example 2: Divide subjects into different age groups, such as ¡30, 30-35, 36-40, ...

Design/Analysis: stratification

- Control for confounding by creating two or more categories or subgroups (strata) in which the confounding variable does not vary.
  - Example 1: Divide subjects with and without periodontal disease into groups based on smoking status: smokers and non-smokers.
  - Example 2: Divide subjects into different age groups, such as ¡30, 30-35, 36-40, ...
- Limitations
  - Unable to control simultaneously for multiple confounding variables
  - Limits sample size
  - Time consuming

Why not always randomized controlled experiments

In practice, though Randomized controlled experiments/trials is the best choice, sometimes it is hard to carry out.

Why not always randomized controlled experiments

In practice, though Randomized controlled experiments/trials is the best choice, sometimes it is hard to carry out.

- Unethical: often in biomeical studies and clinical trials
- Infeasible
- Not scientifically or financially justified

Causal inference

- Observational studies are always easy to carry out, but hard to draw causal inference.

Causal inference

- Observational studies are always easy to carry out, but hard to draw causal inference.
- Treatment selection is influenced by subject characteristics, which is a nonrandomized and uncontrolled process

Causal inference

- Observational studies are always easy to carry out, but hard to draw causal inference.
- Treatment selection is influenced by subject characteristics, which is a nonrandomized and uncontrolled process
- How to account for this systematic difference?

Propensity score

Propensity score

- Propensity score is a method which has been grown in popularity. It creates, in the analysis phase, similar groups, somewhat similar to post hoc stratification.

Propensity score

- Propensity score is a method which has been grown in popularity. It creates, in the analysis phase, similar groups, somewhat similar to post hoc stratification.
- Propensity score is the probability of a subject being assigned to the treatment group conditional on observed baseline covariates (Rosebaum and Rubin, 1983)

Propensity score

- Propensity score is a method which has been grown in popularity. It creates, in the analysis phase, similar groups, somewhat similar to post hoc stratification.
- Propensity score is the probability of a subject being assigned to the treatment group conditional on observed baseline covariates (Rosebaum and Rubin, 1983)
- It uses logistic regression or other binary classification approaches

Propensity score: implementation

- Logistic regression (or other binary classification models capable of predictive probability)
  - Response variable: $z = 1$ for Treatment and 0 for Control
  - Covariates: all features $\boldsymbol{x}$, including all possible confounders

Propensity score: implementation

- Logistic regression (or other binary classification models capable of predictive probability)
  - Response variable: $z = 1$ for Treatment and 0 for Control
  - Covariates: all features $\boldsymbol{x}$, including all possible confounders
- For each subject i with $\boldsymbol{x}_i$, the propensity score is $\hat{p}(z_i = 1 \mid \boldsymbol{x}_i)$

Propensity score: rationale

- The propensity score is the probability of treatment assignment conditional on observed baseline covariates.

Propensity score: rationale

- The propensity score is the probability of treatment assignment conditional on observed baseline covariates.
- If subjects have the same propensity score (probability to be assigned to the treatment group that is determined by a logistic regression model), overall, they will have the same features.

Propensity score: rationale

- The propensity score is the probability of treatment assignment conditional on observed baseline covariates.
- If subjects have the same propensity score (probability to be assigned to the treatment group that is determined by a logistic regression model), overall, they will have the same features.
- The propensity score allows one to design and analyze an observational (nonrandomized) study so that it mimics some of the characteristics of a randomized controlled trial.

Propensity score: rationale

- The propensity score is the probability of treatment assignment conditional on observed baseline covariates.
- If subjects have the same propensity score (probability to be assigned to the treatment group that is determined by a logistic regression model), overall, they will have the same features.
- The propensity score allows one to design and analyze an observational (nonrandomized) study so that it mimics some of the characteristics of a randomized controlled trial.
- In particular, the propensity score is a balancing score: conditional on the propensity score, the distribution of observed baseline covariates will be similar between treated and untreated subjects.

Propensity score: rationale

- The propensity score is the probability of treatment assignment conditional on observed baseline covariates.
- If subjects have the same propensity score (probability to be assigned to the treatment group that is determined by a logistic regression model), overall, they will have the same features.
- The propensity score allows one to design and analyze an observational (nonrandomized) study so that it mimics some of the characteristics of a randomized controlled trial.
- In particular, the propensity score is a balancing score: conditional on the propensity score, the distribution of observed baseline covariates will be similar between treated and untreated subjects.
- The propensity score as a scalar represents an overall status of high-dimensional covariates

Propensity score

- Three simple methods using propensity score
  - Matching on the propensity score
  - Stratification on the propensity score
  - Covariate adjustment using the propensity score

Propensity score matching

- Forming matched sets of treated and untreated subjects who share a similar value (up to some level of difference) of the propensity score.

Propensity score matching

- Forming matched sets of treated and untreated subjects who share a similar value (up to some level of difference) of the propensity score.
- The most common implementation of propensity score matching is one-to-one or pair matching.

Propensity score matching

- Forming matched sets of treated and untreated subjects who share a similar value (up to some level of difference) of the propensity score.
- The most common implementation of propensity score matching is one-to-one or pair matching.
- Once a matched sample has been formed, the treatment effect can be estimated by directly comparing outcomes between treated and untreated subjects in the matched pair.

Propensity score matching

- Forming matched sets of treated and untreated subjects who share a similar value (up to some level of difference) of the propensity score.
- The most common implementation of propensity score matching is one-to-one or pair matching.
- Once a matched sample has been formed, the treatment effect can be estimated by directly comparing outcomes between treated and untreated subjects in the matched pair.
  - If the outcome is continuous (e.g., a depression scale), the effect of treatment can be estimated as the difference between the mean outcome for treated subjects and the mean outcome for untreated subjects

Propensity score matching

- Forming matched sets of treated and untreated subjects who share a similar value (up to some level of difference) of the propensity score.
- The most common implementation of propensity score matching is one-to-one or pair matching.
- Once a matched sample has been formed, the treatment effect can be estimated by directly comparing outcomes between treated and untreated subjects in the matched pair.
  - If the outcome is continuous (e.g., a depression scale), the effect of treatment can be estimated as the difference between the mean outcome for treated subjects and the mean outcome for untreated subjects
  - If the outcome is binary (e.g., heart disease or not), the effect of treatment can be estimated as the difference between the proportion of subjects experiencing the event in each of the two groups (treated vs. untreated)

Propensity score matching

Matched pairs are often formed without replacement; and there are two different ways to achieve balanced matched samples: greedy or optimal matching.

Propensity score matching

Matched pairs are often formed without replacement; and there are two different ways to achieve balanced matched samples: greedy or optimal matching.

- In **greedy matching**, a treated subject is first selected at random. The untreated subject whose propensity score is closest to that of this randomly selected treated subject is chosen for matching to this treated subject. This process is then repeated until one has exhausted the list of treated subjects.

Propensity score matching

Matched pairs are often formed without replacement; and there are two different ways to achieve balanced matched samples: greedy or optimal matching.

- In **greedy matching**, a treated subject is first selected at random. The untreated subject whose propensity score is closest to that of this randomly selected treated subject is chosen for matching to this treated subject. This process is then repeated until one has exhausted the list of treated subjects.
- This process is called greedy because at each step in the process, the nearest untreated subject is selected for matching to the given treated subject, even if that untreated subject would better serve as a match for a subsequent treated subject.

Propensity score matching

- In **optimal matching**, matches are formed to minimize the total within-pair difference of the propensity score.
- Gu and Rosenbaum, 1993 (Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics 2, 405–420) compared greedy and optimal matching and found that optimal matching did no better than greedy matching in producing balanced matched samples.

Propensity score matching

- In **optimal matching**, matches are formed to minimize the total within-pair difference of the propensity score.
- Gu and Rosenbaum, 1993 (Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics 2, 405–420) compared greedy and optimal matching and found that optimal matching did no better than greedy matching in producing balanced matched samples.

The processes of greedy and optimal matchings are different but the performance and results are similar

Stratification on the propensity score

- Stratification on the propensity score involves stratifying subjects into mutually exclusive subsets based on their estimated propensity scores

Stratification on the propensity score

- Stratification on the propensity score involves stratifying subjects into mutually exclusive subsets based on their estimated propensity scores
- Subjects are ranked according to their estimated propensity scores, and then stratified into subsets based on predefined thresholds of the estimated propensity score

Stratification on the propensity score

- Stratification on the propensity score involves stratifying subjects into mutually exclusive subsets based on their estimated propensity scores
- Subjects are ranked according to their estimated propensity scores, and then stratified into subsets based on predefined thresholds of the estimated propensity score
- A common approach is to divide subjects into five roughly equal sized groups using the quintiles (i.e., 20%, 40%, 60%, and 80% quantiles) of the estimated propensity score.

Stratification on the propensity score

- Cochran, 1968 (The effectiveness of adjustment by sub- classification in removing bias in observational studies; Biometrics 24, 295–313) demonstrated that stratifying on the quintiles of **a continuous confounding variable** eliminated approximately 90% of the bias due to that variable

Stratification on the propensity score

- Cochran, 1968 (The effectiveness of adjustment by sub- classification in removing bias in observational studies; Biometrics 24, 295–313) demonstrated that stratifying on the quintiles of **a continuous confounding variable** eliminated approximately 90% of the bias due to that variable
- Rosenbaum and Rubin, 1984 (Reducing bias in observational studies using sub-classification on the propensity score; Journal of the American Statistical Association 79, 516–524) extended this result to **stratification on the propensity score**, stating that stratifying on the quintiles of the propensity score eliminates approximately 90% of the bias due to measured **confounders** when estimating a linear treatment effect

Stratification on the propensity score

- Stratification on the propensity can be conceptualized as an analysis of a set of (five) randomized controlled trials. Within each stratum, the effect of treatment on outcomes can be estimated by comparing outcomes directly between treated and untreated subjects. The stratum-specific estimates of treatment effect can then be pooled across strata to estimate an overall treatment effect (e.g. using weighted average)

Covariate adjustment using the propensity score

A Regression approach

- The outcome is regressed on an indicator variable denoting treatment status and the estimated propensity score

Covariate adjustment using the propensity score

A Regression approach

- The outcome is regressed on an indicator variable denoting treatment status and the estimated propensity score
- The choice of regression model would depend on the nature of the outcome (linear regression or logistic regression)

Covariate adjustment using the propensity score

A Regression approach

- The outcome is regressed on an indicator variable denoting treatment status and the estimated propensity score
- The choice of regression model would depend on the nature of the outcome (linear regression or logistic regression)
- The treatment effect is determined using the estimated regression coefficient from the fitted model

Covariate adjustment using the propensity score

A Regression approach

- The outcome is regressed on an indicator variable denoting treatment status and the estimated propensity score
- The choice of regression model would depend on the nature of the outcome (linear regression or logistic regression)
- The treatment effect is determined using the estimated regression coefficient from the fitted model
  - For a linear model, the treatment effect is interpreted an adjusted difference in means

Covariate adjustment using the propensity score

A Regression approach

- The outcome is regressed on an indicator variable denoting treatment status and the estimated propensity score
- The choice of regression model would depend on the nature of the outcome (linear regression or logistic regression)
- The treatment effect is determined using the estimated regression coefficient from the fitted model
  - For a linear model, the treatment effect is interpreted an adjusted difference in means
  - For a logistic model the treatment effect is interpreted as an adjusted odds ratio

Covariate adjustment using the propensity score

- Propensity score matching and stratification are model-free

Covariate adjustment using the propensity score

- Propensity score matching and stratification are model-free
- Covariate adjustment using the propensity score requires a linear or logistic regression; this method assumes that the relationship between the propensity score and the outcome has been correctly modeled

Covariate adjustment using the propensity score

- Propensity score matching and stratification are model-free
- Covariate adjustment using the propensity score requires a linear or logistic regression; this method assumes that the relationship between the propensity score and the outcome has been correctly modeled
- Instead of using multiple predictors in a regression, all baseline characteristics are combined (through propensity score) into one "1index" (the propensity score). This makes it simpler to check for model assumptions

Comparison of the three propensity score methods

- Propensity score matching eliminates a **bigger** proportion of the systematic differences in baseline characteristics between treated and untreated subjects than stratification on the propensity score or covariate adjustment using the propensity score (Austin, P. C. Type I error rates, coverage of confidence intervals, and variance estimation in propensity-score matched analyses. The International Journal of Biostatistics, 5, Article 13; 2009)

Comparison of the three propensity score methods

- Propensity score matching eliminates a **bigger** proportion of the systematic differences in baseline characteristics between treated and untreated subjects than stratification on the propensity score or covariate adjustment using the propensity score (Austin, P. C. Type I error rates, coverage of confidence intervals, and variance estimation in propensity-score matched analyses. The International Journal of Biostatistics, 5, Article 13; 2009)
- However, propensity score matching is more time-consuming and potentially limits the sample size since some subjects may not able to match

Balance diagnostics

- The true propensity score is a balancing score: given the true propensity score, the distribution of measured baseline covariates is independent of treatment assignment

Balance diagnostics

- The true propensity score is a balancing score: given the true propensity score, the distribution of measured baseline covariates is independent of treatment assignment
- In an observational study the true propensity score is not known. It must be estimated using the data

Balance diagnostics

- The true propensity score is a balancing score: given the true propensity score, the distribution of measured baseline covariates is independent of treatment assignment
- In an observational study the true propensity score is not known. It must be estimated using the data
- An important component of any propensity score analysis is to examine whether the propensity score model has been adequately specified

Variable selection for the propensity score model

- A binary classification model is used to estimate the propensity score

Variable selection for the propensity score model

- A binary classification model is used to estimate the propensity score
- Due to the small sample size, NNs or other richly parameterized models may not be suitable (overfit)

Variable selection for the propensity score model

- A binary classification model is used to estimate the propensity score
- Due to the small sample size, NNs or other richly parameterized models may not be suitable (overfit)
- Variable selection is often needed in a logistic model if used. Possible sets of variables include
  - All measured baseline covariates
  - All baseline covariates that are associated with treatment assignment
  - All potential and true confounders

Variable selection for the propensity score model

- A binary classification model is used to estimate the propensity score
- Due to the small sample size, NNs or other richly parameterized models may not be suitable (overfit)
- Variable selection is often needed in a logistic model if used. Possible sets of variables include
  - All measured baseline covariates
  - All baseline covariates that are associated with treatment assignment
  - All potential and true confounders
- In practice, it may be difficult to determine true confounders among baseline variables
- In many settings, most subject-level baseline covariates likely affect both treatment assignment and the outcome. Therefore, it is likely that one can safely include all measured baseline characteristics in the propensity score model.

Balance diagnostics

- Covariate balance is the degree to which the distribution of covariates is similar across levels of the treatment (binary treatment for treatment-control trials)

Balance diagnostics

- Covariate balance is the degree to which the distribution of covariates is similar across levels of the treatment (binary treatment for treatment-control trials)
- For propensity score matching, assessing whether the propensity score model has been adequately specified involves comparing treated and untreated subjects within the propensity score matched pairs

Balance diagnostics

- Covariate balance is the degree to which the distribution of covariates is similar across levels of the treatment (binary treatment for treatment-control trials)
- For propensity score matching, assessing whether the propensity score model has been adequately specified involves comparing treated and untreated subjects within the propensity score matched pairs
- For stratification on the propensity score, the assessment includes comparing treated and untreated subjects within strata of the propensity score

Balance diagnostics

- Covariate balance is the degree to which the distribution of covariates is similar across levels of the treatment (binary treatment for treatment-control trials)
- For propensity score matching, assessing whether the propensity score model has been adequately specified involves comparing treated and untreated subjects within the propensity score matched pairs
- For stratification on the propensity score, the assessment includes comparing treated and untreated subjects within strata of the propensity score
- `R` package: `MatchIt`

Unconfoundedness

- If balance diagnostics are successful, does this make Propensity Score as good as Randomization?

Unconfoundedness

- If balance diagnostics are successful, does this make Propensity Score as good as Randomization?
- A short answer is NO

Unconfoundedness

- If balance diagnostics are successful, does this make Propensity Score as good as Randomization?
- A short answer is NO
- A key assumption for propensity score methods to be valid is *unconfoundedness*; it is violated if there exist unobserved confounders

Unconfoundedness

- If balance diagnostics are successful, does this make Propensity Score as good as Randomization?
- A short answer is NO
- A key assumption for propensity score methods to be valid is *unconfoundedness*; it is violated if there exist unobserved confounders
- Using propensity score methods, one assumes all confounders are observed as covariates in the data.

Unconfoundedness

- If balance diagnostics are successful, does this make Propensity Score as good as Randomization?
- A short answer is NO
- A key assumption for propensity score methods to be valid is *unconfoundedness*; it is violated if there exist unobserved confounders
- Using propensity score methods, one assumes all confounders are observed as covariates in the data.
- In comparison, randomization controls both known and unknown confounders

Treatment effect

After Propensity Scores are generated, it is possible to estimate the treatment effect

- If Matching is used, one can estimate within each pair, the average them out

Treatment effect

After Propensity Scores are generated, it is possible to estimate the treatment effect

- If Matching is used, one can estimate within each pair, the average them out
- If Stratification is used, one can similarly estimate within each stratum, then weighted average them out

Treatment effect

After Propensity Scores are generated, it is possible to estimate the treatment effect

- If Matching is used, one can estimate within each pair, the average them out
- If Stratification is used, one can similarly estimate within each stratum, then weighted average them out
- If Covariate Adjustment is used, the effect of treatment is determined using the estimated regression coefficient from the fitted regression model

Other causal inference methods: instrumental variable

What if there are unobserved confounders?

Other causal inference methods: instrumental variable

What if there are unobserved confounders?
One can try to find an instrumental variable (IV), which is often used when covariates (observed features) are correlated with error terms (unobserved information).

Other causal inference methods: instrumental variable

What if there are unobserved confounders?
One can try to find an instrumental variable (IV), which is often used when covariates (observed features) are correlated with error terms (unobserved information).
Such correlation occurs when

- changes in the dependent variable change the value of at least one of the covariates ("reverse" causation)
- there are omitted variables that affect both the outcome and covariates (unobserved confounders)

Other causal inference methods: instrumental variable

What if there are unobserved confounders?
One can try to find an instrumental variable (IV), which is often used when covariates (observed features) are correlated with error terms (unobserved information).
Such correlation occurs when

- changes in the dependent variable change the value of at least one of the covariates ("reverse" causation)
- there are omitted variables that affect both the outcome and covariates (unobserved confounders)

Correlation between covariates and error terms is also known as *endogeneity*. In this situation, ordinary least squares produces biased and inconsistent estimates

Instrumental variable (IV)

An IV

- must be correlated with the endogenous explanatory variables, conditionally on the other covariates
- cannot be correlated with the error term in the explanatory equation, conditionally on the other covariates.

Instrumental variable (IV)

An IV

- must be correlated with the endogenous explanatory variables, conditionally on the other covariates
- cannot be correlated with the error term in the explanatory equation, conditionally on the other covariates.

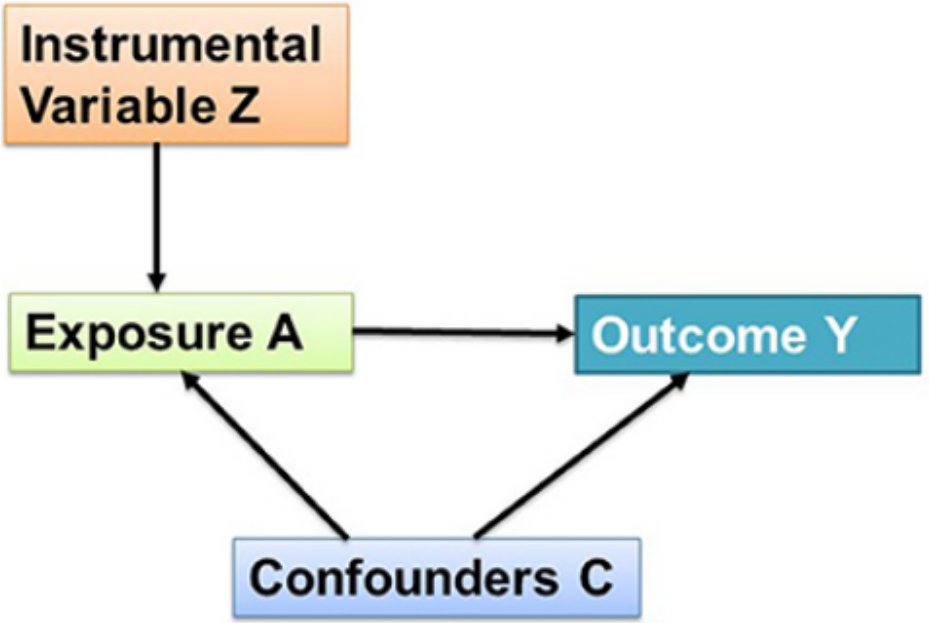
In attempting to estimate the causal effect of an exposure A on outcome Y, an IV is a third variable Z which affects Y only through its effect on A.

Instrumental variable (IV)

An IV

- must be correlated with the endogenous explanatory variables, conditionally on the other covariates
- cannot be correlated with the error term in the explanatory equation, conditionally on the other covariates.

In attempting to estimate the causal effect of an exposure A on outcome Y, an IV is a third variable Z which affects Y only through its effect on A.



Instrumental variable: example

- Research goal: estimate the causal effect of smoking on general health

Instrumental variable: example

- Research goal: estimate the causal effect of smoking on general health
- Problems: correlation does not imply causation
  - Unobserved counfounders: other variables, such as depression, may affect both health and smoking
  - Reverse causation: health may affect smoking

Instrumental variable: example

- Research goal: estimate the causal effect of smoking on general health
- Problems: correlation does not imply causation
  - Unobserved counfounders: other variables, such as depression, may affect both health and smoking
  - Reverse causation: health may affect smoking
- Unethical to conduct randomized trials

Instrumental variable: example

- Research goal: estimate the causal effect of smoking on general health
- Problems: correlation does not imply causation
  - Unobserved counfounders: other variables, such as depression, may affect both health and smoking
  - Reverse causation: health may affect smoking
- Unethical to conduct randomized trials
- Expensive to collect matched pairs of data

Instrumental variable: example

- Research goal: estimate the causal effect of smoking on general health
- Problems: correlation does not imply causation
  - Unobserved counfounders: other variables, such as depression, may affect both health and smoking
  - Reverse causation: health may affect smoking
- Unethical to conduct randomized trials
- Expensive to collect matched pairs of data
- Try to find an IV whose effect on health is mediated through smoking

Instrumental variable: example

- Research goal: estimate the causal effect of smoking on general health
- Problems: correlation does not imply causation
  - Unobserved counfounders: other variables, such as depression, may affect both health and smoking
  - Reverse causation: health may affect smoking
- Unethical to conduct randomized trials
- Expensive to collect matched pairs of data
- Try to find an IV whose effect on health is mediated through smoking $\rightarrow$ the tax rate for tobacco

Instrumental variable: example

- Depression $\stackrel{?}{\rightarrow}$ smoking

Instrumental variable: example

- Depression $\stackrel{?}{\rightarrow}$ smoking
  Potential confounders: some genes make people depress and smoke
  Reverse causation is also possible

Instrumental variable: example

- Depression $\stackrel{?}{\rightarrow}$ smoking
  Potential confounders: some genes make people depress and smoke
  Reverse causation is also possible
  IV?

Instrumental variable: example

- Depression $\stackrel{?}{\rightarrow}$ smoking
  Potential confounders: some genes make people depress and smoke
  Reverse causation is also possible
  IV?
  lack of job $\rightarrow$ depression $\stackrel{?}{\rightarrow}$ smoking

Instrumental variable: example

Good IVs are not always available

- Years of education $\stackrel{?}{\rightarrow}$ income

Instrumental variable: example

Good IVs are not always available

- Years of education $\stackrel{?}{\rightarrow}$ income
  This is a long-standing question in labor economics
  Confounders: ability

Instrumental variable: example

Good IVs are not always available

- Years of education $\stackrel{?}{\rightarrow}$ income
  This is a long-standing question in labor economics
  Confounders: ability
  IV?

Instrumental variable: example

Good IVs are not always available

- Years of education $\stackrel{?}{\rightarrow}$ income
  This is a long-standing question in labor economics
  Confounders: ability
  IV?
  Distance from home to a college/university $\rightarrow$ Years of education $\stackrel{?}{\rightarrow}$ income

Instrumental variable: example

Good IVs are not always available

- Years of education $\stackrel{?}{\rightarrow}$ income
  This is a long-standing question in labor economics
  Confounders: ability
  IV?
  Distance from home to a college/university $\rightarrow$ Years of education $\stackrel{?}{\rightarrow}$ income
  Any problem with the IV?

Instrumental variable: treatment effect estimates

Given IVs $\boldsymbol{Z}$,

- Stage 1: Regress each column of $\boldsymbol{X}$ on $\boldsymbol{Z}$; predict $\boldsymbol{X}$ by $\hat{\boldsymbol{X}}$ given $\boldsymbol{Z}$
- Stage 2: Regress $\boldsymbol{Y}$ on $\hat{\boldsymbol{X}}$
- The regression coefficients from stage 2 are the estimated effects

This method is only valid in linear models