



SAPIENZA
UNIVERSITÀ DI ROMA

SEMINARI INFERENZA CAUSALE

Dipartimento di Sanità Pubblica e Malattie Infettive
Aula Scuola di Specializzazione in Statistica Sanitaria e Biometria
AULA C1 ore 12:00 - 14:00

Martedì 29 Novembre: Introduzione alla causalità

Dr.ssa MARGHERITA MORETTI

Dottorato in Scuola di Scienze Statistiche
Sapienza Università di Roma

Dr.ssa LAURA MONTELISCIANI

Dottorato in Sanità Pubblica Biostatistica ed Epidemiologia
Università degli Studi di Milano-Bicocca

Martedì 6 Dicembre: Metodi per l'inferenza causale

Metodi per l'inferenza causale

Margherita Moretti

margherita.moretti@uniroma1.it

Laura Montelisciani

l.montelisciani@campus.unimib.it

PER SCARICARE LE SLIDE:

https://github.com/MMargherita/seminari_DSPMI



Nei **random trials** la scelta su quali unità trattare avviene tramite **randomizzazione**. Ciò permette che i due gruppi di **trattati** e **non trattati** siano **omogenei** per tutte le caratteristiche (misurabili e non).

→ La **differenza nell'outcome** che osserviamo **non dipende dalle differenze nelle caratteristiche** dei due gruppi, ma possiamo inferire che sia un **effetto causale del trattamento**



Negli studi **osservazionali** si osserva, per ogni unità, la **presenza** o l'**assenza** del trattamento. Dato che il trattamento **non è assegnato in maniera casuale**, i due gruppi di trattati e non trattati potrebbero essere **sistematicamente diversi** rispetto ad alcune caratteristiche (osservate e non).

Propensity Score

Hp: siamo in grado di osservare tutti i possibili confondenti della relazione fra trattamento T e outcome Y

Propensity score

E' un numero che indica la **probabilità**, per ogni individuo, **di essere trattato**, sulla base delle **caratteristiche di base (pre-trattamento)**

$$\text{Pr} (T = 1 \mid X)$$

In caso di **trattamento binario** (trattato o non trattato), il **propensity score** può essere stimato tramite una **semplice regressione** in cui abbiamo:

- Risposta binaria** $T=1$, $T=0$

- Un serie di **covariate** X che rappresentano tutti i possibili **confondenti** della relazione fra T e Y

I **valori predetti dal modello** rappresentano la **probabilità**, per ogni individuo, **di essere trattato**.

E' una **sintesi delle caratteristiche di base** che possono differire fra chi riceve effettivamente il **trattamento** e **chi non lo riceve**

Il propensity score che abbiamo ottenuto, per ogni unità, ha una proprietà di bilanciamento:

Condizionatamente al propensity score, il gruppo dei trattati e quello dei non trattati hanno la stessa distribuzione dei confondenti!

Condizionatamente al propensity score:

- L'assegnazione al trattamento è indipendente dai confondenti
- Il potential outcome è indipendente dall'assegnazione al trattamento

Condizionatamente al Propensity Score $PS(x)$:

- L'assegnazione al trattamento è indipendente dai confondenti
- Il potential outcome è indipendente dall'assegnazione al trattamento

$$y_1, y_0 \perp x \quad \text{allora} \quad y_1, y_0 \perp PS(x)$$

$$E(Y \mid x, T=1) = E(Y_1 \mid x) = E(Y_1 \mid PS(x))$$

$$E(Y \mid x, T=0) = E(Y_0 \mid x) = E(Y_0 \mid PS(x))$$

Diversi metodi basati sul Propensity Score (PS)

- Weighting (IPTW)**: calcolo del peso in funzione del PS, utilizzato per bilanciare trattati e non trattati e renderli il più possibile simili
- Matching (PSM)**: utilizzare il PS come metrica della distanza tra gli individui, più le unità sono vicine in termini di distanza, più sono simili in termini di confondenti
- Stratificazione/subclassification**: utilizza il PS per classificare le unità

Inverse Probability of Treatment Weighting (IPTW)

Idea simile a quella dei pesi campionari:

I **pesi campionari**, utilizzati nelle indagini statistiche, hanno l'obiettivo di **correggere le differenze**, in termini di caratteristiche osservate, che ci sono fra **campione intervistato** e la **popolazione di riferimento**.

Infatti, i pesi di campionamento vengono utilizzati per **correggere il sovra(sotto)campionamento** di unità con caratteristiche specifiche ed arrivare ad un **campione che riflette la popolazione di interesse**

Utilizzando i pesi campionari è possibile generalizzare all'intera popolazione i risultati di ricerca osservati nel campione

Allo stesso modo, i propensity score:

Se utilizzati come **pesi**, permettono al nostro **campione in studio** di **riprodurre** una “**pseudo-popolazione**” composta di **trattati e non trattati** che **non differiscono nelle distribuzioni dei confondenti**

Proprietà della pseudo-popolazione:

- 1) Trattati e non trattati sono statisticamente indipendenti
- 2) Il mean outcome è uguale alla media standardizzata calcolata sulla popolazione reale

Infatti, l'**IPTW** è una **metodologia di standardizzazione!**

IPTW come metodo di standardizzazione

Utilizzando i PS come pesi, possiamo **eliminare il bias di selezione!**

I pesi **eliminano le differenze di composizione** nei gruppi di trattati e non trattati, **modificando la popolazione campionaria**

Ogni individuo contribuisce alle stime in misura proporzionale all'assegnazione W

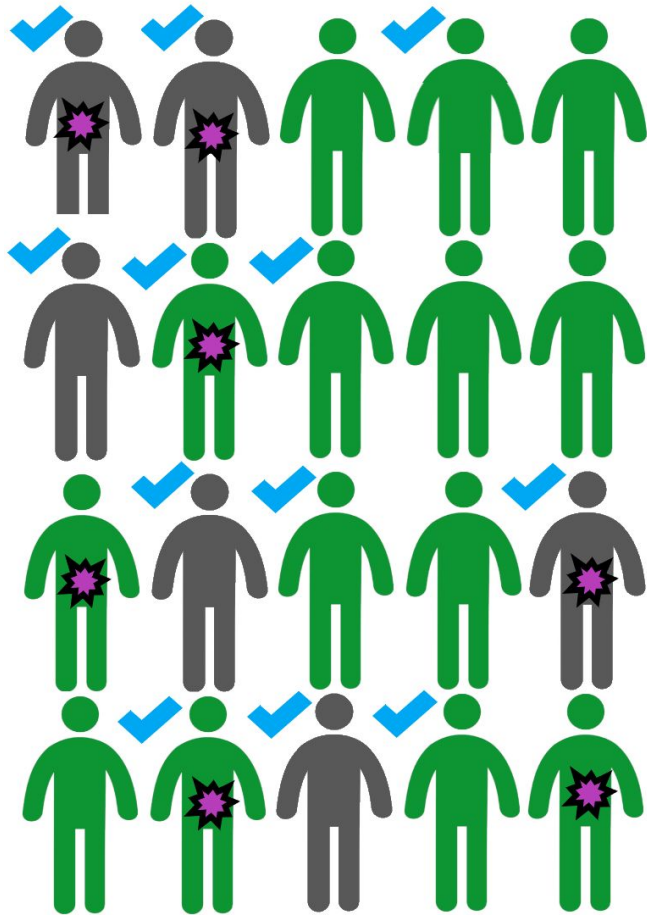
Se vogliamo stimare l'ATE:

dove, la prima parte è per i trattati ($T_i=1$) e la seconda ai non trattati ($T_i=0$)

Se vogliamo stimare l'ATT:

Steps in performing a causal analysis using data preprocessing (Stuart, 2010):

1. Decide on covariates to balance (confounders, mostly)
2. Estimate the distance measure (e.g., propensity score)
3. Condition on the distance measure (e.g., weighting, or stratification, matching)
4. Assess covariates balance; unsatisfactory, repeat steps 2-4 (e.g. adding interaction or consider only variables with common support)
5. Estimate the treatment effect conditioning the sample on the distance measure chosen



Propensity score = $P(T=1 | X)$

$P(\text{essere in studio} | \text{verde}) = 6/14 = 0.43$

$P(\text{essere in studio} | \text{grigio}) = 6/6 = 1$

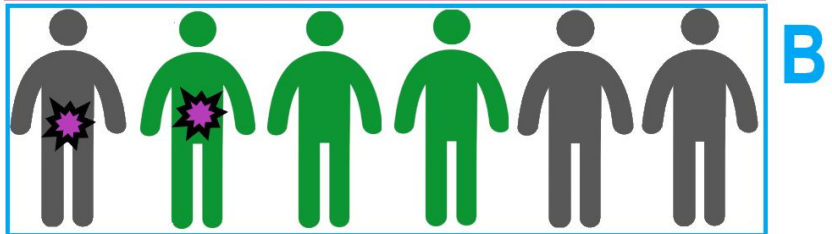
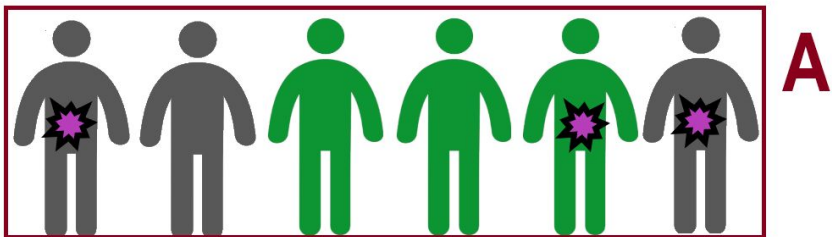
$W_i = 1 / \text{propensity score}$

$W_{\text{verde}} = 1 / 0.43 = 2.33$

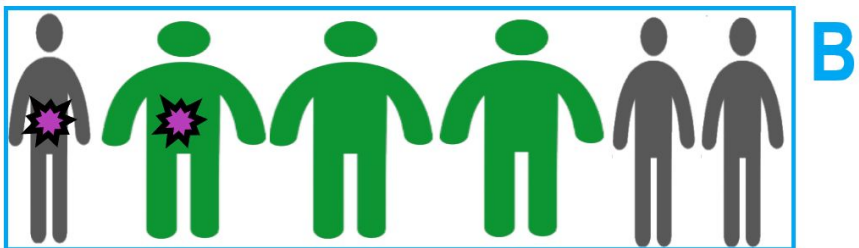


$W_{\text{grigio}} = 1 / 1 = 1$





$$\begin{aligned} \text{Rischio}_A &= 3/6 = 50 \% \\ \text{Rischio}_B &= 2/6 = 33.3 \% \\ \text{Risk Difference} &= 16.2 \% \\ \text{Risk Ratio} &= 0.50/0.333 = 1.5 \end{aligned}$$



$$\begin{aligned} \text{Rischio}_A &= (2 \times 1 + 2.33)/6 = 72.17 \% \\ \text{Rischio}_B &= (1 + 2.33)/6 = 55.5 \% \\ \text{Risk Difference} &= 16.67 \% \\ \text{Risk Ratio} &= 1.3 \end{aligned}$$

Problems (and remedies) with weights:

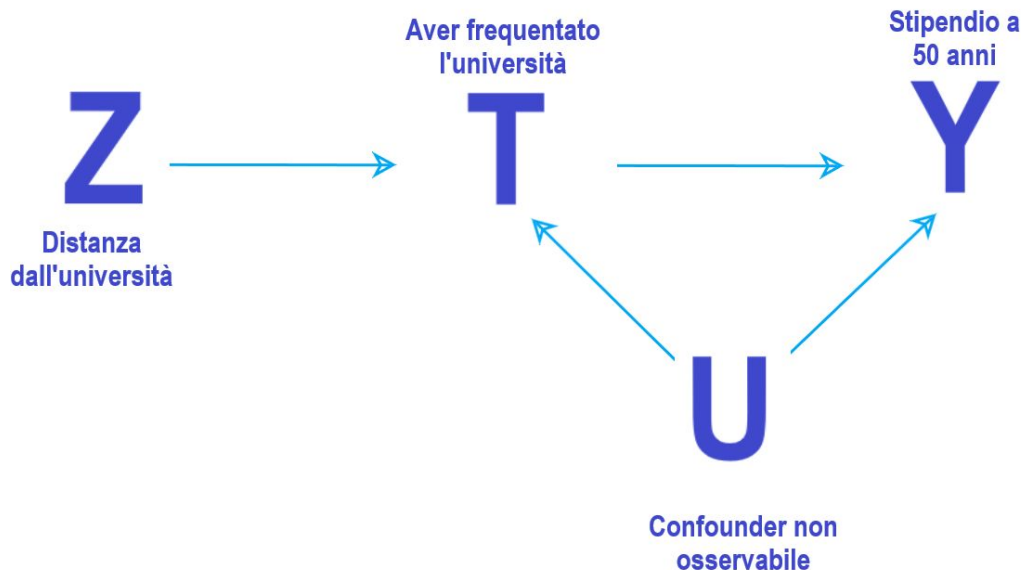
Extreme weights:

1. Truncate weights (starting 99th percentile)
2. Stabilized Weights (obtained multiplying by a constant treated and controls units by the constants representing the number of units among T and C)
3. Try a different PS model
4. Use a different weighting approach (e.g. MMW)

Instrumental Variable

Instrumental Variable

Le **Variabili Strumentali** (IV) permettono di aggiustare per tutti i confondenti, osservabili e non, presenti tra il trattamento e l'outcome.



Un buon strumento rispetto 3 importanti condizioni:

1. La variabile strumentale Z ha un effetto causale diretto sul trattamento T;
2. La variabile strumentale Z ha un effetto sull'outcome Y solo attraverso D, ossia non c'è effetto causale diretto da Z a Y;
3. La variabile strumentale Z non condivide cause comuni con l'outcome Y.

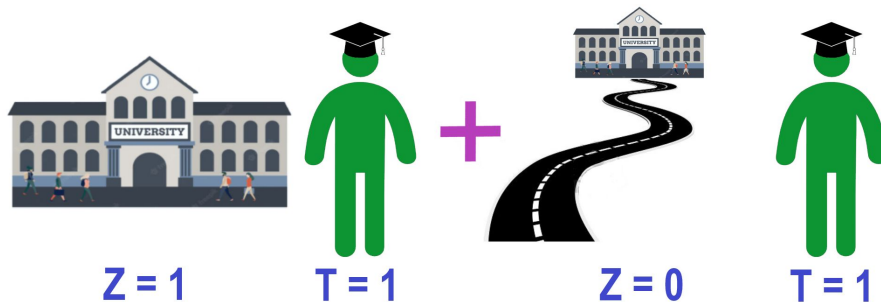
Instrumental Variable

Siano T e Z dicotomiche e sia T_z il potential outcome di T rispetto a Z , allora si generano 4 sottogruppi della popolazione:

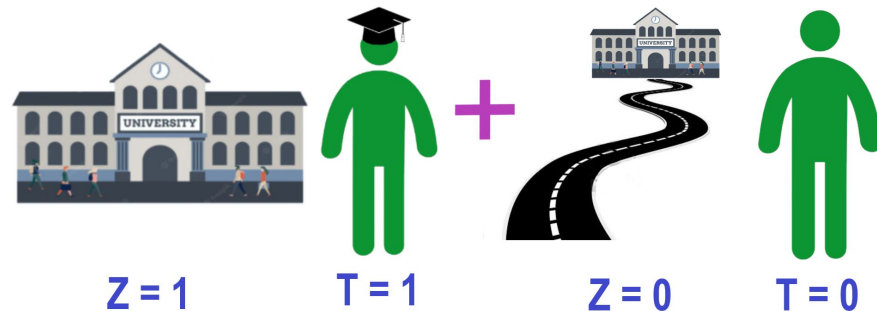
- a) **always takers** $\rightarrow T_0 = 1 \ \& \ T_1 = 1$, ossia qualunque sia il valore che assume lo strumento le unità saranno comunque trattate.
- b) **never takers** $\rightarrow T_0 = 0 \ \& \ T_1 = 0$, ossia qualunque sia il valore che assume lo strumento le unità non saranno trattate.
- c) **compliers** $\rightarrow T_0 = 0 \ \& \ T_1 = 1$, ossia se ricevono lo strumento allora ricevono anche il trattamento, oppure, se non ricevono il trattamento non ricevono neanche il trattamento.
- d) **defiers** $\rightarrow T_0 = 1 \ \& \ T_1 = 0$, ossia se ricevono lo strumento allora non ricevono anche il trattamento, oppure, se non ricevono lo strumento ricevono il trattamento.

Instrumental Variable

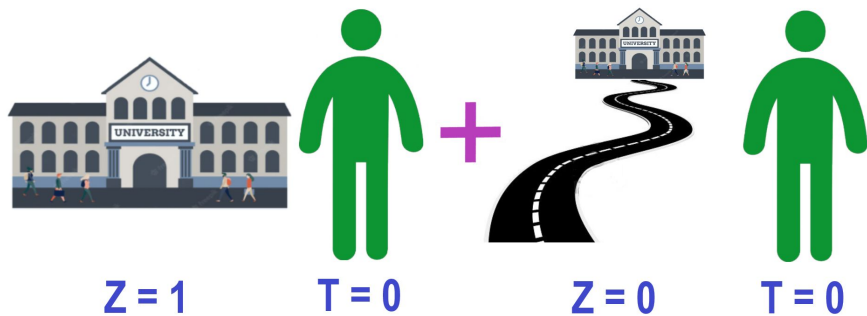
always takers



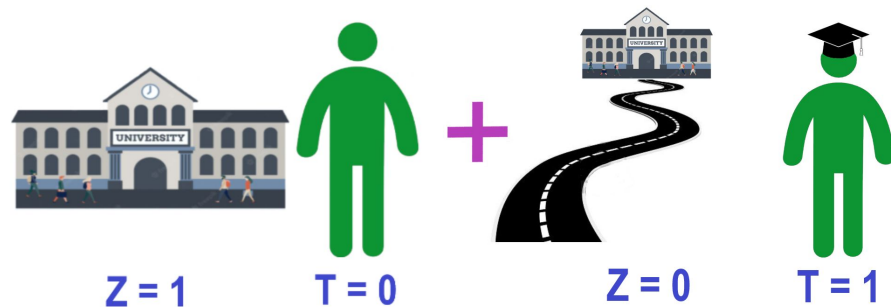
compliers



never takers



defiers



Instrumental Variable: LATE

Ragioniamo sull'effetto causale di ogni strato:

- **always takers**
 - **never takers**
- } Non è possibile stimare alcun effetto causale perché il trattamento è insensibile allo strumento;
- **defiers**: assumo che non esistono perché hanno un comportamento controintuitivo;
 - **compliers**: posso stimare il **Local Average Treatment Effect (LATE)**.



I potential treatment T_z non sono osservabili, ad esempio se osservo $T_1 = 1$ non so distinguere se si tratta di un always takers o di un compliers.

Instrumental Variable: LATE

1. **Monotonicity** (non esistono i defiers):

$$P(T_0 > T_1) = 0 \text{ ossia } P(T_0 = 1 \ \& \ T_1 = 0) = 0$$

2. **Esistenza dei compliers:**

$$P(T_0 < T_1) > 0 \text{ ossia } P(T_0 = 0 \ \& \ T_1 = 1) > 0$$

3. **Uncounfunded Instrument:** (T_0, T_1) La distribuzione dello strumento Z negli strati del Trattamento T è sempre la stessa.

$$P(Z = z \mid T_0 = t_0, T_1 = t_1) = P(Z = z)$$

4. **Mean exclusion restrict:** In media i potential outcome, se condizionati allo strato g , non dipendono dallo strumento Z .

$$P(Y_T \mid Z=0, G=g) = P(Y_T \mid Z=1, G=g)$$

Se le assunzioni sono plausibili si può stimare il LATE utilizzando delle quantità osservabili:

Effetto dello strumento Z sull'outcome Y

$$LATE = E[Y_1 - Y_0 \mid G=c] = \frac{E[Y \mid Z=1] - E[Y \mid Z=0]}{E[T \mid Z=1] - E[T \mid Z=0]}$$

Effetto dello strumento Z sul trattamento T

Consideriamo un trattamento continuo T:

- $E[Y | Z=1] - E[Y | Z=0]$ Effetto dello strumento Z sull'outcome Y

$$\rightarrow Y = \mu_Y + \beta_{YT}T + \varepsilon_Y$$

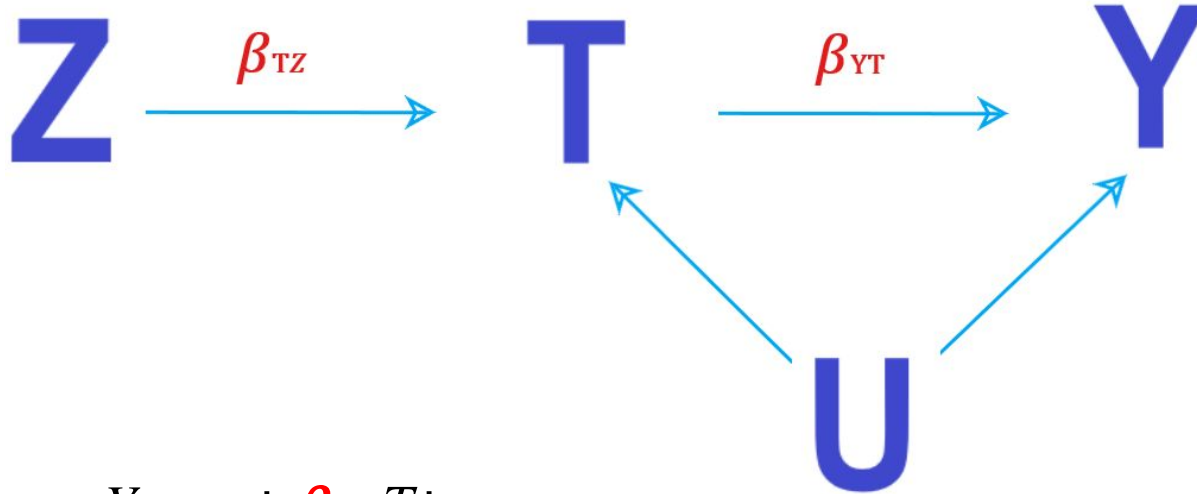
- $E[T | Z=1] - E[T | Z=0]$ Effetto dello strumento Z sul trattamento T

$$\rightarrow T = \mu_T + \beta_{TZ}Z + \varepsilon_T$$

La variabile **T** è **endogena**, ossia è spiegata da altre variabili nel modello, in questo caso dallo strumento Z. **Z** è **esogena** perché non è spiegata da alcuna variabile nel modello.



Stimare il LATE di T su Y $\rightarrow \beta_{YT}$



$$Y = \mu_Y + \beta_{YT} T + \varepsilon_Y$$

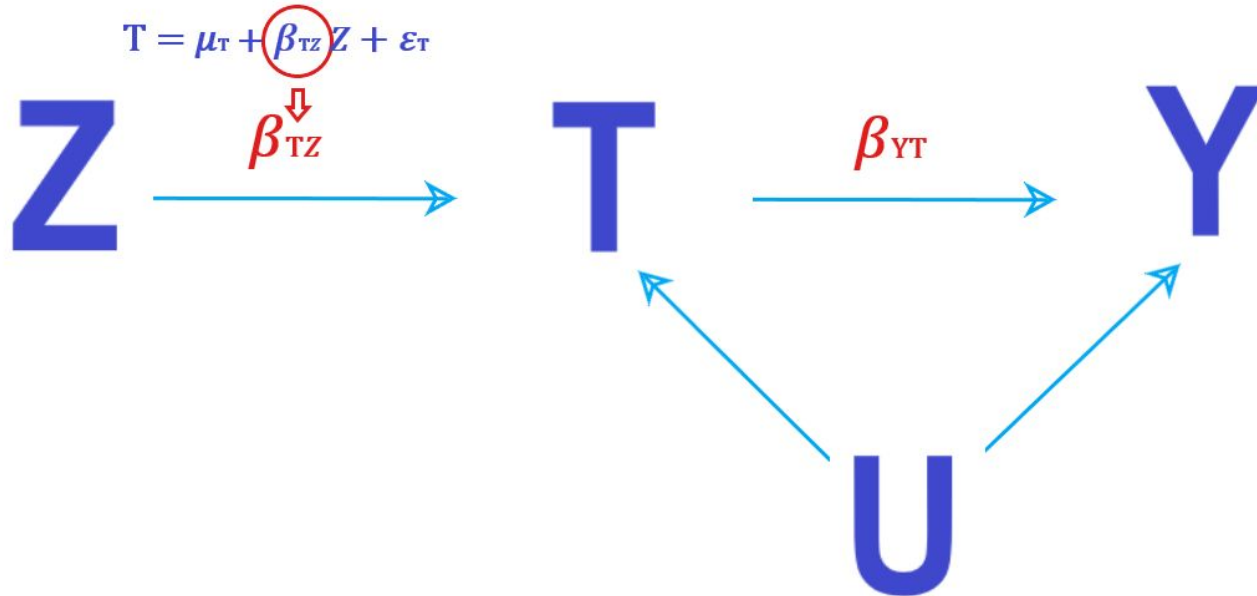
\rightarrow Distorto da U (impossibile aggiustare perché U è non osservabile)

Instrumental Variable: 2SLS



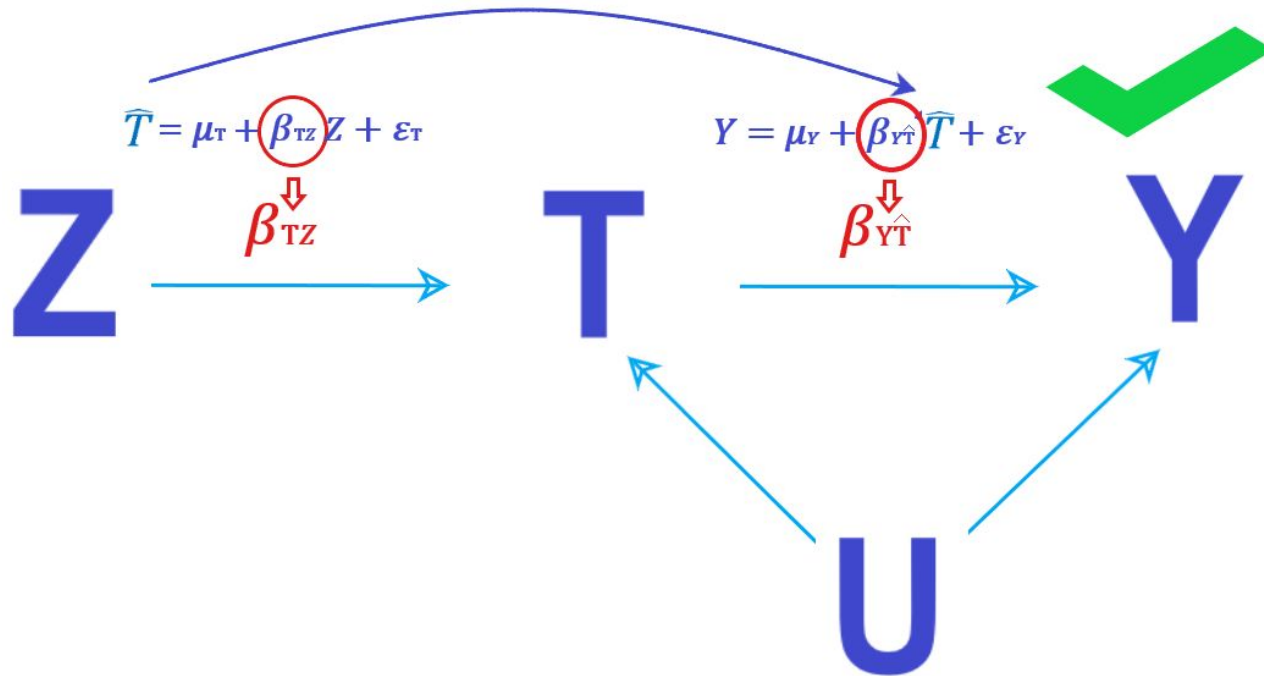
2 stage least square

Stage 1: Stimo β_{TZ} con una prima regressione di Z su T



Instrumental Variable: 2SLS

Stage 2: utilizzo $\widehat{\beta}_{TZ}$ stimato allo Stage 1 per stimare \widehat{T} (non distorto) che utilizzerò come variabile in una seconda regressione per stimare $\widehat{\beta}_{YT}$ (non distorto).



Instrumental Variable: 2SLS

→ Stima del LATE ottenuta col 2SLS:

effetto indiretto dello
strumento sull'outcome:

$$\beta_{TZ} * \beta_{YT}$$


$$\beta_{YT}$$

2° stage (utilizzando la
stima di β_{TZ} del 1° stage)

$$LATE = E[Y_1 - Y_0 \mid G=c] = \frac{E[Y \mid Z=1] - E[Y \mid Z=0]}{E[T \mid Z=1] - E[T \mid Z=0]} = \frac{\beta_{YZ}}{\beta_{TZ}}$$

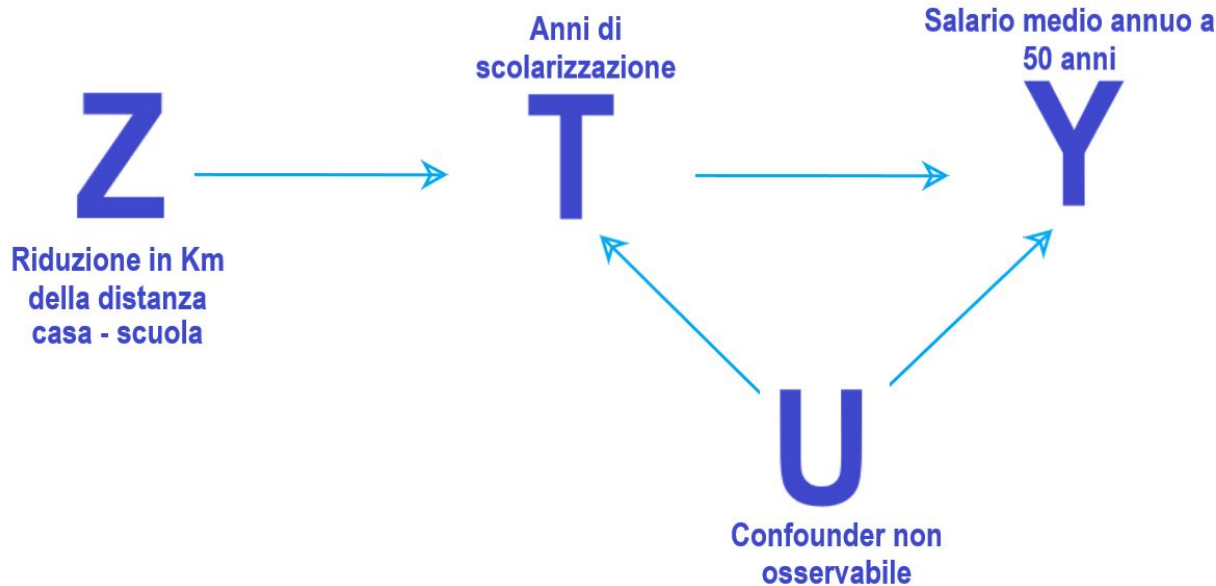
1° stage

Instrumental Variable: 2SLS

→ **Stima del LATE ottenuta col 2SLS:** Esempio



Vogliamo stimare l'effetto di un anno in più di scuola sul salario medio annuo percepito a 50 anni, utilizzando come strumento la vicinanza in Km dalla scuola.



Instrumental Variable: 2SLS

1° stage: $T = \mu_T + \beta_{TZ}Z + \varepsilon_T \rightarrow \beta_{TZ} = 0.2$

Un aumento unitario in Z, ossia di un Km di vicinanza, provoca un aumento di 0.2 di T, ossia anni di scolarizzazione (circa 2 mesi e mezzo).

2° stage: $Y = \mu_Y + \beta_{YT}T + \varepsilon_Y$

$$Y = \mu_Y + \beta_{YT}(\mu_T + \beta_{TZ} * Z) + \varepsilon_Y \rightarrow \beta_{YT} = 2500$$

*Un aumento unitario di T, ossia di un anno la scolarizzazione, provoca un aumento del reddito di 2500\$. Per far aumentare T di uno Z dovrebbe aumentare 5 volte, ossia dovrei ridurre la strada tra la casa e la scuola di 5 Km, ed avrei quindi l'aumento del reddito di $500\$ * 5 = 2500\$$.*

Effetto indiretto dello strumento sull'outcome: $\beta_{TZ} * \beta_{YT} = 2500 * 0.2 = 500$

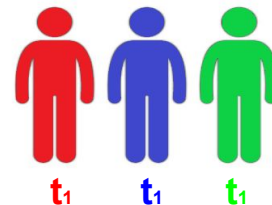
Un aumento unitario di Z, ossia di un Km di vicinanza, provoca un aumento del salario medio annuo di 500 dollari.

Difference-in-difference

Difference-in-difference

Fino ad ora abbiamo parlato di ...

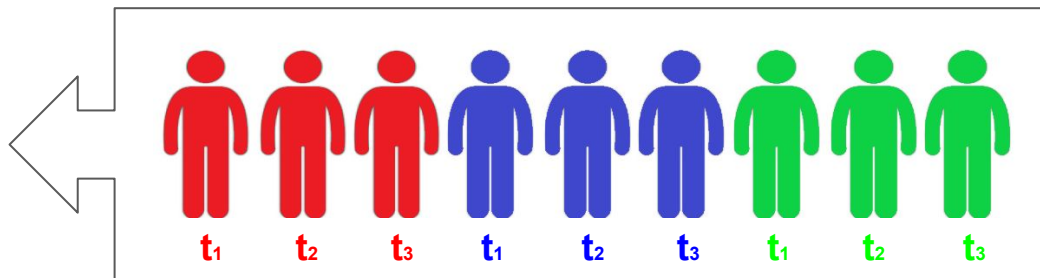
Cross sectional: l'outcome e le covariate sono state osservate nello stesso istante di tempo.



Da ora in poi parleremo di ...

Panel data o dati longitudinali: ciascun individuo viene osservato in diversi istanti di tempo, quindi non avremo più solo una rilevazione dell'outcome e delle covariate ad individuo, ma diverse rilevazioni dell'outcome e delle covariate per ciascun individuo.

E' come se avessi 3 cross sectional, posso stimare la differenza tra gli individui in ogni istante, ma posso anche vedere se il fenomeno ha un certo andamento nel tempo.



Difference-in-difference

Cross sectional:



Quando stimo l'effetto di una covariata o di un trattamento sull'outcome c'è il problema delle **variabili omesse**, ossia tutte quelle variabili o caratteristiche intrinseche delle unità che non sono nel modello.

Panel data o dati longitudinali



Quando stimo l'effetto di una covariata o di un trattamento sull'outcome posso stimare anche l'effetto delle **variabili omesse**. Ogni soggetto ha una serie di misurazioni che permettono di stimare un livello di partenza individuale dell'outcome (*random intercept*).

Difference-in-difference

Cross sectional

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \forall \text{ unità } i$$

β_0 è il valore atteso di Y per la popolazione generale quando $X = 0$. Ma β_0 può anche essere interpretato come **l'effetto delle variabili omesse che non variano con i**, ossia non variano da soggetto a soggetto.

Panel data o dati longitudinali

$$Y_{it} = \beta_{0i} + \beta_{1i} X_{it} + \varepsilon_{it} \quad \forall \text{ unità } i \quad \forall \text{ tempo } t$$

β_{0i} è il valore atteso di Y per l'unità i quando $X = 0$. Posso stimare **l'effetto delle variabili omesse che variano con i**, perché ho più osservazioni per ogni unità.

Difference-in-difference

Cross sectional: Sia Y il costo della casa ed X il numero di servizi di lusso di cui dispone (portineria, ascensore, vista, videosorveglianza, pannelli solari, ecc).

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \forall \text{ unità } i$$

β_0 costo medio di una casa senza servizi di lusso;

β_1 incremento di costo medio all'aumentare unitario del numero di servizi di lusso;

Panel data o dati longitudinali: ipotizziamo che ci sia una sola variabile omessa, se la casa è collocata al centro città o se è in periferia.

$$Y_{it} = \beta_{0i} + \beta_{1i} X_{it} + \varepsilon_{it} \quad \forall \text{ unità } i \quad \forall \text{ tempo } t$$

Avrò due β_{0i} ossia due livelli di partenza possibili a seconda che la casa sia dentro o fuori dal centro: mi aspetto che, a parità di servizi, la casa in centro sia più cara;



Con la stessa covariata riesco a stimare anche l'effetto della variabile omessa

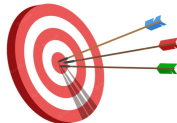
Difference-in-difference: ATE

Supponiamo di avere due istanti di tempo: t e $t-1$. Al tempo $t-1$ nessuno è stato trattato, al tempo t alcuni riceveranno il trattamento mentre altri no.



Stimare L'**ATE** di T su Y : $E[Y_{1,t} - Y_{0,t}]$

Ma possiamo osservare l'effetto causale del trattamento solo nei trattati, dunque ...



Stimare L'**ATET** di T su Y : $E[Y_{1,t} - Y_{0,t} | T=1]$

Ma l'effetto che vedo al tempo t , potrebbe essere costituito dall'effetto causale di T ma anche dall'effetto temporale!



Stimare L'**ATET** di T su Y **al netto dell'effetto del tempo** utilizzando l'effetto temporale osservato nei non trattati

Difference-in-difference: ATET

$$\text{ATE} = E[Y_{1,t} - Y_{0,t}] = \boxed{E[Y_{1,t} - Y_{0,t-1}]} - \boxed{E[Y_{0,t} - Y_{0,t-1}]}$$

↓

differenza
dell'outcome tra
t-1 e t nei trattati

\tilde{Y}_1

↓

differenza
dell'outcome tra t-1
e t nei non trattati

\tilde{Y}_0

Se $\tilde{Y}_1, \tilde{Y}_0 \perp T$

$$\text{ATE} = E[Y_{1,t} - Y_{0,t-1} \mid T=1] - E[Y_{0,t} - Y_{0,t-1} \mid T=0] =$$

$$E[Y_t - Y_{t-1} \mid T=1] - E[Y_t - Y_{t-1} \mid T=0] \rightarrow \text{quantità osservabili}$$

Difference-in-difference

Se non si può assumere $\tilde{Y}_1, \tilde{Y}_0 \perp T$ allora utilizzo

l'ATET

$$ATET = E[Y_{1,t} - Y_{0,t} | T = 1] = \underbrace{E[Y_{1,t} | T=1]}_{\text{Osservato}} - \underbrace{E[Y_{0,t} | T = 1]}_{\substack{\text{Non osservato} \\ \text{Da stimare}}}$$

Per stimare $E[Y_{0,t} | T = 1]$ devo fare delle assunzioni:

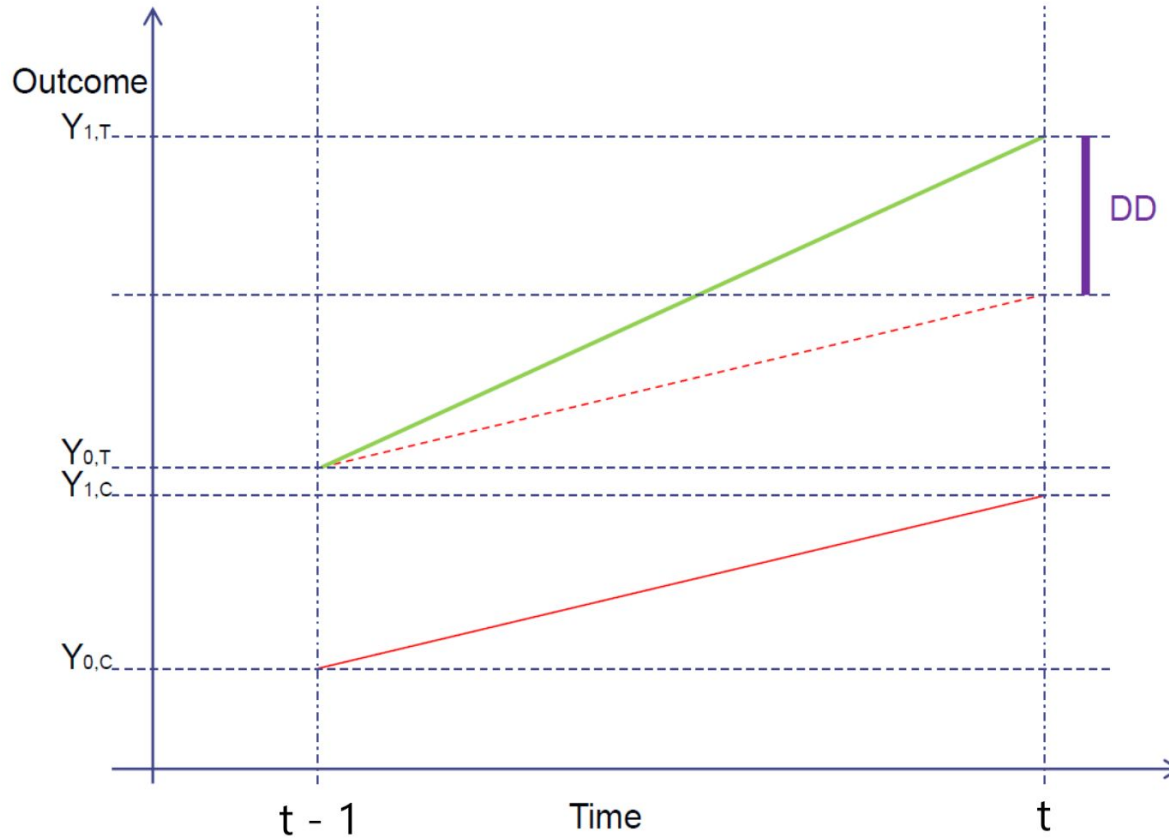
- **Common Trend o Parallel trend:** L'evoluzione del fenomeno nel tempo, senza essere intaccato dal trattamento, è lo stesso nei due gruppi.

$$E[Y_{0,t} - Y_{0,t-1} | T = 1] = E[Y_{0,t} - Y_{0,t-1} | T = 0]$$

- **Bias stability:** Il Bias tra trattati e non trattati è presente ma è stabile nel tempo.

$$E[Y_{0,t} | T=1] - E[Y_{0,t} | T=0] = E[Y_{0,t-1} | T = 1] - E[Y_{0,t-1} | T = 0]$$

Difference-in-difference



Common Trend o Parallel trend: In assenza del trattamento, la differenza tra i trattati e i non trattati rimane la stessa sia al tempo $t-1$ ed che al tempo t

Difference-in-difference

Utilizzando l'assunzione di Common Trend ...

$$E[Y_{0,t} - Y_{0,t-1} | T = 1] = E[Y_{0,t} - Y_{0,t-1} | T = 0]$$

$$E[Y_{0,t} | T = 1] - E[Y_{0,t-1} | T = 1] = E[Y_{0,t} - Y_{0,t-1} | T = 0]$$

$$E[Y_{0,t} | T = 1] - E[Y_{0,t-1} | T = 1] + E[Y_{0,t-1} | T = 1] = E[Y_{0,t} - Y_{0,t-1} | T = 0] + E[Y_{0,t-1} | T = 1]$$

$$E[Y_{0,t} | T = 1] = E[Y_{0,t} - Y_{0,t-1} | T = 0] + E[Y_{0,t-1} | T = 1]$$

... riesco a stimare l'elemento non osservato dell'ATET attraverso delle misure osservabili:

$$ATET = E[Y_t - Y_{t-1} | T = 1] - E[Y_t - Y_{t-1} | T = 0]$$

Questa stima coincide con la stima del coefficiente di regressione della differenza tra gli

outcome $\rightarrow Y_{i,t} - Y_{i,t-1} = \beta_0 + \beta_D D_i + e_i$

Difference-in-difference

Se non si può assumere **Bias Stability/Common trend**

$$\mathbf{E}[\mathbf{Y}_{0,t} \mid \mathbf{T} = \mathbf{1}] = E[Y_{0,t} - Y_{0,t-1} \mid T = 0] + E[Y_{0,t-1} \mid T = 1] + \mathbf{B}_{t,t-1}$$

E' necessario stimare $B_{t,t-1}$: per farlo serviranno dei dati di un periodo precedente.

Ipotesi $\rightarrow \mathbf{B}_{t,t-1} = \mathbf{B}_{t-1,t-2}$

E stimo il bias del periodo precedente come segue

$$\mathbf{B}_{t-1,t-2} = E[Y_{0,t-1} - Y_{0,t-2} \mid T = 1] = E[Y_{0,t-1} - Y_{0,t-2} \mid T = 0]$$

In cui tutto è osservabile perché nessuno è ancora stato trattato, il trattamento sarà assegnato solo al tempo t .

Regression Discontinuity

Regression Discontinuity (sharp)

$$\text{se } Z > z_0 \Rightarrow T = 1$$

$$Y = Y_0 + T * (Y_1 - Y_0)$$

ATE



$$\text{se } Z < z_0 \rightarrow T = 0 \text{ e } Y = Y_0$$

$$\text{se } Z \geq z_0 \rightarrow T = 1 \text{ e } Y = Y_0 + (Y_1 - Y_0)$$

Consideriamo il valore atteso $E[Y | Z = z] = E[Y_0 + T * (Y_1 - Y_0) | Z = z]$

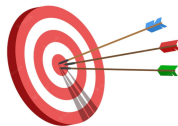
La funzione $E[Y | Z = z]$ è una funzione discontinua, con punto di discontinuità in z_0 .

Infatti:

$$\text{se } z < z_0 \text{ allora } E[Y | Z = z] = E[Y_0 | Z = z]$$

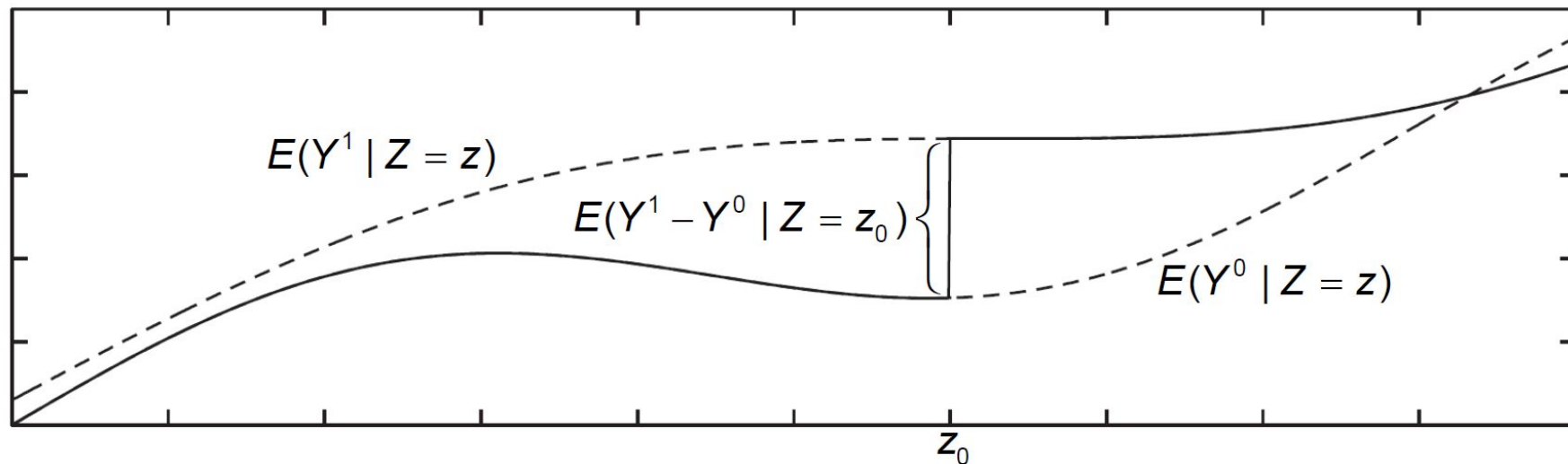
$$\text{se } z \geq z_0 \text{ allora } E[Y | Z = z] = E[Y_1 | Z = z]$$

Regression Discontinuity (sharp)



L'altezza del salto è l'**ATE condizionato a $Z=z_0$** , ed è ciò che vogliamo stimare.

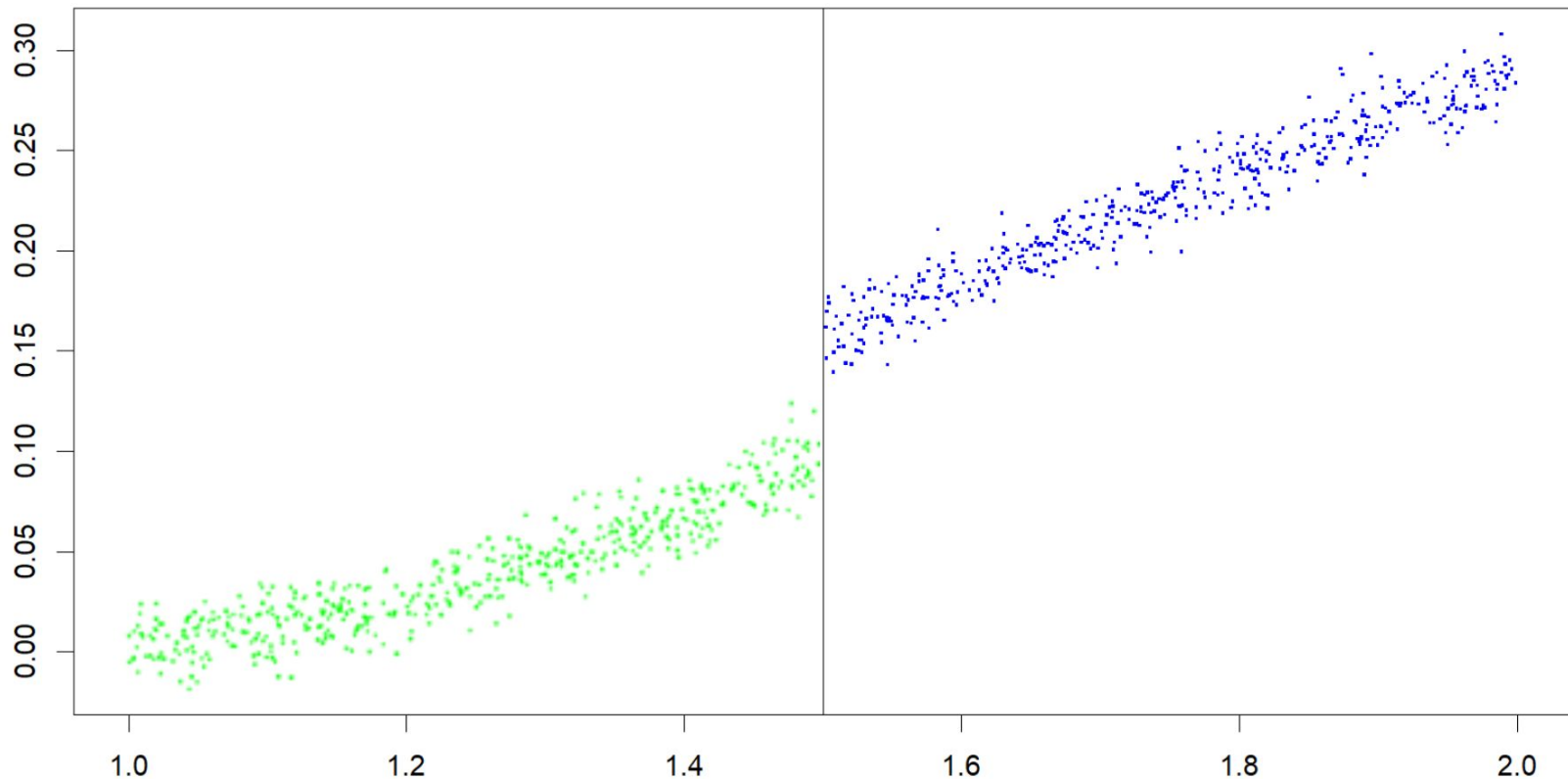
Assunzione: continuità di $E[Y_T | Z = z]$ in $Z=z_0$, per $T = 1, 0$.



— osservato
- - - non osservato

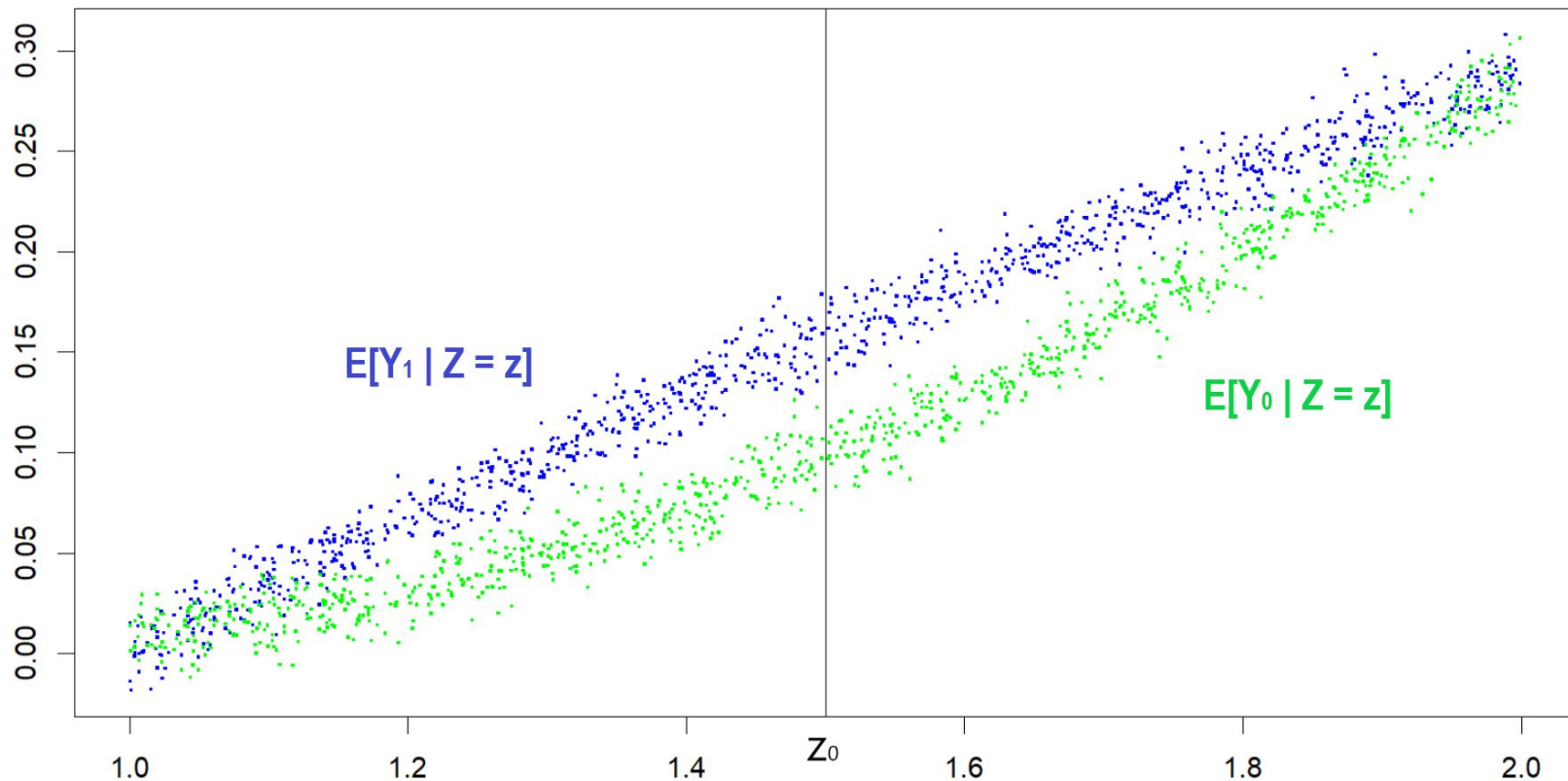
Regression Discontinuity (sharp)

Ciò che osserviamo ($z_0 = 1.5$)



Regression Discontinuity (sharp)

Ciò che vogliamo confrontare (due regressioni): è come se avessimo i dati completi per il verde ed il blu



Regression Discontinuity (sharp)

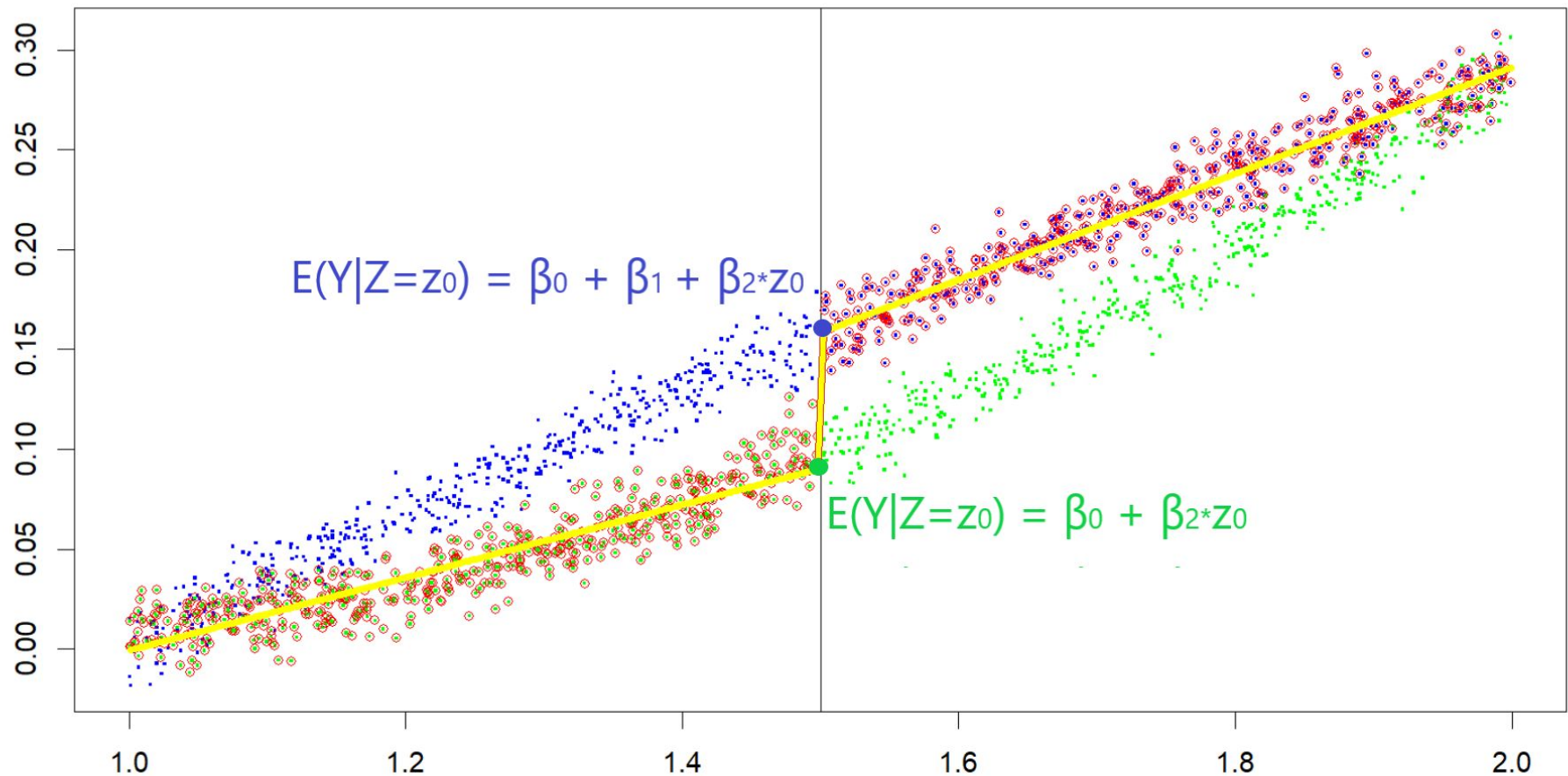
Per stimare l'ATE consideriamo la seguente equazione della retta di regressione per la funzione $E[Y|Z=z]$:

$$E[Y|Z=z] = \beta_0 + \beta_1 * T + \beta_2 * Z$$

Per $Z < z_0$ ossia con $T=0$ sarà: $E[Y_0|Z=z] = \beta_0 + \beta_2 * Z$

Per $Z > z_0$ ossia con $T=1$ sarà: $E[Y_1|Z=z] = \beta_0 + \beta_1 + \beta_2 * Z$

Regression Discontinuity (sharp)





Regression Discontinuity (sharp)

$$ATE = E(Y_1 - Y_0 | Z=z) = E(Y_1 | Z=z) - E(Y_0 | Z=z) =$$

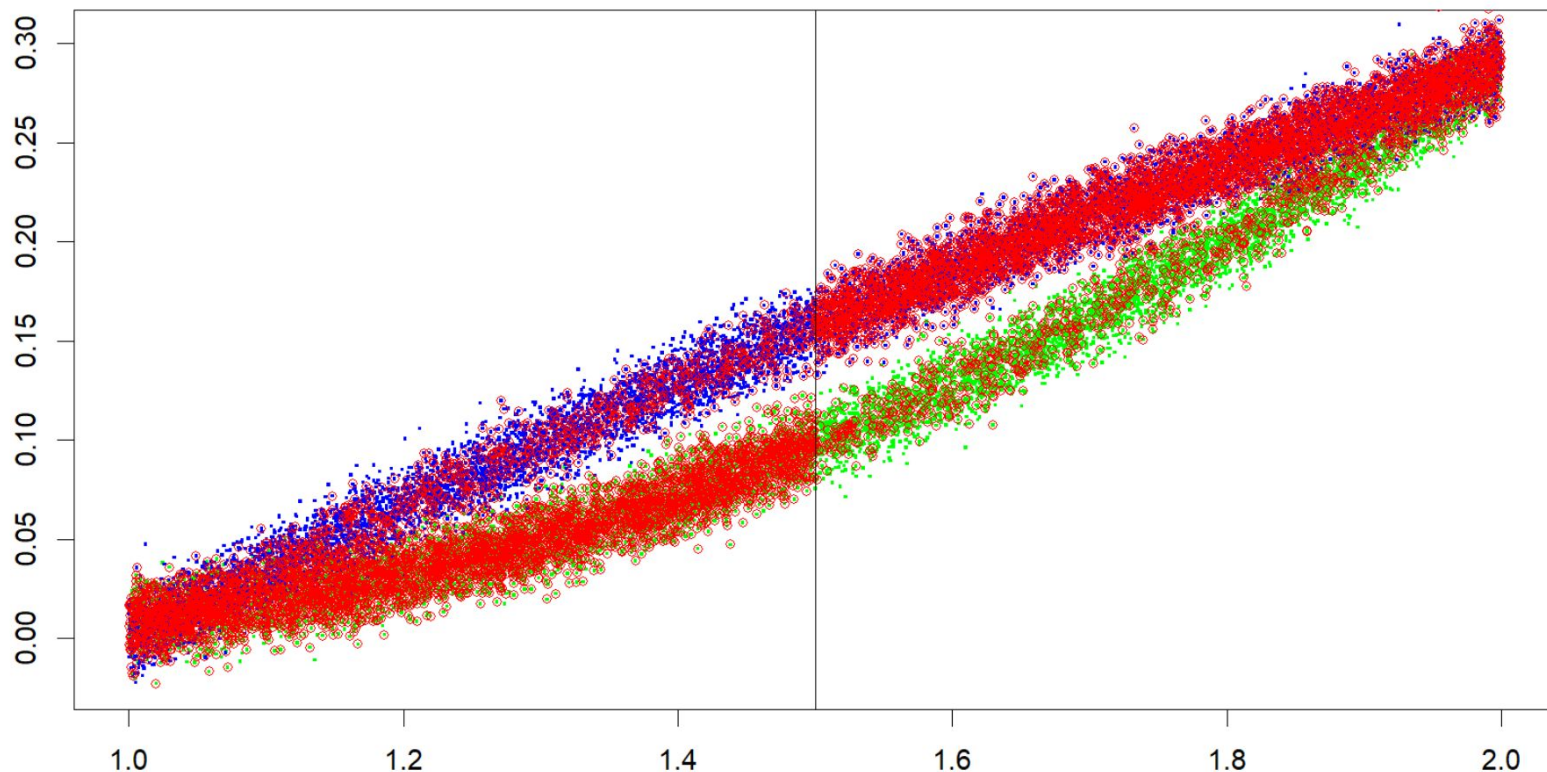
$$\cancel{\beta_0} + \beta_1 + \cancel{\beta_2 * z_0} - \cancel{\beta_0} - \cancel{\beta_2 * z_0} = \beta_1$$

Un altro modo di pensare all'ATE è come differenza del limite destro e del limite sinistro in z_0 della funzione $E[Y|Z=z]$ ossia:

$\lim_{\varepsilon \downarrow 0} E(Y Z = z_0 + \varepsilon)$	$\lim_{\varepsilon \downarrow 0} E(Y Z = z_0 - \varepsilon)$
	
$E(Y_1 Z=z)$	$E(Y_0 Z=z)$
parte destra $Z > z_0$ e $T = 1$	parte sinistra $Z < z_0$ e $T = 0$

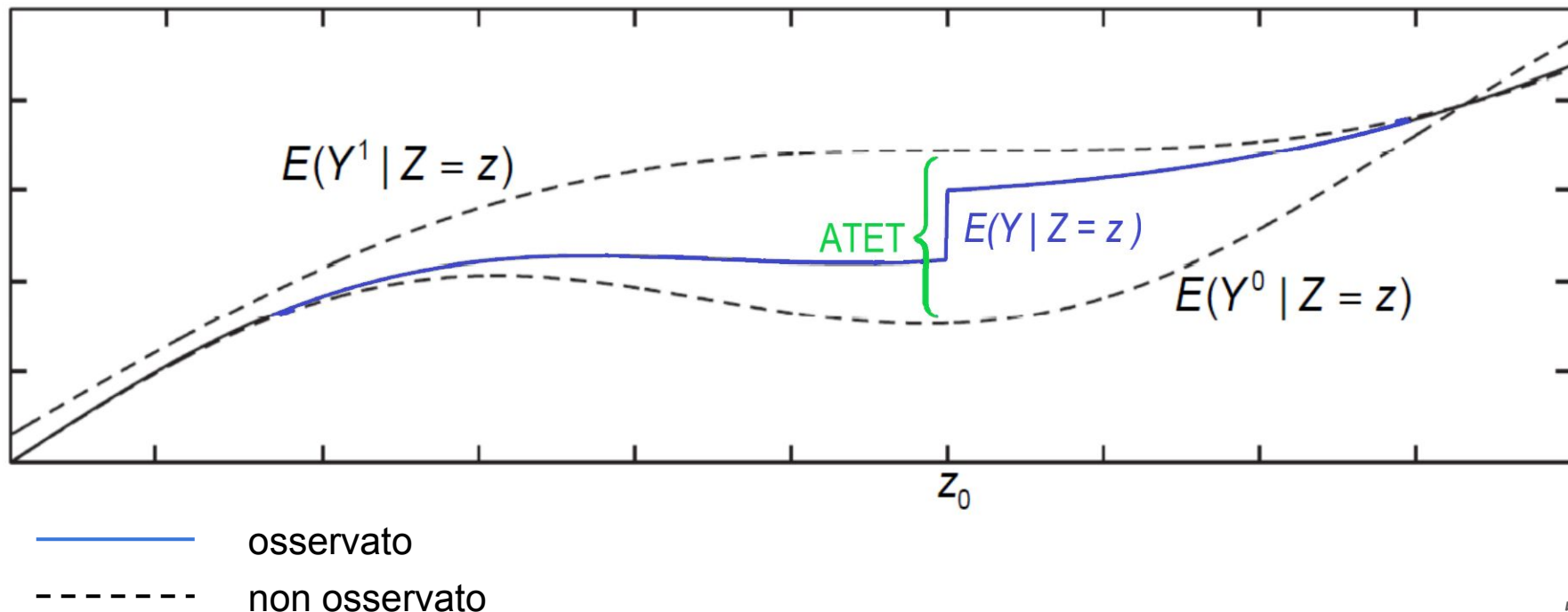
Regression Discontinuity (fuzzy)

Può capitare un altro caso in cui la soglia non definisce perfettamente l'assegnazione del trattamento, ma crea comunque una discontinuità nella probabilità di essere trattati.



Regression Discontinuity (fuzzy)

Possiamo fare un ragionamento affine all'instrumental variable: dato un valore $\varepsilon > 0$ piccolo a piacere, possiamo identificare quattro gruppi latenti distinti nella popolazione per $Z = z$ $[z_0 - \varepsilon, z_0 + \varepsilon]$: Always Takers, Never Takers, Compliers e Defiers.



Regression Discontinuity (fuzzy)

Per gli stessi motivi visti nell'IV considereremo solo i compliers, ossia solo a chi si presume che sarà trattato in base al superamento della soglia



Il **LATE** condizionato a $Z=z_0$ è ciò che vogliamo stimare.

$$\text{LATE} = E(Y^1 - Y^0 | Z = z_0, G = c) = \frac{\lim_{\varepsilon \downarrow 0} [E(Y | Z = z_0 + \varepsilon) - E(Y | Z = z_0 - \varepsilon)]}{\lim_{\varepsilon \downarrow 0} [E(T | Z = z_0 + \varepsilon) - E(T | Z = z_0 - \varepsilon)]}$$

- Il numeratore esprime l'LATE moltiplicato per la probabilità di essere compliers.

$$\lim_{\varepsilon \downarrow 0} [E(Y | Z = z_0 + \varepsilon) - E(Y | Z = z_0 - \varepsilon)] = E(Y^1 - Y^0 | Z = z_0, G = c) P(G = c | Z = z_0)$$

- Per questo aggiungiamo il denominatore che esprime la probabilità di essere compliers condizionatamente a z_0 . Il denominatore serve a purificare il numeratore.

$$\lim_{\varepsilon \downarrow 0} E(T | Z = z_0 + \varepsilon) - E(T | Z = z_0 - \varepsilon) = P(G = c | Z = z_0)$$

Regression Discontinuity (fuzzy)

Nel caso della RD fuzzy non c'è una concordanza perfetta tra Z e T , dunque β_1 non esprime più l'ATE ma esprime l'intent-to-treat ossia l'intento iniziale di trattare l'unità ma non la ricezione effettiva del trattamento.

Il LATE può essere stimato con il 2SLS visto in ambito dell'IV:

$$1^\circ \text{ stage} \rightarrow \mathbf{P} = \beta_0 + \beta_1 * T + \beta_2 * Z$$

$$2^\circ \text{ stage} \rightarrow Y = \alpha_0 + \alpha_1 * \mathbf{P} + \alpha_2 * Z$$

dove P è una dummy che indica l'effettiva partecipazione dell'unità al programma / trattamento.

$$\begin{aligned} \mathbf{LATE} &= E(Y_1 - Y_0 | Z=z, G=c) = E(Y_1 | Z=z, G=c) - E(Y_0 | Z=z, G=c) = \\ &\alpha_0 + \alpha_1 + \alpha_2 * Z - \alpha_0 + \alpha_2 * Z = \alpha_1 \end{aligned}$$

Parte delle slide presentate derivano dal materiale dei corsi:

- Roberto Rocci [*“Inferenza Causale”*](#)
- Miguel Hernan [*“Causal Diagrams: Draw Your Assumptions Before Your Conclusions”*](#)
- Bruno Arpino [*“Course on Causal Inference”*](#)
- Peter Eibich & Angelo Lorenti [*“Advanced Methods for Causal Inference”*](#)
- Linda Valeri [*“Mediation Analysis and Causal Inference”*](#)

Libri “divulgativi” sull’inferenza causale:

-Miguel Hernan [“Causal Inference: What If \(the book\)”](#)

-Scott Cunningham [“Causal Inference The Mixtape”](#)

-Judea Pearl [“The book of why”](#)

Altri: <https://www.bradyneal.com/which-causal-inference-book>



SAPIENZA
UNIVERSITÀ DI ROMA

SEMINARI INFERENZA CAUSALE

Dipartimento di Sanità Pubblica e Malattie Infettive
Aula Scuola di Specializzazione in Statistica Sanitaria e Biometria
AULA C1 ore 12:00 - 14:00

Martedì 29 Novembre: Introduzione alla causalità

Dr.ssa MARGHERITA MORETTI

Dottorato in Scuola di Scienze Statistiche
Sapienza Università di Roma

Dr.ssa LAURA MONTELISCIANI

Dottorato in Sanità Pubblica Biostatistica ed Epidemiologia
Università degli Studi di Milano-Bicocca

Martedì 6 Dicembre: Metodi per l'inferenza causale

Margherita Moretti

margherita.moretti@uniroma1.it

Laura Montelisciani

l.montelisciani@campus.unimib.it

PER SCARICARE LE SLIDE:

[https://github.com/MMargherita/
seminari_DSPMI](https://github.com/MMargherita/seminari_DSPMI)