

Matching statistico

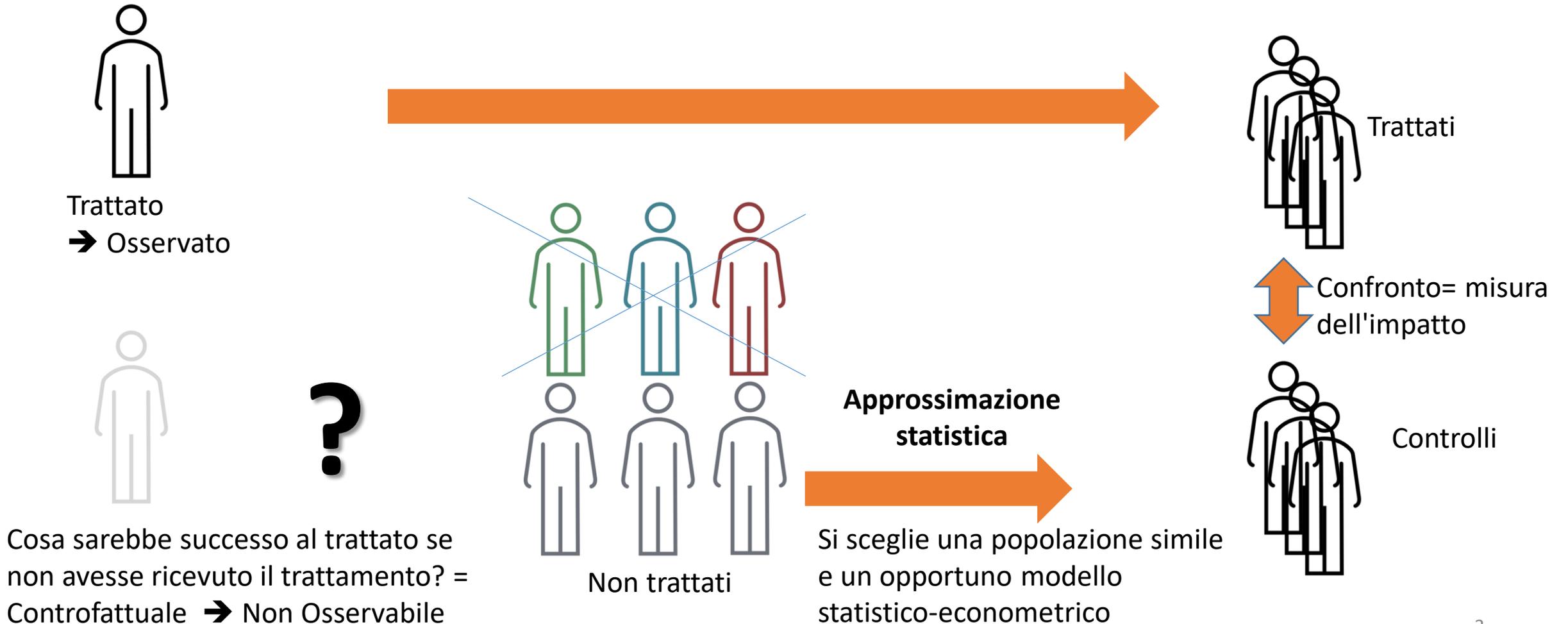
Valutazione di politiche

Simona Comi
Management e Design dei servizi
AA 2021/2022

MATCHING

- Lo scopo principale delle tecniche di Matching è quello di riprodurre il gruppo dei trattati utilizzando i non trattati, selezionando cioè un gruppo di controllo ex-post, e quindi ristabilire le condizioni sperimentali in un contesto non sperimentale. Sotto alcune ipotesi, il metodo riesce a ricostruire il corretto campione controfattuale.
- Le ipotesi che si fanno garantiscono che l'unica differenza che rimane tra trattati e controlli è la partecipazione al trattamento.
- Il matching può essere usato sia con cross-section che con dati longitudinali. Nella sua formula standard, però, di solito la dimensione longitudinale non viene utilizzata.

Approssimare il controfattuale



Matching

- Usando la logica della selezione sulle caratteristiche osservabili, possiamo stimare l'effetto causale.
- Bisogna costruire un campione di confronto usando i controlli che abbiano le stesse caratteristiche dei trattati
- Ovvero, bisogna accoppiare ad ogni trattato il controllo più simile
- **Esempio:** vogliamo stimare l'effetto causale di un corso di formazione sul reddito di un individuo
- **Ipotizziamo** che l'età sia l'unico *counfounding factor* che influenza sia il reddito che la decisione di fare formazione.

Trainees			Non-trainees			Matched Sample		
unit	Age	Earnings	Unit	age	earnings	unit	age	Earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000			
3	29	14400	3	30	21000			
4	25	20800	4	27	9300			
5	29	6100	5	54	41100			
6	23	28600	6	48	29800			
7	33	21900	7	39	42000			
8	27	28800	8	28	8800			
9	31	20300	9	24	25500			
10	26	28100	10	33	15500			
11	25	94000	11	26	400			
12	27	14300	12	31	26600			
13	29	12500	13	26	16500			
14	24	19700	14	34	24200			
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	23	9500			
Avg	28.5	16426		33.1	20605			

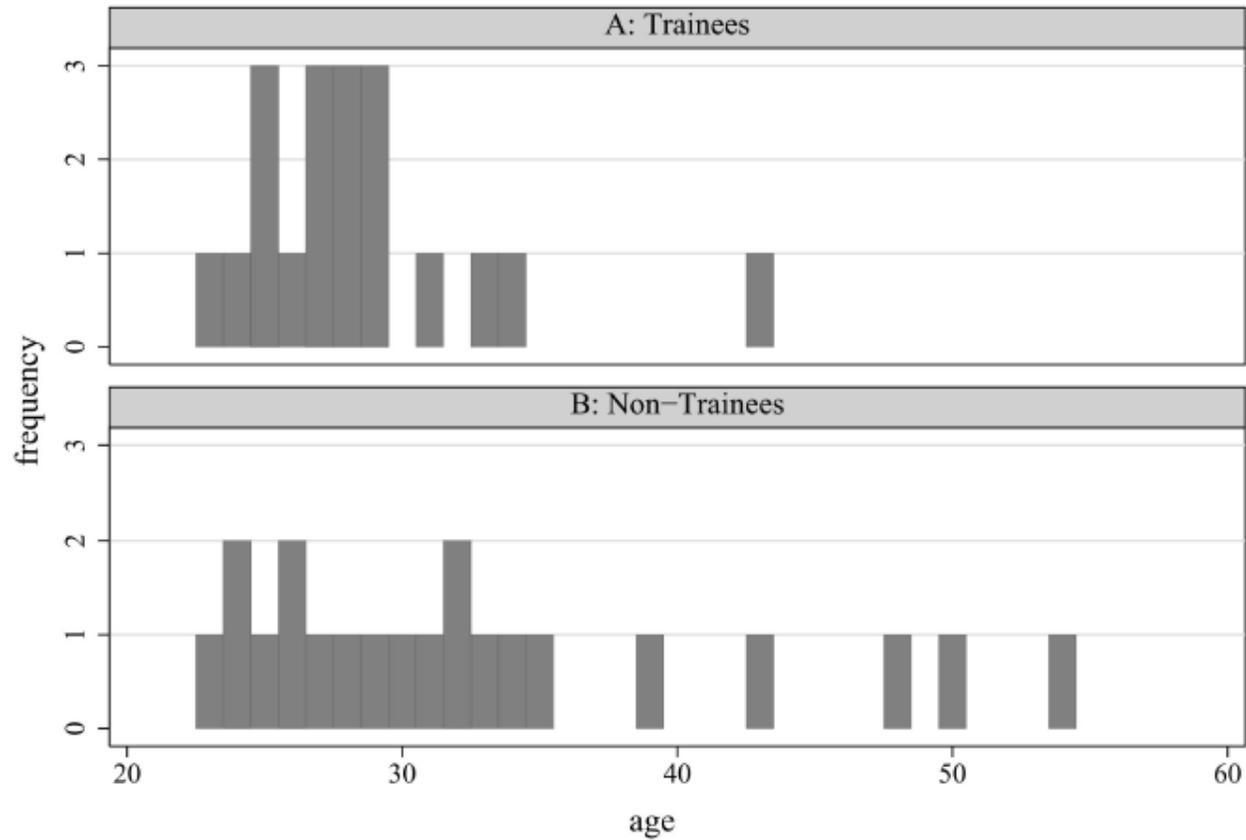
Trainees			Non-trainees			Matched Sample		
unit	Age	Earnings	Unit	age	earnings	unit	age	Earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000			
4	25	20800	4	27	9300			
5	29	6100	5	54	41100			
6	23	28600	6	48	29800			
7	33	21900	7	39	42000			
8	27	28800	8	28	8800			
9	31	20300	9	24	25500			
10	26	28100	10	33	15500			
11	25	94000	11	26	400			
12	27	14300	12	31	26600			
13	29	12500	13	26	16500			
14	24	19700	14	34	24200			
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	23	9500			
Avg	28.5	16426		33.1	20605		18.5	13.982

Trainees			Non-trainees			Matched Sample		
unit	Age	Earnings	Unit	age	earnings	unit	age	Earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300			
5	29	6100	5	54	41100			
6	23	28600	6	48	29800			
7	33	21900	7	39	42000			
8	27	28800	8	28	8800			
9	31	20300	9	24	25500			
10	26	28100	10	33	15500			
11	25	94000	11	26	400			
12	27	14300	12	31	26600			
13	29	12500	13	26	16500			
14	24	19700	14	34	24200			
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	23	9500			
Avg	28.5	16426		33.1	20605			

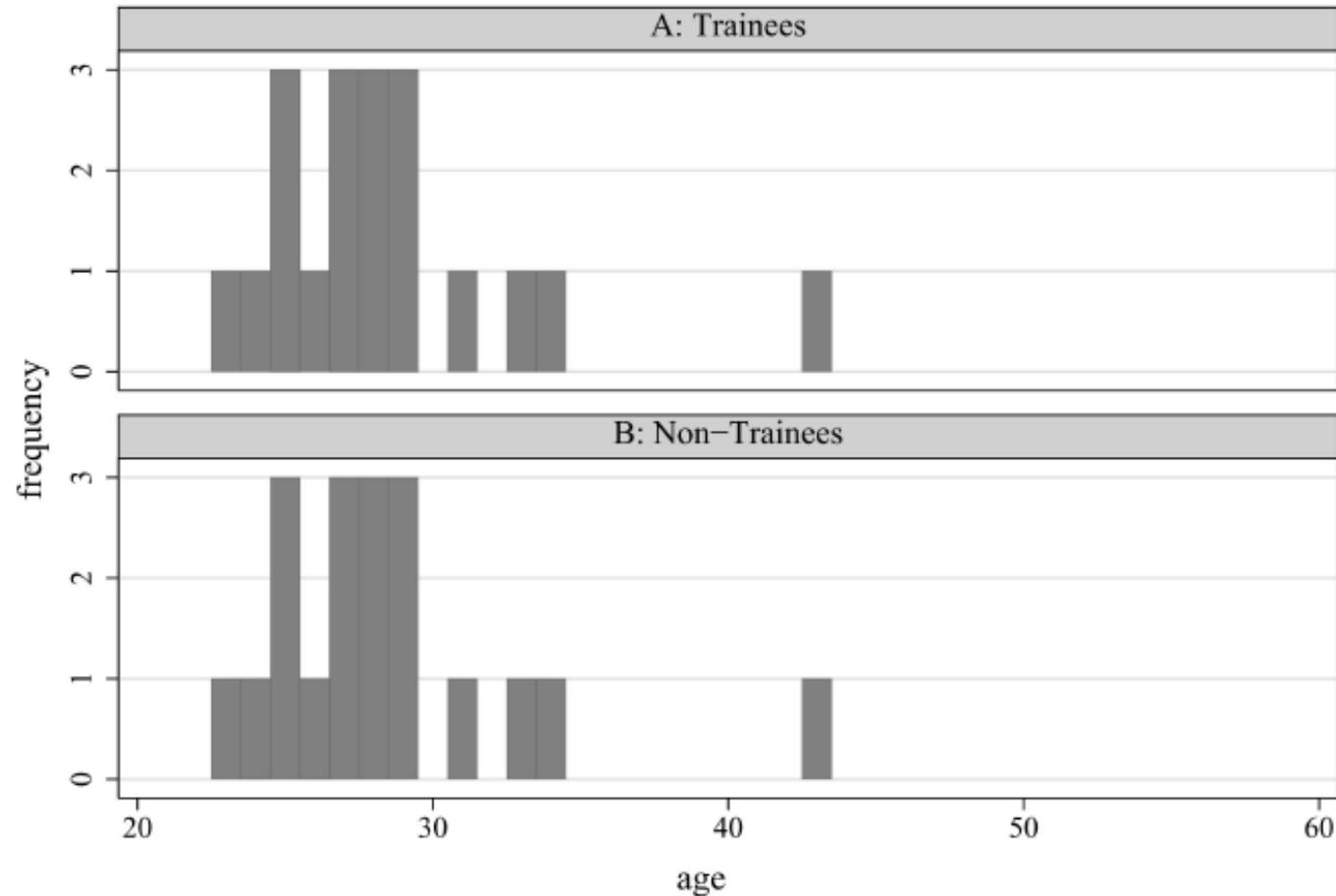
Trainees			Non-trainees			Matched Sample		
unit	Age	Earnings	Unit	age	earnings	unit	age	Earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300	15	25	23300
5	29	6100	5	54	41100	17	29	6200
6	23	28600	6	48	29800	20	23	9500
7	33	21900	7	39	42000	10	33	15500
8	27	28800	8	28	8800	4	27	9300
9	31	20300	9	24	25500	12	31	26600
10	26	28100	10	33	15500	11,13	26	8450
11	25	94000	11	26	400			
12	27	14300	12	31	26600			
13	29	12500	13	26	16500			
14	24	19700	14	34	24200			
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	23	9500			
Avg	28.5	16426		33.1	20605			

Trainees			Non-trainees			Matched Sample		
unit	Age	Earnings	Unit	age	earnings	unit	age	Earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300	15	25	23300
5	29	6100	5	54	41100	17	29	6200
6	23	28600	6	48	29800	20	23	9500
7	33	21900	7	39	42000	10	33	15500
8	27	28800	8	28	8800	4	27	9300
9	31	20300	9	24	25500	12	31	26600
10	26	28100	10	33	15500	11,13	26	8450
11	25	94000	11	26	400	15	25	23300
12	27	14300	12	31	26600	4	27	9300
13	29	12500	13	26	16500	17	29	6200
14	24	19700	14	34	24200	9,16	24	17700
15	25	10100	15	25	23300	15	25	23300
16	43	10700	16	24	9700	1	43	20900
17	28	11500	17	29	6200	8	28	8800
18	27	10700	18	35	30200	4	27	9300
19	28	16300	19	23	9500	8	28	8800
Avg	28.5	16426		33.1	20605		18.5	13.982

Disribuzione dell'età prima del matching



Distribuzione dell'età dopo il matching



In questo caso si dice che i campioni sono "bilanciati" sulla caratteristica osservabile età. Il bilanciamento viene di solito testato (con semplici ttest).

Stima dell'effetto del training sul reddito

- Differenza nel reddito medio tra trattati e controlli:
 - Prima del matching:

$$16428 - 20605 = -4177$$

- Dopo il matching:

$$16426 - 13982 = 2444$$

- La scelta delle X , che rappresentano le variabili con cui vengono accoppiati gli individui, è abbastanza delicata.
- Poiché lo scopo del metodo di valutazione è controllare per la selezione, le informazioni più corrette sono quelle disponibili per gli individui al momento della scelta di partecipare o meno al trattamento.
- Si lavora quindi sotto l'ipotesi che ciò che rimane non spiegato sia random rispetto all'essere trattati o no.
- La soluzione suggerita dal matching si basa sulla seguente ipotesi:

- **M1: (conditional independence assumption - CIA)**

Condizionatamente alle variabili X , gli outcomes dei non trattati sono indipendenti dal non aver partecipato al trattamento.

Conditional independence assumption (CIA)

- Condizionatamente alle X (conditional on X) gli outcome potenziali sono indipendenti dall'assegnazione al trattamento:

$$(Y_i^1, Y_i^0) \perp\!\!\!\perp D_i | X_i$$

- Dove (Y_i^1, Y_i^0) sono gli outcome potenziali (1 se trattati, 0 se non trattati e D_i è la dummy del trattamento.
- Per individui con lo stesso valore di X , l'assegnazione al trattamento è "as good as random"
- La probabilità di ricevere il trattamento è la stessa per individui con le stesse X .

Richiamo: Effetto della cia sulla stima

- Ricordiamo come avevamo definito l'ATT (o TOT)

$$\begin{aligned} & E\{Y_i \mid D_i = 1\} - E\{Y_i \mid D_i = 0\} \\ &= E\{Y_i(1) \mid D_i = 1\} - E\{Y_i(0) \mid D_i = 1\} \\ &+ E\{Y_i(0) \mid D_i = 1\} - E\{Y_i(0) \mid D_i = 0\} \\ &= \tau + E\{Y_i(0) \mid D_i = 1\} - E\{Y_i(0) \mid D_i = 0\} \\ &\text{dove } \tau = E\{\Delta_i \mid D_i = 1\} \text{ `e ATT.} \end{aligned}$$

- Per effetto della CIA avremo che il selection bias è nullo:

$$E\{Y_i(0) \mid D_i = 1\} - E\{Y_i(0) \mid D_i = 0\} = 0$$

τ Prende anche il nome di Conditional ATT

Ulteriore interpretazione della CIA

- Lavorare sotto la CIA equivale ad ipotizzare che le caratteristiche non osservabili dei non trattati nell'equazione dell'outcome siano indipendenti dalla partecipazione, conditional on X :

$$(u_i - u(X_i)) \perp d_i \mid X_i.$$

- Ovvero, conditional on X , i trattati e i non trattati sono confrontabili per ciò che concerne l'outcome in caso di non trattamento.
- Di conseguenza, secondo questa ipotesi, non può esserci selezione sul termine non osservabile u_i

- L'ipotesi M1 implica una sorta di versione condizionale dell'ipotesi di randomizzazione:

$$E[u_i | d_i, X_i] = E[u_i | X_i]$$

- Che, sotto l'ipotesi di esogeneità dell'X ci permette di ottenere $E[u_i]$
- L'implicazione pratica della M1 è che per ogni osservazione appartenente al gruppo di trattamento (y^1) possiamo cercare un (gruppo) individuo non trattato (y^0) con le stesse caratteristiche X ed essere certi che questo y^0 rappresenta un controfattuale corretto.
- Di solito la M1 viene testata con balancing tests, che possono essere semplici test di differenza delle medie delle X tra i due campioni (trattati e controlli).

- Il matching quindi consiste nel ricostruire un data-set che sia sperimentale. La sua capacità di farlo dipende dalla disponibilità del controfattuale. Cioè dobbiamo essere sicuri di poter riprodurre ogni trattato tra i non trattati.
- Questo fatto è illustrato nella seconda ipotesi del matching:
- M2 o **Common Support**: tutti gli individui trattati hanno una controparte nella popolazione dei non trattati che potenzialmente avrebbero potuto partecipare:
$$0 < P(d_i = 1 \mid X_i) < 1$$
- Per ogni valore di X, c'è una probabilità positiva diversa da 0 o 1 di essere trattati o no.
- In altre parole, non è possibile predire esattamente il trattamento per nessuno specifico valore di X

- Indichiamo con S il common support di X , cioè il sottospazio della distribuzione di X che rappresenta sia i trattati che i controlli.
- Sotto l'ipotesi M2 S è l'intero dominio di X .
- L'ATT può allora essere definito come:

$$\alpha^{ATT}(S) = E[y^1 - y^0 | d = 1, X \in S]$$

- Dove $\alpha^{ATT}(S)$ è l'impatto medio del trattamento sui partecipanti che hanno caratteristiche osservabili X e appartengono a S .

Lo stimatore matching per l'ATT

- Il matching perfetto non sempre è fattibile, a volte non si trovano "gemelli in X" per i trattati.
- Quindi in generale scriviamo:

$$\hat{\alpha}_{ATT} = \frac{1}{N_1} \sum_{D_i=1} (Y_i - Y_{j(i)})$$

- Dove $Y_{j(i)}$ è l'outcome dell'individuo j non trattato che ha il valore r più vicino possibile a X_i

Lo stimatore matching per l'ATT

- A volte invece di un solo non trattato "matchato" con un trattato se ne usano di più, prendendo la media dell'outcome:

$$\hat{\alpha}_{ATT} = \frac{1}{N_1} \sum_{D_i=1} \left\{ Y_i - \left(\frac{1}{M} \sum_{m=1}^M Y_{jm(i)} \right) \right\}$$

- Dove $M=2$ o a volte anche $M=5$

Come misurare la vicinanza

- Quando il vettore delle covariate (X) contiene più di una variabile (definiamo k il numero delle covariate, con $K > 1$)

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{pmatrix}$$

- E' opportuno trovare una misura adeguata per trovare i controlli più simili (o vicini) ai trattati

Distanza euclidea

- Una possibile misura è la Distanza Euclidea, definita come:

$$\begin{aligned} \|X_i - X_j\| &= \sqrt{(X_i - X_j)'(X_i - X_j)} \\ &= \sqrt{\sum_{n=1}^K (X_{ni} - X_{nj})^2}. \end{aligned}$$

- PB: la distanza euclidea è sensibile alla scala di misura delle X.
- Può quindi essere "normalizzata" per risolvere questo problema (vi risparmio la formula)

The curse of dimensionality

- Matching diventa problematico con troppe covariate
- Ciò avviene anche se dividiamo ogni covariata in *coarse categories*.
- Per esempio, se avessimo k covariate divise in 3 *coarse categories* (l'età potrebbe essere giovane, mezza età, vecchio; il reddito potrebbe essere basso, medio, alto etc) il numero di celle che otterremo pre matchare le osservazioni sarebbe 3^k , che per $k=10$ ci darebbe $3^{10} = 59049$ celle, all'interno delle quali dovremmo osservare sia trattati che controlli.

Propensity score matching

- Un alternativa più semplice è di accoppiare gli individui utilizzando una funzione delle X .
- Di solito il matching viene fatto sulla propensione alla partecipazione dato l'insieme delle caratteristiche X , ovvero sul propensity score:

$$P(X_i) = P(d_i = 1 | X_i)$$

- Definito come la probabilità di essere trattati conditional on a set of observed variables X .
- Intuitivamente, il $P(X_i)$ sintetizza tutte le informazioni contenute nelle X in un unico valore.
- Quindi possiamo semplicemente controllare per $p(X)$ per eliminare il selection bias

Propensity score matching

- Rosenbaum and Rubin(1983) hanno dimostrato che l'ipotesi CIA rimane valida quando si controlla per $P(X)$ invece che per X .

$$(Y_i^1, Y_i^0) \perp\!\!\!\perp D_i | p(X_i)$$

- Quando si utilizza $P(X)$, il gruppo di confronto per ogni individuo trattato viene scelto con un criterio pre-definito di vicinanza tra propensity scores.
- Dopo aver definito la vicinanza tra ogni osservazione trattata, il passo successivo è scegliere gli appropriati pesi per associare l'insieme di non trattati ad ogni trattato.
- Esistono diverse possibilità:
- **Nearest Neighbor matching** : assegna un peso pari a 1 all'osservazione dei non trattati più vicina e 0 a tutte le altre.
- **Kernel matching** definisce un insieme di osservazioni vicine ad ogni trattato e costruisce il controfattuale utilizzando tutti i non trattati vicini, non solo il più vicino. Il peso sarà uguale a 1 per tutti i non trattati nelle vicinanze del trattato e zero per tutti gli altri.