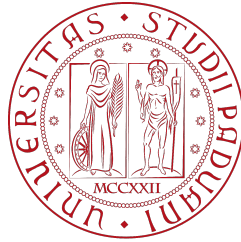


Università degli Studi di Padova
Dipartimento di Psicologia Generale
Corso di Laurea in
Scienze Psicologiche Cognitive e Psicobiologiche



**Inferenza causale
utilizzando tecniche di propensity score:
un'analisi del database SHARE**

Relatore: Prof. Massimo Nucci
Dipartimento di Psicologia Generale

Correlatore: Dott. Paolo Girardi
DAIS - Università Ca' Foscari Venezia

Laureanda: Laura Sità
Matricola N. 1221179

Anno Accademico 2021/2022

Indice

1	Introduzione	1
1.1	Euristiche e bias	1
1.2	Causalità e correlazione	2
1.3	Definizione di causalità	3
1.4	Obiettivo	4
2	Diversi tipi di studi	6
	Directed acyclic graphs (DAG)	6
2.1	Studi randomizzati	7
2.1.1	Il concetto di outcome potenziali	8
2.2	Studi osservazionali	9
2.2.1	Esempio di studio osservazionale	10
2.3	Studi quasi sperimentali	11
2.3.1	Aggiustamento additivo tramite regressione lineare	11
2.3.2	Matching e propensity score	14
3	Applicazione pratica	17
3.1	Metodo statistico	17
3.1.1	Regressione logistica	17
3.1.2	Propensity score (PS)	19
3.2	Effetto del fumo di sigaretta sul rischio di decadimento cognitivo	21
3.2.1	Ipotesi	22
3.2.2	Risultati della regressione logistica con e senza propensity score	23
3.2.3	Confronto dei risultati	25
3.3	Discussione e limiti dello studio	25
4	Conclusione	27

Ringraziamenti

Appendice A. Directed acyclic graphs

Appendice B. Codice dell'analisi riferita al capitolo 3.2

Statistiche descrittive stratificate per genere

Statistiche inferenziali

Bibliografia

1 Introduzione

Ho svolto questa tesi durante un periodo di tirocinio interno presso il Dipartimento di Psicologia dello Sviluppo e della Socializzazione, dove ho avuto modo di vivere in prima persona l'ambiente della ricerca scientifica all'interno del nostro Ateneo. Nel periodo compreso fra ottobre e dicembre 2021, sono stata seguita dal Prof. Paolo Girardi che, in qualità di tutor aziendale e correlatore esterno, mi ha introdotto nell'ambito, presentandomi nozioni teoriche di psicomètria, che sono andate a complementare quelle già acquisite tramite il mio percorso di studi, e fornendomi del materiale - paper e ricerche - che ha costituito fin da subito un esempio tangibile di quanto apprendevo. Questa esperienza è stata per me un'ottima occasione sia per iniziare a scoprire dall'interno il mondo della ricerca, nella speranza di poter continuare a farlo anche in futuro, che per applicare le conoscenze già acquisite e quelle apprese in itinere a delle reali ipotesi di ricerca.

Il tema di questa tesi è incentrato sullo studio dell'inferenza causale in psicologia e su come condurre una buona analisi statistica indagando questo particolare tipo di relazione. Anche se riportare degli esempi rappresenta sicuramente la via più immediata per comunicare quanto si prefigge questa trattazione, partiamo prendendo in considerazione degli aspetti teorici, a cominciare dalla definizione delle circostanze in cui si può parlare di inferenza causale e delle difficoltà che questo ragionamento comporta nella scienza, ma anche nella vita di ogni giorno.

1.1 Euristiche e bias

Produrre inferenze è una sfida per le persone, perché possiedono una memoria a breve termine limitata con cui lavorare (Hogg & Vaughan, 2016), ma possono, d'altra parte, fare affidamento su una memoria a lungo termine estremamente capace: risulta particolarmente conveniente sfruttare questa funzione, conservando le informazioni che man mano acquisiamo con l'esperienza, sotto forma di schemi nella memoria a lungo termine, che vengono richiamati in seguito, quando necessario. L'inferenza, il ragionamento logico che garantisce la possibilità di generalizzare la conoscenza a partire da dati posseduti, risulta quindi un processo largamente guidato dagli schemi, poiché permette di trarre conclusioni a sostegno di conoscenze di cui già disponiamo. L'uso di tali schemi sarebbe regolato sulla base di strategie cognitive definite euristiche, delle scorciatoie di pensiero che permettono di individuare la soluzione a un problema avente caratteristiche complesse a partire da più semplici operazioni di giudizio.

Anche se la maggior parte delle volte i nostri quotidiani processi inferenziali sembrano adeguati, può capitare che si dimostrino sbagliati e che ne rimaniamo vittime, con il risultato di commettere ragionamenti erronei: in questi casi, le scorciatoie si rivelano distorsioni, talvolta inevitabili, che prendono il nome di bias cognitivi. Un esempio di bias cognitivo è dato dalla tendenza umana che induce a cercare le cause degli eventi e i loro collegamenti

superficiali: questa inclinazione, nella sua inconsapevole urgenza, porta a confondere i due diversi tipi di relazione che stanno alla base degli eventi, correlazione e causazione. Questo fenomeno è stato largamente oggetto di interesse e studi di ampia portata, perché, in psicologia come in tutte le scienze, la distinzione fra correlazione e causazione ne costituisce un aspetto fondamentale (Bleske-Rechek et al., 2015).

1.2 Causalità e correlazione

Correlazione implica il cambiamento di una prima variabile al variare di una seconda, senza ulteriori precisazioni. Una cosa diversa è, invece, la causalità: per poter affermare che una variabile X causa una variabile Y , è necessario mostrare che X precede Y . Per questo motivo si tratta di due concetti diversi: correlazione non implica causazione. Ad esempio, la correlazione fra altezza e peso non implica che il peso di una persona sia dovuto esclusivamente alla sua altezza. D'altro canto, se tracciare questa distinzione a parole richiede semplicemente un'attenzione ai termini con cui ci si esprime, riconoscere il collegamento fra due variabili nella realtà non è sempre così immediato, soprattutto nell'ambito delle scienze sociali, dove intuire la presenza relazioni causali, anche dove non ci sono, avviene in modo così naturale da compromettere talvolta la logicità di un ragionamento.

A questo proposito, Bleske-Rechek e collaboratori (Bleske-Rechek et al., 2015) hanno provato ad indagare il grado con cui le persone in generale tendono a compiere inferenze causali a partire da ipotetiche descrizioni di studi sperimentali e non sperimentali riguardo il comportamento umano. Nel caso degli studi sperimentali, i soggetti avrebbero potuto derivare giustamente inferenze causali, al contrario di quanto sarebbe legittimo fare nel caso di studi non sperimentali, in virtù dell'impostazione dei primi che mira ad indagare l'effetto di una o più variabili manipolate dallo sperimentatore su una o più variabili oggetto di misurazione.

In particolare, sono stati individuati 4 gruppi di soggetti, composti da 30 persone ciascuno, ad ognuno dei quali veniva mostrata una diversa versione della vignetta presentante l'ipotetica ricerca. I due elementi modificati per ogni vignetta, che andavano a determinare le quattro differenti condizioni, erano

- il fatto che la ricerca fosse sperimentale o non sperimentale,
- una relazione positiva o negativa fra le variabili di interesse di ciascuna ricerca.

Siccome ciascuna di queste quattro condizioni veniva declinata per 3 diversi temi di ricerca, quali videogame e aggressività, autostima e performance accademica, consumazione di materiale pornografico e soddisfazione coniugale, si ottenevano 12 gruppi sperimentali.

Dopo aver letto la vignetta, ai partecipanti veniva chiesto di

1. selezionare la rappresentazione grafica che secondo loro descriveva la condizione sperimentale che avevano analizzato,
2. leggere sei affermazioni e scegliere quali avrebbero costituito delle inferenze appropriate che i ricercatori avrebbero potuto fare sulla base dei risultati ottenuti.

Dallo studio è emersa, in primo luogo, la tendenza da parte dei soggetti a tracciare relazioni causali a partire da dati non sperimentali tanto spesso quanto dai dati sperimentali: ad esempio, partecipanti che leggevano riguardo uno studio in cui *era rilevato* oppure *era manipolato* il tempo dedicato a giocare ai videogiochi erano ugualmente predisposti a inferire che giocare ai videogiochi influenzasse l'aggressività, quando tale inferenza sarebbe stata lecita solo nel secondo caso, in virtù dell'intervento, ovvero della manipolazione della variabile, da parte dello sperimentatore. In secondo luogo, è emersa la tendenza a favorire ragionamenti che coincidevano con nozioni già possedute, a prescindere dal contenuto dei fatti oggettivi presentati: i partecipanti erano molto più portati a inferire l'effetto causale dell'autostima sulla performance accademica se descritto come un effetto positivo piuttosto che nei casi in cui era descritto come un effetto negativo, proprio perché è opinione comune pensare che avere un'alta autostima stia a monte di un buon rendimento accademico (e non il contrario).

In conclusione, lo studio documenta come le persone traggano frequentemente inferenze causali a partire da dati non causali e come la direzione delle loro inferenze subisca inevitabilmente l'influenza delle nozioni già possedute, indipendentemente dai risultati presentati. Queste tendenze inconsapevoli, per quanto siano frutto di meccanismi volti a semplificare i nostri ragionamenti quotidiani, rappresentano un rischio per la corretta comprensione della scienza. Pertanto, risulta necessario rendere sempre chiara la propria esposizione soprattutto quando si intende estendere il proprio contributo ad un pubblico più ampio possibile e, inoltre, precisare cosa si intende per causalità.

1.3 Definizione di causalità

Molte situazioni sociali sono in grado di innescare analisi causali di tipo implicito, delle quali possiamo non essere consapevoli. Ci sono, però, situazioni in cui queste analisi diventano intenzionali, soprattutto quando un evento ci appare inaspettato e sorprendente. Le analisi che siamo portati a fare, che rischiano talvolta di rivelarsi erranee perché soggette a bias, sono dette attribuzioni causali. Per attribuzione causale intendiamo il processo messo in atto al fine di cercare spiegazioni per il proprio e per l'altrui comportamento, inferendo le cause che stanno dietro specifiche azioni e sentimenti. Ad esempio, facendo riferimento all'esperienza di un dato individuo coinvolto in un fenomeno, potremmo essere interessati a sapere se tale fenomeno abbia avuto delle conseguenze, quindi abbia influito causalmente, sull'individuo o meno: in altre parole, ci chiediamo se ciò che abbiamo di fronte sia l'effetto di una relazione causale. Ne consegue che la definizione di causalità consiste nel contrasto

fra l'esito (che chiameremo outcome) del fenomeno a cui l'individuo è stato esposto (ad esempio un trattamento ricevuto) in un particolare momento e lo stato che presenterebbe nello stesso particolare momento lo stesso individuo se non fosse stato esposto al fenomeno. Di conseguenza, ottenere inferenza causale risulta difficile proprio perché ci è dato osservare solo uno dei due scenari: quello che si è effettivamente verificato (Rosenbaum, 2018).

Un chiaro esempio è fornito dalla situazione che si era delineata all'inizio del ProCESS Trial, un caso riguardante la cura dello shock settico, nato dalla collaborazione fra ricercatori che lavoravano in 31 dipartimenti di emergenza degli Stati Uniti, riportato da uno studio pubblicato nel 2014 sulla rivista scientifica "New England Journal of Medicine" (Lilly, 2014). Lo shock settico è una patologia grave che insorge quando, in seguito alla diffusione di un'infezione, la pressione sanguigna tende ad abbassarsi notevolmente ed è spesso letale per bambini e anziani.

I risultati dello studio sono emersi dal confronto tra i due protocolli a disposizione per trattare la malattia: alcuni pazienti hanno ricevuto il protocollo più aggressivo dei due, mentre altri quello meno aggressivo. Ogni paziente è stato messo nella condizione di avere pari possibilità di ricevere un trattamento piuttosto che l'altro e, logicamente, per ognuno di loro si è potuto osservare solo l'effetto del trattamento ricevuto. In altre parole, in casi come questi, a livello del singolo individuo, non è possibile comparare gli esiti di entrambi i trattamenti e quindi di fare attribuzioni causali (Rosenbaum, 2018).

1.4 Obiettivo

Per quanto sia un concetto difficile a cui approcciarsi, l'inferenza causale descrive un tipo di legame fra variabili particolarmente informativo rispetto ad una semplice correlazione. La correlazione, infatti, rappresenta una via più sicura nello studio di due fenomeni, ma sicuramente meno ricca di significato. Di conseguenza, l'obiettivo di questa tesi consiste nell'illustrare, ricorrendo anche a riferimenti concreti, i limiti e le possibilità entro i quali è possibile fare inferenza causale e quindi nell'incoraggiarne la diffusione all'interno del panorama sperimentale, soprattutto della psicologia, offrendo essa stessa degli esempi.

In linea con questo proposito, nel seguente capitolo verranno illustrati più tipi di studi, differenziati sulla base della loro impostazione iniziale, assieme alle relative modalità di condurre inferenza causale. Il capitolo successivo, invece, sarà dedicato all'esposizione teorica e quindi pratica, eseguita su un dataset europeo a partire da una personale ipotesi causale, di una particolare tecnica, il propensity score, utile per l'adattamento del campione a disposizione e, di conseguenza, per operare legittime inferenze causali.

2 Diversi tipi di studi

In psicologia, esistono molte analisi che hanno affrontato lo studio di variabili dal punto di vista correlazionale. D'altro canto, è sentita la necessità di ampliare le conoscenze in questo campo potendo rispondere a condivise domande di ricerca che mirano a identificare le cause a monte dei fenomeni studiati da questa disciplina. Inoltre, l'inferenza causale non è di centrale importanza solo per la psicologia, ma è, come afferma Foster, inevitabile (Foster, 2010). Egli, infatti, sostiene l'essenzialità dell'inferenza causale in tre modi: prima di tutto, siccome lo scopo principale della psicologia consiste nel migliorare le vite umane, è fondamentale che si avvalga di uno strumento che permetta di intervenire a priori; in secondo luogo, l'analisi a livello causale è un tipo di ragionamento inevitabile a cui il pensiero umano tende inconsapevolmente; infine, siccome spesso delle semplici associazioni (come lo sono le correlazioni) sono confuse con le relazioni causali, può capitare che il pensiero di coloro che non sono del mestiere ricada in inferenze erranee, comportando conseguenze anche gravi.

Pertanto, per quanto ostica possa risultare questa via, è necessario proporsi di stabilire una plausibile inferenza causale come obiettivo della ricerca empirica in psicologia. In tal senso, per spostarsi dalla correlazione alla causalità, Foster suggerisce di farsi aiutare da due strumenti: il digrafo aciclico direzionato (o, in inglese, *directed acyclic graph* - DAG) e il concetto di outcome potenziali (Foster, 2010).

Directed acyclic graphs (DAG)

I DAG costituiscono uno strumento particolarmente utile per chiarire ipotetiche relazioni causali e identificare se e in che modo possono essere testate: grazie ad essi, è possibile rappresentare graficamente gli elementi che compongono uno studio. In particolare, i cosiddetti 'nodi' o 'vertici' del grafo stanno ad indicare le variabili, mentre la loro interazione diretta è definita dalle linee che li uniscono, dette 'archi' o 'connessioni'. Se come archi vengono utilizzati dei semplici segmenti, allora il grafo è detto non direzionato e codifica solo il "grado di affinità" (ovvero la correlazione) fra le variabili: in questo caso il grafo rappresenta solo in che modo le variabili si influenzano a vicenda, senza però poter stabilire cause o effetti. Se invece come archi vengono utilizzate delle frecce, il grafo è detto direzionato proprio perché codifica relazioni di parentela fra le variabili: la direzione di queste frecce rappresenta l'ipotetico rapporto causale fra di esse. Pertanto, i DAG costituiscono un esempio di grafo direzionato di tipo aciclico, caratteristica necessaria per rappresentare nessi causali fra le variabili, siccome la causa deve precedere il suo effetto.

Facendo riferimento al Grafico 1 (si veda l'Appendice A), possiamo identificare una variabile X , che rappresenta un fattore di rischio, avente un effetto causale sul rispettivo outcome Y . Tale relazione causale presentata una condizione 'ideale' che, in quanto tale, coinvolge

esclusivamente le due variabili di interesse; d'altra parte, siccome i DAG servono anche ad illustrare i fenomeni appartenenti alla realtà, quasi sempre vanno inserite anche altre variabili, per l'influenza che esercitano sulle prime. Nel diagramma, infatti, possono essere anche incluse le variabili, rappresentate da Z , dette confondenti, che possono comparire in diversi punti del DAG (e legate in diversi modi alle altre) in base al ruolo, che può rivelarsi critico per lo studio.

Prendendo in esame un caso specifico come quello presentato nella rivista *Annual Review* (D'Onofrio et al., 2020), come è possibile osservare nel Grafico 2.1 (si veda l'Appendice A), Z agisce da confondente, in quanto rappresenta tutta la serie di fattori che causano sia X che Y . Pertanto, si definisce spuria o non-causale l'associazione che sorge quando il fattore di rischio (X) e l'outcome (Y) condividono cause comuni (Z). Per risolvere questa condizione e ri-ottenere la relazione causale 'ideale', è possibile condizionare, ovvero fissare ad un valore costante, Z tramite vari metodi cosicché l'unico percorso aperto (alla verifica sperimentale) rimanga quello che congiunge direttamente X con Y (si veda il Grafico 2.2 in Appendice A).

Per concludere, poter affermare l'esistenza di una relazione causale presuppone che vengano bloccati tutti i percorsi non casuali che intercorrono fra X e Y : il procedimento che lo rende possibile è detta associazione condizionale. Un modo per avere questo 'blocco' delle variabili confondenti in modo naturale è costituito dall'impostazione di un particolare tipo di studi, quelli randomizzati.

2.1 Studi randomizzati

La randomizzazione è la tecnica di elaborazione dei gruppi su cui condurre l'indagine in grado di rendere sperimentale lo studio, perché ogni soggetto ha pari possibilità di ricevere il trattamento, e quindi di essere parte del cosiddetto gruppo sperimentale, o non ricevere il trattamento (oppure di ricevere un trattamento diverso dal primo), e quindi di rientrare nel gruppo di controllo.

Questa particolare condizione implica che i soggetti appartenenti ai due gruppi siano equilibrati dal punto di vista delle caratteristiche individuali, sia delle variabili covariate osservate (le variabili che potrebbero influire nello studio e di cui pertanto si tiene conto all'interno del modello che andrà testato), sia delle covariate non osservate, siccome entrambe tendono a distribuirsi 'per caso' in maniera omogenea in entrambi i gruppi. Questa proprietà degli studi randomizzati è definita simmetrica.

Facendo riferimento ad un caso già introdotto (Lilly, 2014), il ProCESS Trial fornisce un esempio di studio randomizzato. In questo studio, le variabili covariate prese in considerazione erano l'età, il genere, se il paziente era ricoverato in una casa di riposo o meno e il suo stato di salute pre-trattamento (rappresentato dal punteggio al test "Acute Physiology and Chronic Health Evaluation" - APACHE II - usato per la valutazione di

malattie gravi). Essendo determinata prima del trattamento assegnato, una covariata non influisce sul tipo di trattamento assegnato al paziente, al contrario dell'outcome, che si rileva dopo l'assegnazione al trattamento. In linea con quanto esposto in precedenza, il valore dell'outcome esiste in due versioni: l'esito a cui va incontro il paziente avendo ricevuto il protocollo aggressivo o avendo ricevuto il protocollo meno aggressivo. In base alle circostanze del ProCESS Trial, è vista come successo la situazione in cui il paziente viene dimesso vivo dall'ospedale nel giro di 60 giorni (dal compimento del trattamento), come insuccesso il verificarsi della dipartita in ospedale del paziente all'interno dello stesso arco di tempo. Paragonando i potenziali outcome che ogni paziente manifesterebbe sotto trattamenti alternativi si potrebbe ottenere l'effetto causale del trattamento.

2.1.1 Il concetto di outcome potenziali

Nel concetto di outcome potenziali giace il cuore dell'inferenza causale. Questo framework, nella sua basilare importanza per questo tipo di ragionamento (Foster, 2010), permette di concentrarsi su un trattamento (X) e il suo outcome (Y), assegnando $X = 1$ alla condizione di trattamento e $X = 0$ a quella di controllo. A caratterizzare ogni individuo ci sono quindi tre variabili: Y_1 (l'effetto del trattamento), Y_0 (l'effetto della condizione di controllo) e X (il trattamento ricevuto). D'altro canto, risulta impossibile poter osservare tutte e tre queste variabili, ma soltanto X assieme a Y_1 oppure Y_0 . Quindi, sostituendo X con 1 o 0 a seconda del singolo caso, si può definire l'outcome osservato per ciascun soggetto come

$$Y_{obs} = X \cdot Y_1 + (1 - X) \cdot Y_0$$

Oggetto di interesse è come definire idealmente l'effetto atteso del trattamento, τ , ovvero la differenza fra Y_1 e Y_0 per ciascun individuo i . Con E che rappresenta il valore atteso, si ha

$$\tau = E[Y_{1,i} - Y_{0,i}]$$

Siccome risulta materialmente impossibile ottenere sia Y_1 che Y_0 per uno stesso individuo, il ragionamento va trasposto nei termini del campione di cui disponiamo. Pertanto, l'effetto del trattamento sarà dato dalla differenza pesata fra gli outcome dei soggetti trattati e gli outcome dei soggetti non trattati. In altre parole, l'effetto complessivo del trattamento (o, in inglese, *avarage effect of treatment* - ATE) risulterà dalla differenza fra l'effetto complessivo del trattamento per i trattati (o, in inglese, *avarage effect of treatment for the treated* - ATT) e l'effetto complessivo del trattamento per i non trattati (o, in inglese, *avarage effect of treatment for the untreated* - ATU):

$$ATE = ATT - ATU$$

Ripercorrendo lo stesso ragionamento nei termini concreti del ProCESS Trial, secondo quanto riportato da Rosenbaum (Rosenbaum, 2018), gli outcome potenziali a cui possono

andare incontro i pazienti dello studio sono costituiti dalla condizione di assegnazione al trattamento aggressivo, X_1 , e al trattamento meno aggressivo, X_0 . Pertanto, trattandosi di uno studio randomizzato, la probabilità che un paziente qualsiasi, il paziente i , riceva il trattamento aggressivo, quindi $Pr(X_1)$, che chiamiamo π_i , è del 50%:

$$\pi_i = Pr(X_i = 1) = 1/2$$

A questo punto, è possibile verificare la presenza di una relazione causale fra protocollo subito e mortalità di ciascun soggetto, confrontando, secondo la definizione stessa di effetto causale, i due outcome potenziali per lo stesso paziente i : r_{Ti} , l'effetto del protocollo più aggressivo (dove T sta per trattamento), e r_{Ci} , l'effetto del protocollo meno aggressivo (dove C sta per controllo). Pertanto si ha $r_{Ti} = 1$ oppure $r_{Ci} = 1$ a seconda che il paziente venga dimesso vivo dall'ospedale nel giro di 60 giorni (condizione di successo), $r_{Ti} = 0$ oppure $r_{Ci} = 0$ nel caso della sua dipartita prima che venga dimesso dall'ospedale (condizione di insuccesso). Prendendo, ad esempio, in considerazione il diciassettesimo paziente coinvolto nel trial, i due possibili esiti sono definiti r_{T17} e r_{C17} , però, avendo tale paziente - per caso - ricevuto il trattamento più aggressivo, possiamo sapere effettivamente il valore solo del primo: per questo motivo gli outcome potenziali sono talvolta chiamati controfattuali.

Proseguendo secondo il ragionamento di Foster, l'effetto causale del trattamento, δ_i , è dato dalla differenza fra i due potenziali outcome per lo stesso soggetto i , $\delta_i = r_{Ti} - r_{Ci}$. D'altro canto, come già evidenziato, non ci è dato osservare entrambi i termini della differenza e, di conseguenza, nemmeno δ_i .

Trattandosi, però, il ProcESS trial di uno studio randomizzato ed essendo quindi ragionevolmente sicuri che non ci siano differenze significative fra i gruppi, possiamo osservare gli effetti del protocollo più aggressivo per metà del campione e gli effetti del protocollo meno aggressivo per l'altra metà. In questo modo risulta possibile effettuare legittimamente delle inferenze circa l'effetto causale dei protocolli sui pazienti a livello della popolazione, senza però poter fare altrettanto riguardo il singolo soggetto.

2.2 Studi osservazionali

Contrariamente a quanto è avvenuto nell'esempio presentato, nella realtà dei fatti che la scienza si propone di indagare, soprattutto la psicologia, gli eventi non possono sempre essere indotti artificialmente dallo sperimentatore (per motivi etici e/o perché si tratta di fenomeni al di fuori della sua portata) e, di conseguenza, i soggetti non possono essere caratterizzati da una stessa probabilità di appartenere al gruppo sperimentale (condizione asimmetrica). Questo è un altro problema (o una conseguenza del problema di fondo) che caratterizza l'inferenza causale negli studi cosiddetti osservazionali. Secondo Foster, quando non è possibile randomizzare gli individui nei gruppi di trattamento, come in questi casi, la soluzione consiste nel fare solide assunzioni riguardo l'assegnazione alle condizioni sperimentali (Foster, 2010). Una tale assunzione è l'ignorabilità: essa permette di sostenere che,

fra individui con un simile profilo di covariate, l'assegnazione alle condizioni sperimentali avviene come se fosse stata possibile la randomizzazione. In altre parole, sotto questa assunzione, lo studio non risente degli effetti di variabili confondenti non osservabili (e quindi non osservate) e il valore che assume l'outcome per i diversi livelli del trattamento è legato dal tipo di trattamento.

Per capire il ruolo che ricoprono le covariate non osservate, prendiamo in esame un esempio di relazione fra X e Y dove la variabile Z funge da mediatore fra trattamento e outcome (si veda il Grafico 3.1 in Appendice A). Come già affermato, per ottenere un diretto effetto da parte di X su Y , basterebbe condizionare Z . D'altro canto, come illustra il DAG (si veda il Grafico 3.2), aggiustare per il mediatore, induce una relazione spuria fra il trattamento e W , una variabile non osservata. A causa della presenza stessa di W , la stima dell'effetto diretto di X su Y subisce una distorsione.

2.2.1 Esempio di studio osservazionale

L'esempio pratico di uno studio osservazionale si delinea facendo riferimento a come si prospettava l'emergenza sanitaria mondiale dovuta al virus COVID-19 durante i mesi di marzo e aprile 2021, quando, a breve distanza dall'avvio, la campagna vaccinale si prospettava l'unica soluzione per la risoluzione della situazione pandemica. Allora, la popolazione italiana era spaccata fra coloro che, riportando priorità di diverso tipo (dovute ad età, tipo di lavoro, rischi dovuti al virus), avevano potuto ricevere la prima dose del vaccino, coloro che, non rientrando all'interno delle categorie soprascritte, non avevano ancora avuto la possibilità di ricevere il vaccino, e, infine, coloro che, pur rientrando nelle categorie in questione, si erano rifiutati di riceverlo. Da metà aprile erano iniziate in Italia le vaccinazioni della fascia over 80, persone ad elevata fragilità, di persone di età compresa tra i 70 e i 79 anni e, a seguire, di quelle di età compresa tra i 60 e i 69 anni. Parallelamente a queste categorie, veniva completata la vaccinazione di tutto il personale sanitario e socio-sanitario e, a seguire, le altre categorie considerate prioritarie dal Piano nazionale, seguendo un ordine indicato: personale scolastico e universitario, forze armate, di polizia e del soccorso pubblico, servizi penitenziari, altre comunità residenziali socio-sanitarie, civili, religiose, ecc.

Naturale è interrogarsi riguardo a cosa sia dovuta la riluttanza e la conseguente negazione dell'offerta vaccinale che ha interessato e continua a interessare buona parte della popolazione. Nello studio di questo fenomeno, la volontà di ricevere il vaccino rappresenta una variabile dipendente, a monte della quale possiamo ipotizzare la presenza di un fattore psicologico, come la fiducia nelle istituzioni sanitarie (variabile indipendente). Ammettiamo quindi di poter indagare l'effetto di quest'ultima variabile su un campione rappresentativo della popolazione italiana per genere, età, area di residenza ed educazione, ma profondamente sbilanciato riguardo all'offerta vaccinale, a causa del piano nazionale stesso di distribuzione del vaccino. Pertanto, associando ad Y la volontà di ricevere il vaccino e

ad X la fiducia nelle istituzioni sanitarie, è possibile ipotizzare un rapporto causale fra le due variabili, rappresentabile tramite una freccia condotta da X a Y . Questa relazione, però, risentirebbe della disomogeneità della popolazione e quindi del campione: siccome la distribuzione del vaccino non è dipesa (o meglio, non è stata manipolata) dai ricercatori, si tratta necessariamente di uno studio osservazionale, anziché randomizzato. Infatti, nel DAG che raffigura l'ipotesi (si veda il Grafico 4 in Appendice A), compaiono anche le covariate che determinano l'offerta del vaccino (ovvero età, genere, lavoro), Z_1 e le variabili confondenti che non dipendono dall'offerta del vaccino (ad esempio, lo scetticismo generale sui vaccini), Z_2 . Per proseguire in questo studio, risulta quindi necessario che il campione utilizzato venga aggiustato in modo tale da non risentire di questa differenza.

2.3 Studi quasi sperimentali

Nel caso degli studi osservazionali, la soluzione consiste nel ricreare una condizione che si avvicini il più possibile alla randomizzazione, quindi ad impostare un disegno detto quasi sperimentale. Questa condizione può essere raggiunta ricorrendo a metodi di aggiustamento delle variabili confondenti, che possono implicare

1. l'utilizzo di un modello, ad esempio di regressione lineare, che integri l'influenza delle variabili confondenti tramite la loro addizione nel modello;
2. l'utilizzo di tecniche basate sul propensity score.

2.3.1 Aggiustamento additivo tramite regressione lineare

L'aggiustamento delle covariate tramite la regressione lineare è preferibile quando si lavora con campioni molto ampi, siccome l'elevata numerosità stessa comporta la compensazione dei bias dovuti alla selezione dei soggetti (Li et al., 2014). Anche se la regressione lineare verrà trattata più approfonditamente nel capitolo successivo, possiamo anticipare la sua utilità in qualità di metodo per stimare il valore atteso di una variabile dipendente, condizionato dai valori di altre variabili indipendenti. Un modello di regressione lineare, infatti, può essere rappresentato come segue

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

dove

- i rappresenta la i -esima osservazione;
- Y_i è la variabile dipendente, precedentemente chiamata anche outcome, che svolge il ruolo di predittore;

- X_i è la variabile indipendente, precedentemente chiamata anche trattamento, che in questo caso svolge il ruolo di regressore;
- $\beta_0 + \beta_1 X$ è l'equazione della retta che descrive il rapporto tra le variabili, detta retta di regressione;
- β_0 è l'intercetta della retta, quindi il valore che assume Y quando X vale 0;
- β_1 è il coefficiente angolare della retta, quindi il valore di cui aumenta Y ad ogni aumento unitario di X ;
- ϵ_i è l'errore statistico con distribuzione normale $\epsilon_i \sim N(0, \sigma^2)$.

Affinché la regressione lineare possa essere in grado di fornire un aggiustamento, deve comprendere nel modello anche le altre variabili contemplate nello studio, ovvero le covariate. In tal senso, possono essere integrate in maniera additiva all'interno del modello precedente, che ora rappresenta una regressione multipla:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \dots + \beta_k X_{ki} + \epsilon_i$$

dove $X_{1i}, X_{2i} + \dots + X_{ki}$ sono le i -esime osservazioni per ciascuno dei k regressori, appunto la variabile indipendente e le covariate.

Nell'impostare il modello di regressione multipla serve attenzione, siccome ne dipende la validità dei risultati: è necessario che siano presenti tutte le covariate che influenzano la relazione indagata per aggiustare il valore dell'outcome, come desiderato, sfruttando il cosiddetto fenomeno di indipendenza condizionale. Se si verifica questa condizione, l'outcome subisce una variazione dovuta al trattamento a cui si va a sommare quella dovuta alle altre variabili in gioco: in sostanza, viene fatto 'regredire' all'effetto combinato di variabile trattamento e covariate.

Z come moderatore

Riprendendo il caso già trattato risalente alla rivista *Annual review* (D'Onofrio et al., 2020), ricordiamo che il problema dell'influenza di Z , variabile confondente, era stato risolto, affermando che l'assegnazione casuale dei soggetti rappresentava un modo per condizionare Z e quindi poter analizzare l'effetto causale di X su Y . Sapendo che negli studi osservazionali la randomizzazione non è possibile, un altro modo per condizionare Z in questi casi consiste nel fissarla ad un valore z . In questo modo, si ottiene una curva che esprime la probabilità che si verifichi un determinato valore di Y a partire da un valore di X in funzione del valore z della confondente. A questo punto la funzione probabilistica diventa una funzione logica dove z ha un valore additivo.

Immaginiamo di voler indagare se un maggior numero di figli (variabile indipendente, X) causa un aumento dello stress nei genitori (variabile dipendente, Y). Nell'analizzare questo

effetto causale, non possiamo prescindere dalla presenza di alcuni elementi fondamentali all'interno del quadro, come ad esempio la coesione familiare. Possiamo ragionevolmente ipotizzare che un maggior numero di figli può comportare, oltre che un aumento di stress genitoriale, un aumento della coesione familiare, che, a sua volta, influisce positivamente sull'umore dei genitori. Si delineano quindi due scenari: nel primo, la coesione familiare (variabile confondente Z) modera l'influenza del numero di figli sullo stress genitoriale, mentre nella seconda si ha solo l'effetto diretto del numero di figli sulla variabile dipendente. Questi scenari si possono rappresentare graficamente come due vie che si dipartono da X e si congiungono su Y (si veda il Grafico 5.1 in Appendice A).

A questo punto, possiamo quindi fissare il livello di coesione familiare ad un certo valore z e considerare l'effetto di X su Y (si veda il Grafico 5.2). Una possibile soluzione per proseguire nell'indagine, è rappresentata, come detto, dall'utilizzo della regressione lineare per costruire un modello in grado di descrivere il fenomeno e stimare Y , dove si opera un aggiustamento del valore outcome tramite l'addizione delle covariate.

$$Y = \alpha + \beta X + \gamma z + \gamma\phi Xz + \epsilon$$

Z come mediatore

Z però può comparire in punti diversi del DAG, in base alle condizioni che si prospettano in ciascuno studio, e può essere sempre integrata nel modello che le descrive, ad esempio tramite il metodo additivo, se si tratta di un modello di regressione lineare. Nel caso presentato sopra, la variabile confondente Z svolge il ruolo di moderatore, ma, nell'impostare graficamente l'ipotesi di ricerca, si può rappresentare Z anche come mediatore: questo è il caso che si verifica quando si ipotizza un unico percorso che connette X ed Y in mezzo al quale si trova Z , che media completamente l'influenza della prima variabile sulla seconda (si veda il Grafico 6.1 in Appendice A).

Ipotizziamo questa volta che una famiglia coesa, per quanto numerosa, comporti più difficilmente stress per i genitori, rispetto ad una famiglia in cui si verificano, ad esempio, frequenti litigi. In questo caso, Z funge da mediatore, perché, in virtù di questa diversa ipotesi, si trova interamente coinvolta nell'influenza da parte di X su Y . Anche in questo caso, il modello di regressione lineare è in grado di descrivere la situazione delineata, mettendo a sistema l'effetto della variabile indipendente prima e dopo l'influenza di Z (si vedano il Grafico 6.2 e 6.3).

$$\begin{cases} Z = \beta_1 X + \epsilon \\ Y = \gamma Z + \epsilon \end{cases} \Rightarrow Y = \gamma\beta_1 X + \epsilon$$

2.3.2 Matching e propensity score

Le tecniche di aggiustamento alternative a quella precedentemente illustrata sono preferibili quando si ha a che fare con campioni composti da pochi soggetti e si è interessati allo studio di poche variabili.

Si inizia raccogliendo delle particolari informazioni (le principali variabili confondenti) per ciascun soggetto: stando alle disposizioni dell'American Psychological Association (APA), convenzionalmente si tratta di età, genere, educazione, status socio-economico ed, eventualmente, nazionalità. Alle confondenti individuate dall'APA si aggiungono poi le variabili di cui uno studio deve tenere necessariamente conto: ad esempio, in uno studio che indaga l'effetto di una dieta sul peso degli individui, sarà necessario raccogliere informazioni (oltre che sulle variabili APA) anche sullo stile di vita dei soggetti, sulla presenza di eventuali malattie croniche e così via.

Successivamente si può usare una delle seguenti tecniche, che agiscono direttamente sul bilanciamento delle categorie menzionate all'interno dei campioni (sempre con l'obiettivo di avvicinarsi il più possibile alla randomizzazione).

Una prima opzione è il matching, secondo Rubin (Rubin, 2006), l'approccio preferenziale negli studi osservazionali (West & Thoemmes, 2010). Associando ad ogni soggetto trattato un soggetto di controllo identico al primo (secondo le variabili confondenti) alla condizione basale, si può stimare un effetto causale a livello individuale che non risente di bias.

Pertanto, l'effetto complessivo del trattamento è dato dalla differenza fra la media degli outcome di ogni unità trattata e la media degli outcome delle rispettive unità di controllo (Foster, 2010).

La singola covariata misurata (COV) non può essere che l'unica responsabile dell'assegnazione sperimentale (West & Thoemmes, 2010). Creando dei match fra un partecipante del gruppo trattamento con uno del gruppo di controllo in base alla covariata a condizione basale, si può ottenere una stima non distorta dell'effetto causale a livello individuale. Più in generale, se COV ha la stessa distribuzione nei due gruppi di soggetti, non può essere la causa dell'effetto del trattamento osservato. Ottenere questa distribuzione omogenea di COV tra i due gruppi permette di non influenzare l'assegnazione. Questa procedura riproduce l'equilibrio raggiunto per specifiche covariate attraverso la randomizzazione avendo a disposizione ampi campioni.

Un'estensione del matching è la stratificazione, siccome permette di creare strati del campione composti da unità simili dal punto di vista delle confondenti. Ne deriva che, più sono numerosi gli strati creati, più simili fra loro risultano le unità di ciascuno strato (Foster, 2010).

Storicamente, il principale ostacolo nell'usare con successo il matching è stata la necessità di creare match per diverse covariate. Anche avendo ampi campioni, è stato difficile identificare match accurati quando erano coinvolte più variabili (West & Thoemmes, 2010).

D'altro canto, lo sviluppo dei metodi di propensity score (PS) fornisce una teoria statistica in grado di estendere i benefici del matching (e non solo) avendo dalle poche alle tante covariate. Il PS, infatti, consiste nella stima della probabilità che ciascun soggetto possiede di ricevere il trattamento sulla base delle covariate che lo contraddistinguono. Tale stima viene quindi applicata in diversi contesti (alcuni dei quali presentano l'unione fra PS e matching) per rendere i campioni equilibrati.

3 Applicazione pratica

Come premesso nei paragrafi introduttivi, gli studi osservazionali costituiscono spesso l'unica opzione percorribile in molti ambiti di ricerca della psicologia che intendono rispondere a quesiti causali, siccome la manipolazione delle variabili indipendenti (ad esempio il consumo di sostanze, come quelle alcoliche) non risulterebbe etico e tanto meno possibile da condurre, praticando l'assegnazione casuale (Harder et al., 2010). In ogni caso, un'altra complicazione degli studi osservazionali sorge sia dalla selezione non probabilistica del campione sia dall'effetto esercitato da variabili confondenti. A questo proposito, l'utilizzo di tecniche che permettono di arginare le distorsioni dovute alle disomogeneità tra i soggetti all'interno di campioni non probabilistici risulta di vitale importanza quando si affronta un'indagine causale.

Possiamo pertanto proseguire l'esposizione, trattando più nello specifico in cosa consistono le due tecniche di correzione dell'outcome illustrate nel precedente capitolo - l'aggiustamento additivo delle covariate e il propensity score, confrontandone i risultati in seguito all'applicazione pratica avendo una reale ipotesi causale.

3.1 Metodo statistico

L'obiettivo finale di ogni analisi causale consiste in un'accurata stima del vero effetto causale, non semplicemente nel bilanciamento delle covariate misurate. Come esposto precedentemente, un'assunzione chiave per stimare il reale effetto causale è quello dell'ignorabilità forte, secondo cui, una volta ottenuto il condizionamento delle covariate, l'analisi non risente di altri effetti confondenti dovuti all'associazione tra trattamento e outcome. In altre parole, le covariate misurate incluse nella stima dell'outcome (in maniera additiva o tramite l'utilizzo di tecniche basate sul propensity score) permettendo di aggiustare le stime tenendo conto di possibili bias nascosti.

3.1.1 Regressione logistica

Come riportato in precedenza, la regressione lineare può essere utilizzata per correggere l'equazione che descrive l'outcome in funzione della variabile indipendente tramite la sommatoria delle covariate. Oltre alla regressione lineare, però, possono essere impiegati in tal senso anche altri modelli che rientrano nella categoria dei cosiddetti modelli lineari generalizzati (o, in inglese, *generalized linear model* - GLM), trattandosi di una generalizzazione dei modelli di regressione lineare. Per definizione, ogni GLM presenta

- una funzione $f()$, che descrive la distribuzione casuale della variabile dipendente;

- un predittore lineare η_i , il cui valore risulta dalla funzione lineare dei regressori messi assieme in maniera addittiva

$$\eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik};$$

- una funzione di collegamento $g()$, così chiamata perché connette il predittore lineare η con il valore predetto della variabile dipendente, $\mathbf{E}[Y_i]$, nella maniera seguente

$$\mathbf{E}[Y_i] = g(\mu_i) = \eta_i$$

Nel capitolo precedente è stata utilizzata la regressione lineare multipla come modello per rappresentare delle situazioni d'esempio. Risulta invece più indicato l'utilizzo di altri modelli nei casi, ricorrenti in psicologia, in cui si ha a che fare con variabili di risposta dicotomiche, trattanti quindi la presenza o l'assenza di una caratteristica in ciascuna delle unità statistiche: ad esempio il genere, una malattia, la presenza di figli, il fatto di fumare o meno. Tali variabili sono codificate in termini di 0 e 1 associando a questi valori le due condizioni in cui si possono presentare: ad esempio, se la variabile dipendente rappresenta l'abitudine di fumare, l'outcome associato ad un soggetto fumatore sarà pari a 1, mentre quello di un non fumatore sarà pari a 0. Un modello in grado di rappresentare questa condizione binomiale è il modello di regressione logistica, che permette di predire la probabilità π_i , dato un valore di X , x_i , che ha la variabile dipendente di essere pari a 1, $Pr(Y = 1)$. Ad esempio, la probabilità che una persona fumi si può calcolare come

$$\pi_i = Pr(Y = 1|X = x_i)$$

Proprio perché le probabilità sono per definizione limitate ad un intervallo di valori compreso tra 0 e 1, l'utilizzo di un modello di regressione lineare non sarebbe appropriato. Infatti, per ottenere π_i dobbiamo applicare al predittore η_i la funzione $\lambda(z)$, chiamata logistica. La funzione logistica non è altro che una funzione inversa della funzione di collegamento (che è appunto invertibile), ovvero $g^{-1}()$, in grado di garantire che il valore di π_i rimanga compreso tra 0 e 1 $\pi_i = \lambda(\eta_i)$. Il modello che otteniamo è detto lineare logistico e $\lambda(z)$ è la funzione canonica di collegamento per la regressione binomiale.

Avendo dato una definizione teorica della regressione logistica, possiamo ora spiegare in che modo può essere impiegata come tecnica di aggiustamento del campione negli studi osservazionali: è possibile aggiustare la probabilità, restituita dalla funzione, che ogni soggetto ha di manifestare la variabile dipendente tramite l'aggiunta delle covariate all'interno della funzione stessa. Ad esempio, la probabilità dell'abitudine al fumo

$$\pi_i = Pr(Y = 1|X = x_i) = \lambda(\beta_0 + \beta_1 X_i)$$

può essere corretta, sommando i valori che presenta il soggetto per tutte le covariate contemplate nello studio

$$\pi_i = Pr(Y = 1|X = x_i) = \lambda(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \dots + \beta_k X_{ki})$$

3.1.2. Propensity score (PS)

Passando invece a descrivere l'utilizzo delle tecniche basate sull'utilizzo del propensity score, possiamo distinguere due fasi: stima e applicazione.

Il PS stesso esprime la stima della probabilità di osservare un dato valore della variabile indipendente, che può essere un intervento o un trattamento. Scegliamo di chiamarla trattamento in linea con il vocabolario usato nella letteratura che riguarda l'inferenza causale. Di conseguenza, il PS predice la probabilità che ogni individuo ha di ricevere il trattamento, rappresentata tramite un modello che vede il trattamento ricevuto come una funzione delle covariate misurate. Queste covariate vengono scelte sulla base della loro potenziale capacità empirica o teorica di confondere la relazione tra trattamento e outcome. Se questo effetto confondente viene inibito, condizionando, ad esempio, tramite PS, le covariate, allora è possibile ottenere conclusioni causali: questo è l'assunto di ignorabilità forte.

La seconda fase, quella di applicazione, riguarda l'uso delle stime ottenute precedentemente con il fine di rendere i gruppi di soggetti trattati e non trattati comparabili, per fare in modo, in altre parole, che lo studio osservazionale si avvicini il più possibile ad uno randomizzato.

Possiamo ora approfondire la trattazione di queste due fasi, seguendo lo studio di Harder e colleghi (Harder et al., 2010) dove viene riportato con precisione quando e in che modo è possibile utilizzare i vari metodi di stima del PS in combinazione con i metodi che ne permettono l'applicazione, valutando la capacità delle varie combinazioni di bilanciare le covariate osservate.

Stima del propensity score

Come afferma la letteratura citata, possiamo prendere in considerazione tre metodi, due parametrici e uno non parametrico, per stimare il PS. Qui approfondiamo solo uno dei due metodi parametrici, la regressione logistica multivariata, siccome verrà impiegata negli esempi a venire. Questo approccio permette di ottenere la stima della probabilità che ogni soggetto riporti o non riporti la caratteristica oggetto di interesse (ovvero la variabile indipendente dello studio, in genere la X) considerando nel modello l'inclusione di tutte le covariate nello studio. Harder e colleghi, ad esempio, hanno calcolato la probabilità che ogni soggetto sperimentale ha di riportare abuso o dipendenza da cannabis (che rappresenta quindi la condizione di trattamento dello studio) sulla base delle condizioni di ognuno a livello demografico, socio-economico e così via. Così facendo, si ottengono due distribuzioni di valori: la probabilità di fumare per i soggetti trattati ($T=1$) e la probabilità che avrebbero di non fumare i soggetti non trattati ($T=0$). In tal senso, le due distribuzioni di probabilità sono rappresentate da due curve che possono essere completamente sovrapposte, per nulla sovrapposte o aventi qualche valore in comune. Nel primo caso, i due gruppi si presentano identici dal punto di vista della distribuzione delle covariate ed è ciò che avviene il caso di uno studio randomizzato. La condizione che si delinea nel secondo caso fa capire che

ci sono troppe differenze pre-esistenti tra individui trattati e non, per cui non avrebbe senso proseguire con l'analisi causale. Il terzo caso, invece, indica la presenza di una porzione di area in comune tra le due distribuzioni ed è la condizione che si presenta più frequentemente, essendo propria degli studi osservazionali. Successivamente descriviamo i metodi di applicazione del PS che mirano ad avvicinare il più possibile il terzo scenario al primo, quindi a rendere uno studio osservazionale quasi sperimentale.

Applicazione del propensity score

Dopo aver stimato le probabilità con il PS, il passo seguente sta nella applicazione ad una regressione con l'obiettivo di casualità. Le probabilità stimate vengono impiegate per pesare i soggetti trattati e non, in modo da rendere casuale il risultato della regressione con l'outcome di interesse. Harder e colleghi hanno passato in rassegna cinque metodi per applicare il propensity score, sottolineando che il compito di ogni ricercatore consiste anche nel decidere quale di questi sia il più appropriato da utilizzare per rispondere al proprio quesito di ricerca, motivandone la scelta. Si tratta di metodi di applicazione diversi tra di loro: i primi tre stimano l'ATT, mentre gli ultimi due stimano l'ATE. Pertanto, la stima del PS può essere impiegata

- nel caso del matching 1:1, per creare delle coppie composte da un'unità trattata e una di controllo con simile PS. Questo tipo di applicazione non risulta il più funzionale perché, da una parte, vengono scartate tutte le unità di controllo che non si uniscono a quelle trattate e, dall'altra, i controlli non scartati - e quindi impiegati nelle coppie - sono spesso degli outlier;
- nel caso del full matching, per creare set di soggetti con PS simili fra cui ci sia almeno un individuo trattato e almeno uno di controllo;
- nel caso del *weighting by the odds*, per assegnare agli individui trattati un peso pari ad 1 e ai controlli un peso pari al loro PS (ρ_i) convertito secondo la scala della probabilità che è pari a $\rho_i/(1 - \rho_i)$;
- nel caso dell'*inverse probability of treatment weights* (IPTW), per pesare ogni individuo con la probabilità inversa di ricevere il trattamento che ha realmente ricevuto, quindi per assegnare agli individui trattati un IPTW pari a $1/\rho_i$ e agli individui di controllo $1/(1 - \rho_i)$;
- nel caso della subclassificazione, per dividere i soggetti entro gruppi (o sottoclassi) in base a valori simili di PS.

Dopo la stima e l'applicazione, il terzo e ultimo passo consiste nell'usare il PS per impostare un modello che sia in grado di restituire delle stime del risultato dello studio più

accurate. Quindi, come anticipato, possiamo riproporre le fasi descritte nel presente paragrafo, facendo questa volta un utilizzo pratico delle tecniche di stima e applicazione del PS più funzionali alla risoluzione di una concreta ipotesi di ricerca.

3.2 Effetto del fumo di sigaretta sul rischio di decadimento cognitivo

Molti degli studi che propongono le proprie analisi come esempi di applicazione delle tecniche di correzione dell'outcome si concentrano su ambiti legati al decadimento cognitivo e ai suoi fattori di rischio, proprio perché questi ultimi rappresentano un punto critico di cui tenere conto durante le analisi. A questo proposito, troviamo in letteratura una testimonianza di come l'utilizzo di tecniche di correzione del bias dovuto all'attrito selettivo possa influenzare notevolmente i risultati di una ricerca. L'attrito selettivo rappresenta, infatti, la tendenza di alcuni soggetti ad avere maggiori probabilità di abbandonare uno studio rispetto ad altri ed è legata alla mortalità selettiva (quindi al verificarsi di una morte prematura da parte di soggetti svantaggiati rispetto ad una popolazione di appartenenza di età simile), ma anche all'abbandono dello studio a causa di ulteriori motivi.

Secondo quanto esposto da Weuve e colleghi (Weuve et al., 2012), da analisi non aggiustate era risultato che il punteggio di decadimento cognitivo dei fumatori diminuiva di 0.11 unità standard per decade più rapidamente rispetto quanto avveniva in coloro che non sono mai stati fumatori, ma anche che, in seguito ad aggiustamento (avendo quindi tenuto conto dell'attrito selettivo), il tasso di decadimento cognitivo nel corso di 10 anni era stimato avvenire di 0.20 unità più rapidamente per i fumatori rispetto che per i non fumatori. In altre parole, era stato possibile individuare il risultato finale dello studio solo grazie alla correzione del bias: in un arco di tempo di 10 anni, i fumatori presentavano un decadimento cognitivo che i non fumatori presentavano in 12.5 anni. In conclusione, le stime dell'effetto del fumo sul decadimento cognitivo rischiano di essere sottostimate a causa dell'attrito selettivo e, in tal senso, risulta fondamentale tenerne conto nelle analisi per ottenere dei risultati più precisi.

Per condurre uno studio mira ad evidenziare gli aspetti causali di un'abitudine nel tempo, è necessario un approccio longitudinale. Ad esempio, nello studio citato, i due gruppi messi a confronto erano rappresentati da fumatori al momento in cui è iniziato lo studio (baseline) e da soggetti che non avevano mai fumato (fino alla baseline). Il termine di confronto tra i due gruppi erano i risultati ad una serie di test che misuravano le funzioni cognitive di base, somministrati per un massimo di 5 volte ad intervalli costanti di 3 anni (per un totale di circa 12 anni di follow-up).

Lo studio di Weuve e colleghi ipotizzava che, paragonando coloro che erano considerati fumatori alla baseline e coloro che non sono mai stati fumatori, i primi sarebbero stati soggetti a un decadimento cognitivo più rapido rispetto ai secondi nei 12 anni di follow-up rispetto alla baseline, ma anche che tale associazione sarebbe risultata più forte dall'analisi che teneva conto dell'attrito selettivo. A tal proposito, possiamo nuovamente testare

l'associazione tra le due variabili - l'essere fumatore e il decadimento cognitivo, ipotizzando questa volta un effetto causale della prima sulla seconda, cosa che è possibile verificare tenendo conto della probabilità che ogni soggetto ha di essere fumatore, quindi una volta aggiustato per una serie di covariate.

3.2.1 Ipotesi

Come anticipato, per osservare gli effetti del fumo sul decadimento cognitivo è necessario un approccio longitudinale, quindi disporre di dati che sono stati raccolti sugli stessi soggetti a distanza di anni. In tal senso, possiamo utilizzare i dati raccolti in una consistente indagine, la *Survey of Health, Ageing and Retirement in Europe* (SHARE). Si tratta di un'infrastruttura di ricerca che raccoglie dati in 28 paesi europei e Israele e studia gli effetti delle abitudini sul piano sociale, della salute, economico e ambientale nel corso della vita. Tenendo conto che sono disponibili altrove dettagliate informazioni su questa indagine (Börsch-Supan et al., 2013), possiamo anticipare che gli individui eleggibili, scelti in base a metodi di selezione della probabilità, erano persone di 50 anni o più anziane che, assieme ai rispettivi partner (indipendentemente dalla loro età), hanno fornito le risposte dalle proprie case tramite colloqui al computer (CAPI), assistiti dal personale. SHARE mette a disposizione dati raccolti sugli stessi soggetti a distanza di due anni (le cosiddette Wave) a partire dal 2004, con la Wave 1, fino alla più recente Wave 8, del 2020.

All'interno della nostra indagine causale, l'essere fumatore alla baseline ("smoking") è considerato come variabile indipendente, mentre la variabile dipendente è rappresentata dal decadimento cognitivo, associato alla differenza in termini di fluenza verbale ("fluency") nel corso di una decade, quindi tra il punteggio ottenuto alla Wave 5, del 2013 (Börsch-Supan, n.d.-b), rispetto al punteggio alla Wave 1, del 2004 (Börsch-Supan, n.d.-a), che identifichiamo come baseline. Di conseguenza, i soggetti di cui disponiamo andranno a comporre il gruppo sperimentale, se nel 2004 hanno risposto in modo affermativo all'item in cui veniva chiesto "Adesso fuma?", oppure il gruppo di controllo in caso di risposta negativa. Di entrambi i gruppi sperimentali siamo interessati a capire se il decadimento cognitivo avviene in maniera diversa, quindi a confrontare i punteggi da loro ottenuti nell'item dedicato alla fluenza verbale nel quale veniva chiesto ai soggetti "Ora vorrei che mi indicasse tutti gli animali a cui riesce a pensare. Ha un minuto a sua disposizione" (veniva assegnato 1 punto per animale).

Inoltre, sono state selezionate anche delle variabili covariate, i cui valori (ovvero i punteggi alla baseline) sono stati inseriti nelle analisi al fine di bilanciare il campione, che, per la natura stessa dall'indagine SHARE, non poteva essere creato a partire da una randomizzazione. Pertanto, le variabili selezionate proprio perché potrebbero condizionare la probabilità per una persona di fumare, sono

- età,

- genere,
- anni di educazione,
- titoli di studio secondo l'ISCED (1997),
- BMI,
- stato di salute,
- mobilità (funzione delle braccia e limitazioni nei movimenti fini),
- limitazioni fisiche nelle attività quotidiane (GALI),
- numero di malattie croniche,
- livello di depressione valutato secondo la scala EURO-D,
- numero di figli,
- stato civile,
- stato lavorativo,
- stato economico,
- costo della propria abitazione,
- bere più di due bicchieri di alcolici quasi ogni giorno,
- inattività fisica.

Anche se SHARE offre dati raccolti in diversi paesi, abbiamo preso in considerazione solo gli individui italiani, le cui caratteristiche sono presentate attraverso le statistiche descrittive, raccolte nelle Tabelle 1 e 2 dell'Appendice B. Ora possiamo proseguire con l'analisi dati all'interno di questo studio quasi sperimentale, basato sull'ipotesi tale per cui coloro che fumano alla baseline presenteranno un decadimento cognitivo più accelerato (quindi un peggioramento in termini di fluenza verbale) nel corso di 10 anni, rispetto a coloro che alla baseline hanno dichiarato di non fumare (si veda il Grafico 7 in Appendice A).

3.2.2 Risultati della regressione logistica con e senza propensity score

Per verificare l'ipotesi, utilizziamo un modello di regressione logistica in due modi diversi: in un caso, impostiamo un modello che predice semplicemente l'influenza della variabile indipendente sulla variabile dipendente, corretta tramite la sommatoria delle covariate; mentre, in un secondo caso, rendiamo più preciso il modello, inserendovi anche la correzione tramite propensity score, ottenuta tramite il metodo IPTW, illustrato in precedenza. Come si può notare dallo script in Appendice B, l'unica variazione apportata alle linee di codice

che riportano la seconda versione del modello di regressione logistica riguarda l'aggiunta dei pesi creati tramite PS.

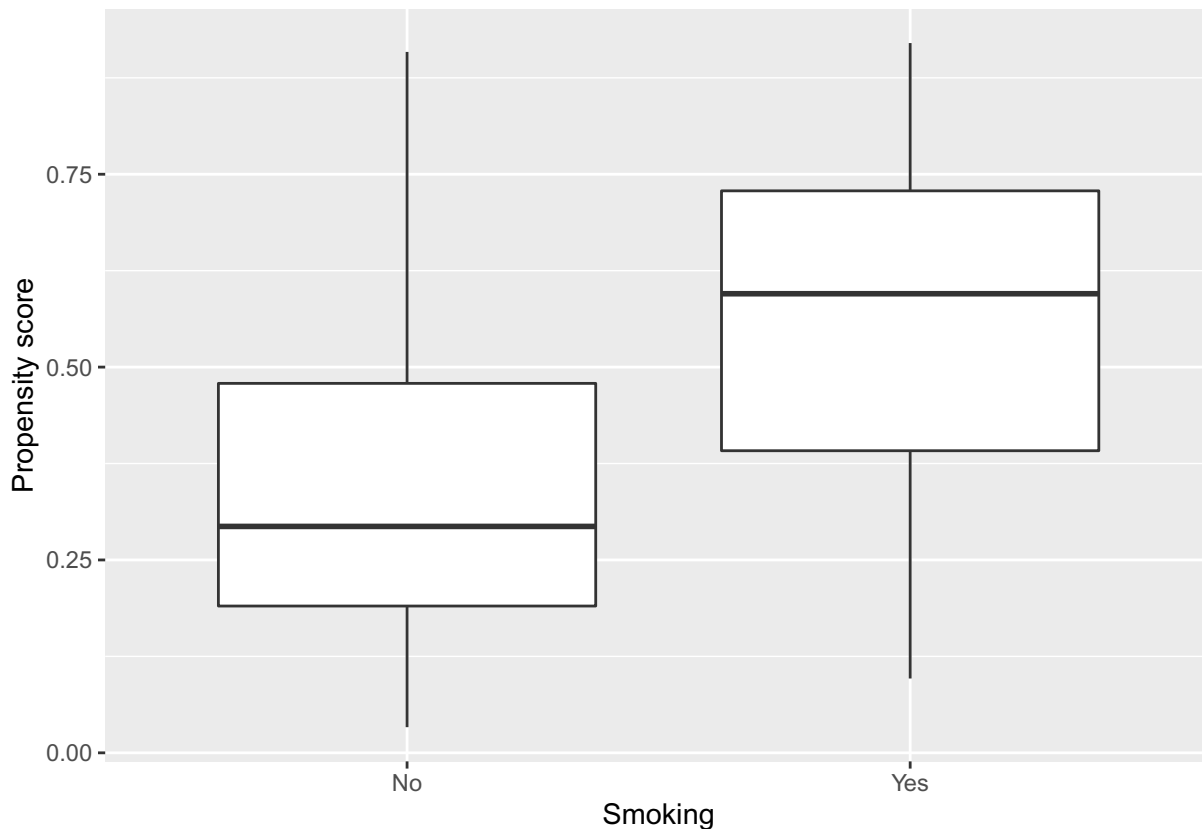


Grafico 8. Propensity score stimate per fumatori e non fumatori

Nella Tabella 3 in Appendice B sono presentati i risultati espressi in termini di odds ratio (OR), ovvero tramite il rapporto tra la frequenza di decadimento cognitivo nel gruppo di fumatori e la frequenza della stessa nel gruppo di non fumatori. Gli OR permettono di osservare con chiarezza i risultati: per valori di odds ratio maggiori ad 1, significa che l'evento osservato (il decadimento cognitivo) si verifica più spesso nel gruppo trattato (i fumatori) rispetto al gruppo di controllo (i non fumatori), viceversa per i valori minori ad 1. In altre parole, la variabile indipendente induce sulla variabile dipendente un effetto che,

- se $OR > 1$, è di rischio;
- se $OR < 1$, è protettivo;
- se $OR \sim 1$, non è presente.

3.2.3 Confronto dei risultati

Osservando la Tabella 3, possiamo notare che si ottengono risultati differenti dal primo e dal secondo utilizzo del modello di regressione logistica. In particolare, facendo riferimento alla variabile indipendente di nostro interesse, emerge, dal GLM non corretto tramite PS, che l'effetto del fumo sul decadimento cognitivo non è significativo (condizione espressa dal p-value, in questo mancante proprio perché non significativo), mentre dal GLM corretto tramite PS, che il medesimo effetto è presente in modo significativo e, come suggerito dal valore di OR maggiore ad 1 (OR=1.05, con un intervallo di confidenza (CI) del 95%: 1.01, 1.08), è un effetto di rischio. Dal fatto che si ottengono risultati significativi solo grazie all'utilizzo del propensity score, capiamo quanto sia importante il ruolo dell'aggiustamento all'interno delle analisi. Nel caso di dati raccolti su fumatori soprattutto, risulta fondamentale avere quest'accortezza, siccome rappresentano una popolazione che tende ad evitare questo tipo di indagini, comportando quindi un bias di selezione negli studi che richiedono la loro partecipazione, proprio come è emerso dallo studio di Weuve e colleghi.

3.3 Discussione e limiti dello studio

Guardando quindi i risultati significativi emersi dalla correzione con PS, possiamo affermare che il fumo rappresenta un fattore di rischio per il decadimento cognitivo. In tal senso risulta possibile avanzare delle spiegazioni biologiche plausibili che riguardano i diversi tipi di effetti della variabile indipendente, sulla base anche di quanto presente in letteratura (Anstey et al., 2007).

Un limite di questo studio quasi sperimentale consiste nel fatto che, per quanto ci sforziamo di tenere conto di tutte le variabili in gioco nell'aggiustamento per le covariate, non saremo mai in grado di controllare per tutti i possibili confondenti, non potendo mai avere la certezza di essere impermeabili all'effetto delle variabili confondenti non misurate. Non è quindi improbabile che, in uno studio quasi sperimentale come quello presentato, venga violata l'assunzione di ignorabilità forte: bisogna ammettere che, nonostante l'applicazione delle tecniche di correzione, gli effetti del trattamento misurati possono comunque risentire di bias. Infatti, da tenere a mente è che, nell'affrontare quesiti di inferenza causale, perfino all'interno di framework corretti tramite PS - che abbiamo visto produrre risultati migliori - ottenere l'equilibrio delle covariate misurate potrebbe non essere sufficiente, se l'assunzione di ignorabilità forte non è soddisfatta. Proprio per questo motivo, i disegni sperimentali causali più utilizzati sono in genere basati su studi caso-controllo.

4 Conclusione

La causalità rappresenta una sfida per ogni ricercatore nell'ambito delle scienze sociali ed è quindi compito anche di coloro che lavorano nel campo della psicologia rispettare l'utilizzo di particolari tecniche e indirizzare degli accorgimenti in linea con questo obiettivo. Per riassumere quanto detto dall'inizio di questa esposizione, nell'impostare uno studio risulta necessario sia avere a disposizione un campione sufficientemente numeroso e rappresentativo della popolazione d'interesse, che saper sviluppare nel modo più preciso possibile la propria analisi. Avendo quindi formulato l'ipotesi sperimentale, è necessario ottenere una chiara panoramica delle variabili dipendenti e indipendenti, del loro rapporto, e delle influenze che su di esse esercitano altri fattori, aiutandosi visivamente attraverso un DAG: come può avvenire quando una o più variabili si rivelano essere dei collider, operarvi un condizionamento può rappresentare una soluzione per conservare l'aspetto causale dell'ipotesi (Foster, 2010). Il passaggio di trasposizione grafica di quanto ipotizzato si rivela dunque critico: la precisione nell'analisi guadagnata tenendo conto delle covariate è minacciata dai percorsi alternativi che si aprono inevitabilmente integrando le covariate stesse nell'ipotesi sperimentale e che, di conseguenza, rendono lo studio vulnerabile all'influenza di altri fattori. Possiamo ragionevolmente affermare che la scelta migliore è selezionare solo le covariate indispensabili, tenendo però presente che l'effetto confondente delle variabili non osservate rimane inevitabile.

Per poter proseguire in modo accurato con l'analisi concreta dei dati, si rivela fondamentale l'utilizzo delle tecniche come il propensity score, perché rappresenta un modo, talvolta l'unico, in grado di rilevare la significatività dei risultati, come nell'esempio riportato nel precedente capitolo. Il PS, infatti, si sta dimostrando un importante strumento statistico, fondamentale per rendere possibile una più precisa valutazione e correzione del disequilibrio tra gruppo di trattamento e di controllo, dovuto alle covariate misurate, negli studi osservazionali. Per quanto riguarda il suo utilizzo, tra le tante tecniche di PS, la scelta non dipende dai valori dell'outcome, né dalla dimensione o dalla significatività statistica dell'effetto del trattamento stimato dai modelli di regressione, quanto più dal dataset, dalla domanda di ricerca e dalla generalizzabilità desiderata dei risultati: una soluzione potrebbe essere confrontare i risultati ottenuti dalle diverse combinazioni di stime e metodi di applicazione del PS (Harder et al., 2010).

Anche se agire in questo modo non garantisce un'accuratezza definitiva all'analisi, siccome non sarà mai possibile tenere conto di tutte le variabili non osservabili, l'utilizzo del propensity score va incoraggiato in quanto oggetto di continue ricerche volte al suo progresso e miglioramento tali da renderlo un mezzo preciso e utile per soddisfare i criteri necessari alla causalità, soprattutto in psicologia, quindi una delle scienze in cui, per la natura stessa dei suoi quesiti di ricerca, è difficile poter impostare uno studio che abbia fin dall'inizio una struttura sperimentale, ma che d'altro canto ne riporta un'urgente necessità.

Ringraziamenti

This paper uses data from SHARE Waves 1, 2, 3, 4, 5, 6, 7, 8 and 9 (DOIs: 10.6103/SHARE.w1.800, 10.6103/SHARE.w2.800, 10.6103/SHARE.w3.800, 10.6103/SHARE.w4.800, 10.6103/SHARE.w5.800, 10.6103/SHARE.w6.800, 10.6103/SHARE.w7.800, 10.6103/SHARE.w8.800, 10.6103/SHARE.w8ca.800, 10.6103/SHARE.w9ca800), see Börsch-Supan et al. (2013) for methodological details.(1) The SHARE data collection has been funded by the European Commission, DG RTD through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982, DASISH: GA N°283646) and Horizon 2020 (SHARE-DEV3: GA N°676536, SHARE-COHESION: GA N°870628, SERISS: GA N°654221, SSHOC: GA N°823782, SHARE-COVID19: GA N°101015924) and by DG Employment, Social Affairs & Inclusion through VS 2015/0195, VS 2016/0135, VS 2018/0285, VS 2019/0332, and VS 2020/0313. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C, RAG052527A) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

Appendice A. Directed acyclic graphs

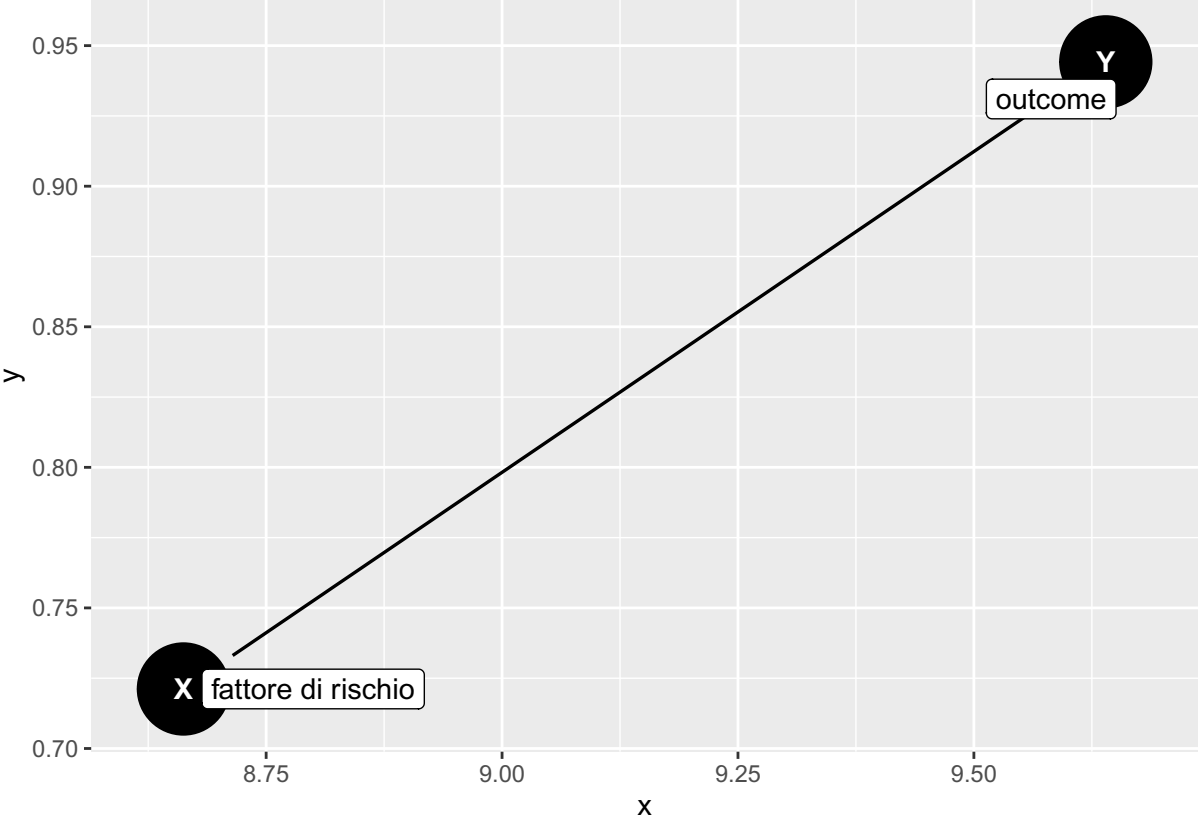


Grafico 1. Effetto causale di X su Y

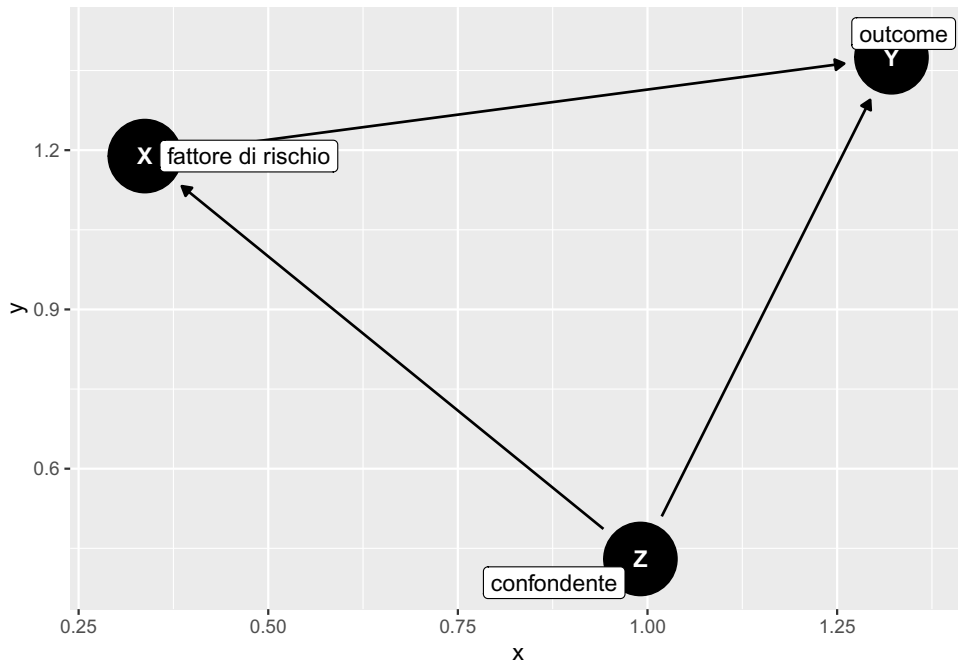


Grafico 2.1. Relazione spuria

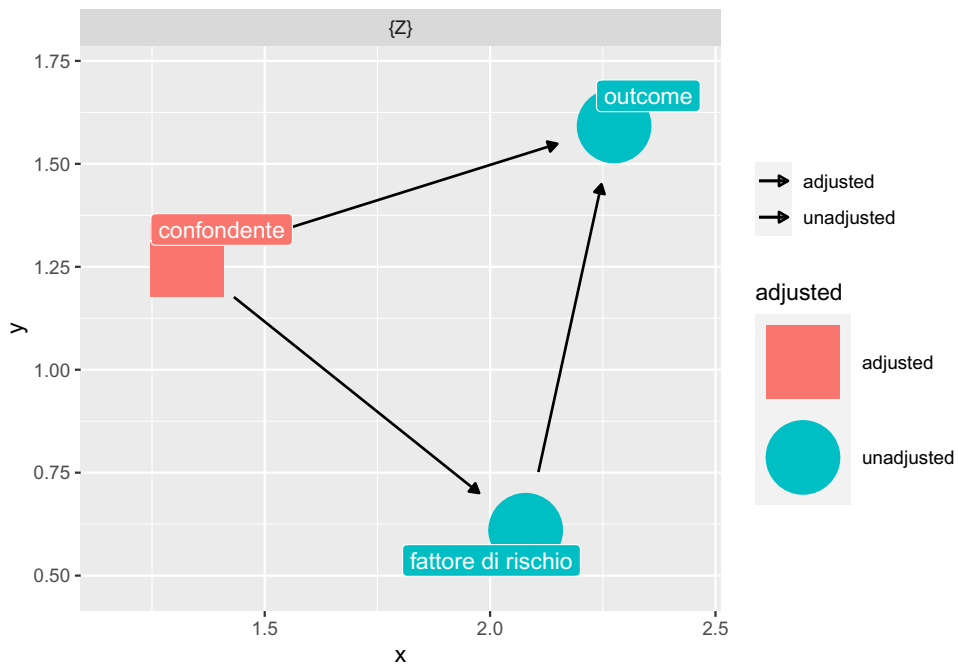


Grafico 2.2. Condizionamento della variabile confondente

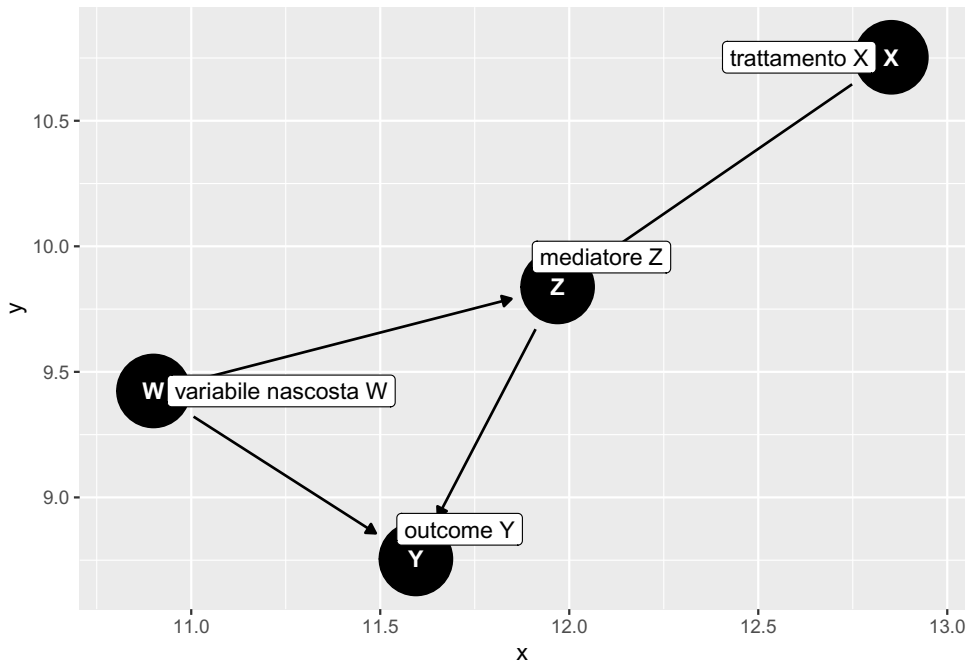


Grafico 3.1. Prima di condizionare il mediatore

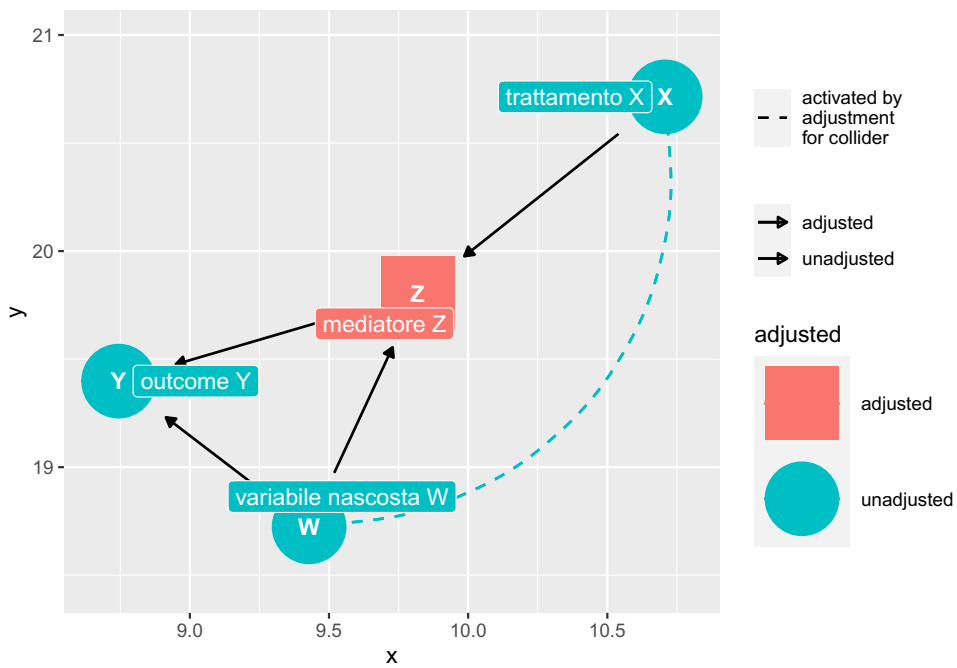


Grafico 3.2. Dopo aver condizionato il mediatore

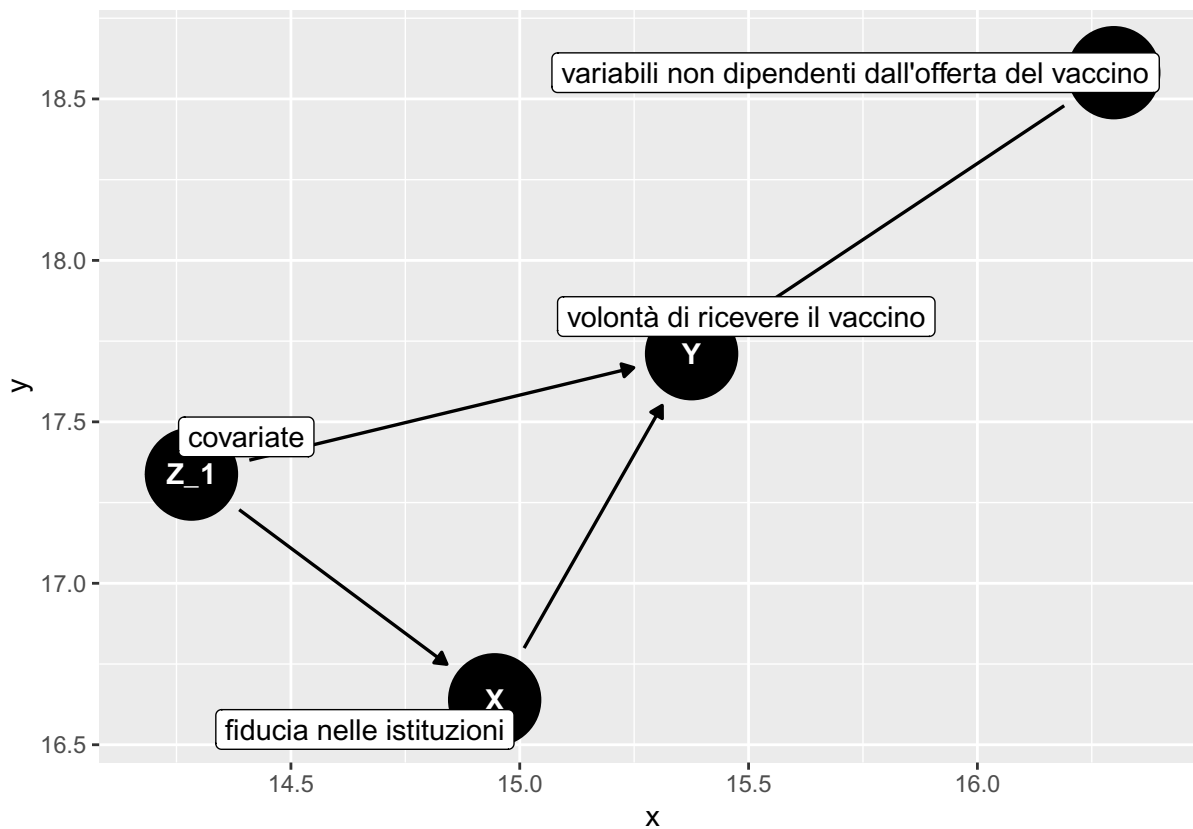


Grafico 4. Variabili coinvolte nella relazione causale tra fiducia nelle istituzioni e volontà di ricevere il vaccino

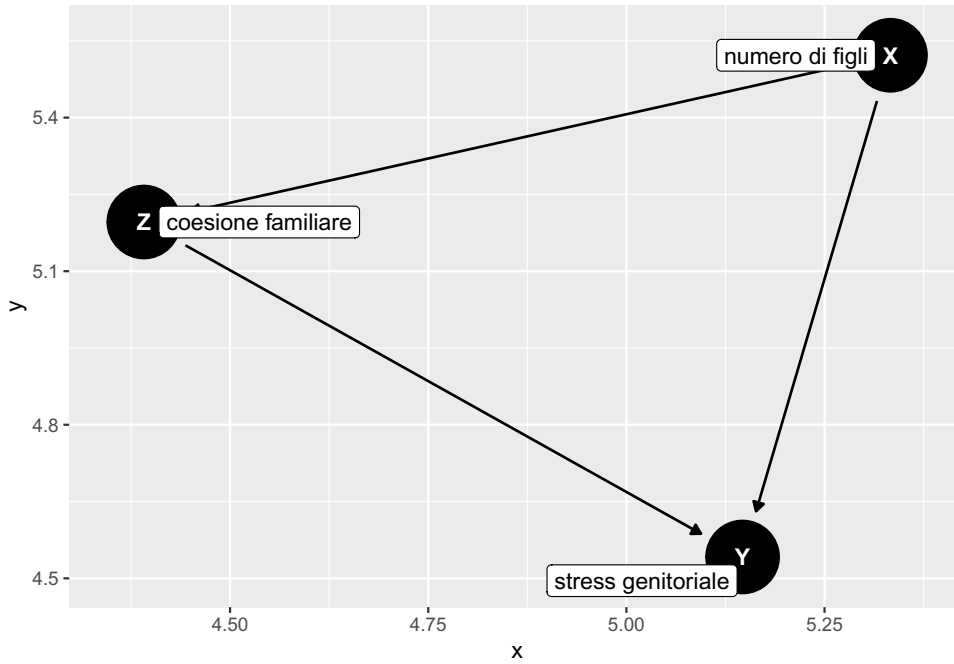


Grafico 5.1. Prima di condizionare il moderatore

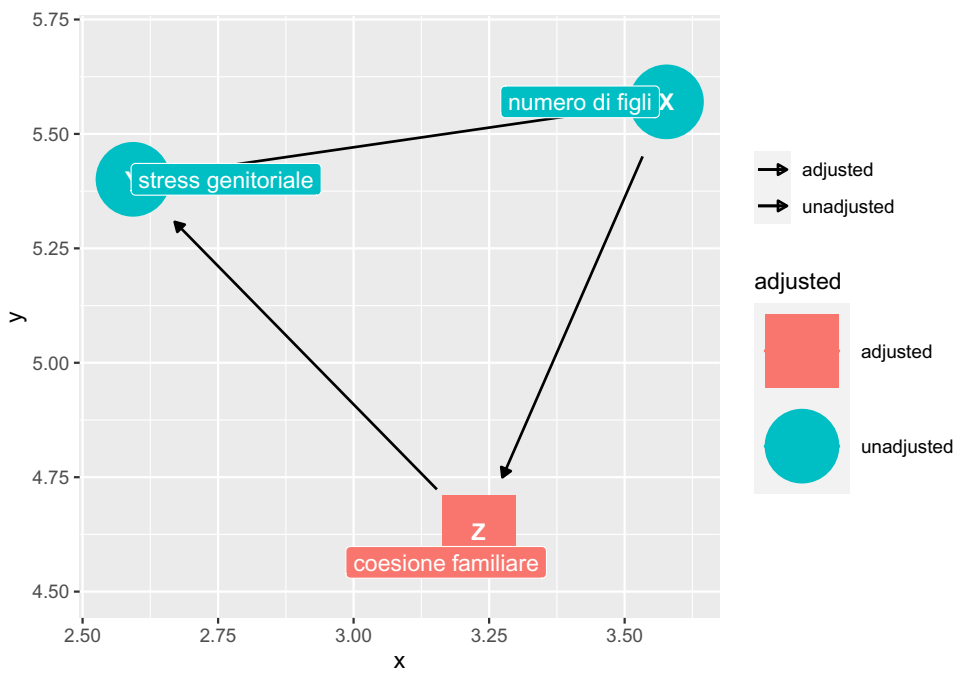


Grafico 5.2. Dopo aver condizionato il moderatore

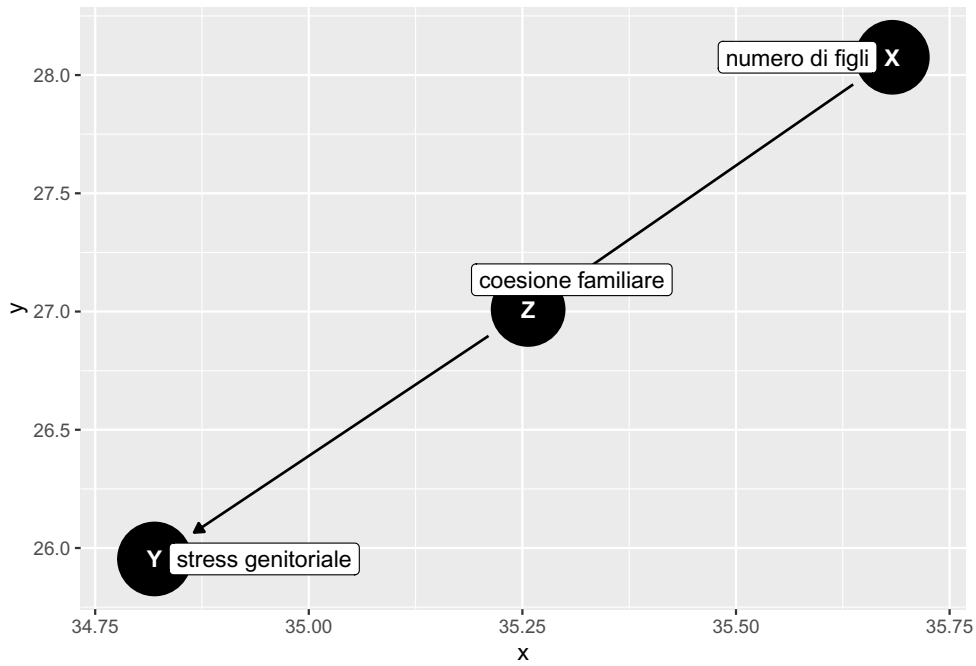


Grafico 6.1. Prima di condizionare il mediatore

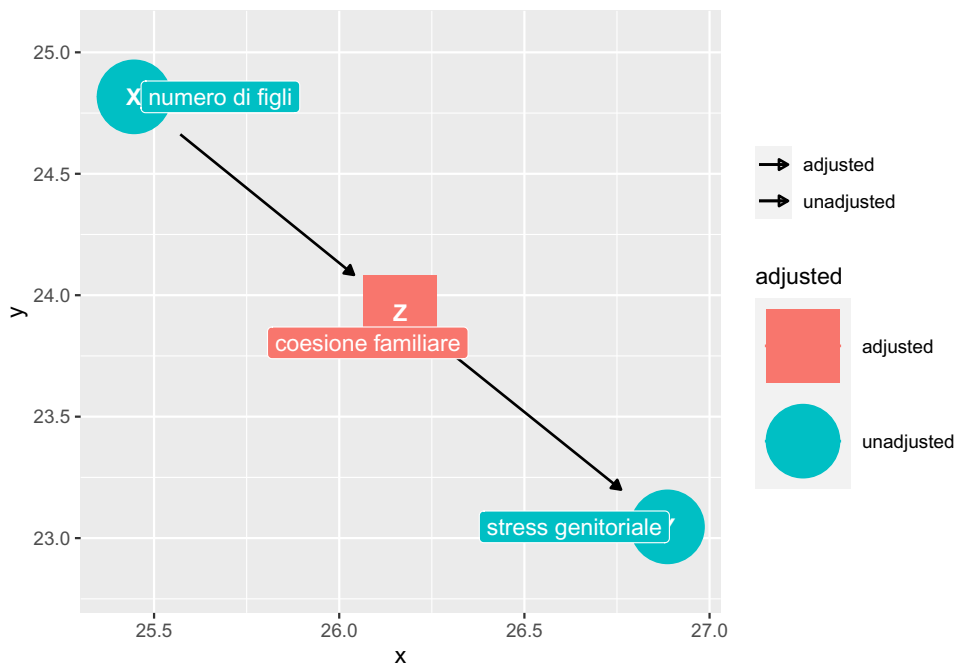


Grafico 6.2. Dopo aver condizionato il mediatore

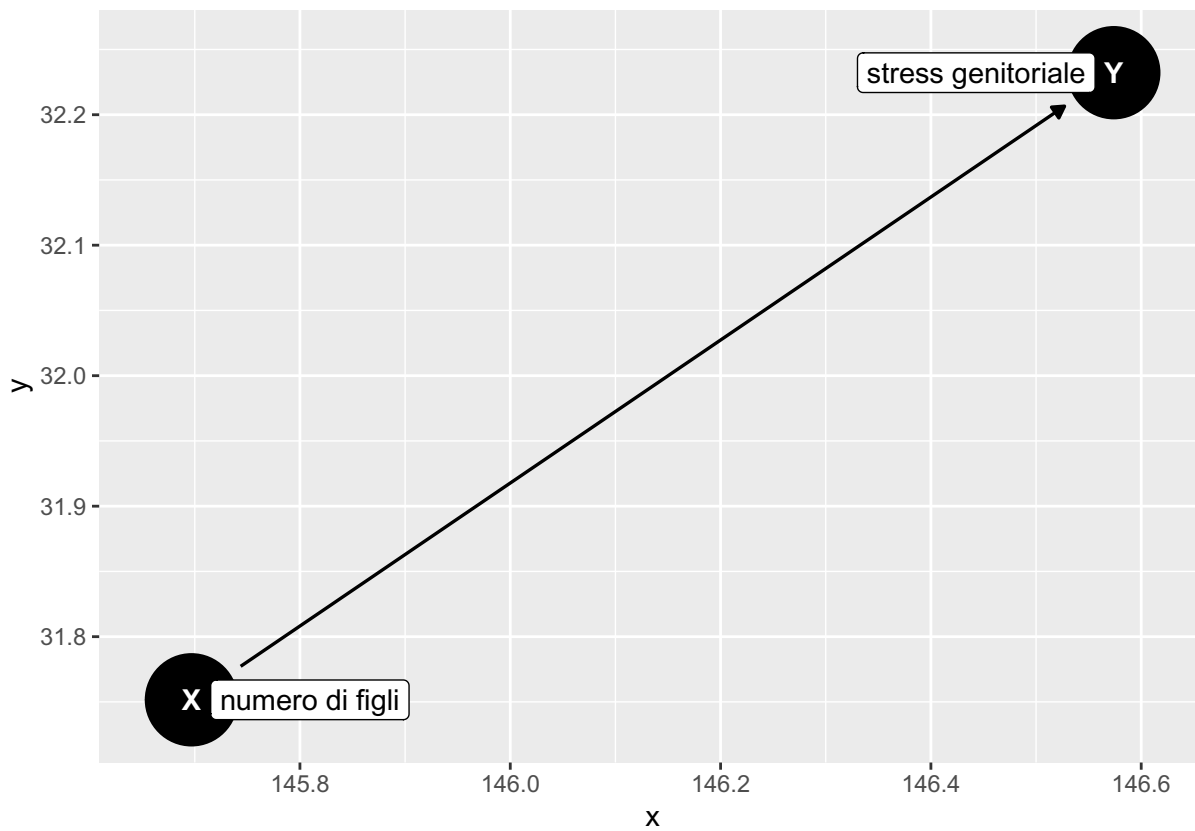


Grafico 6.3. Relazione causale senza il mediatore

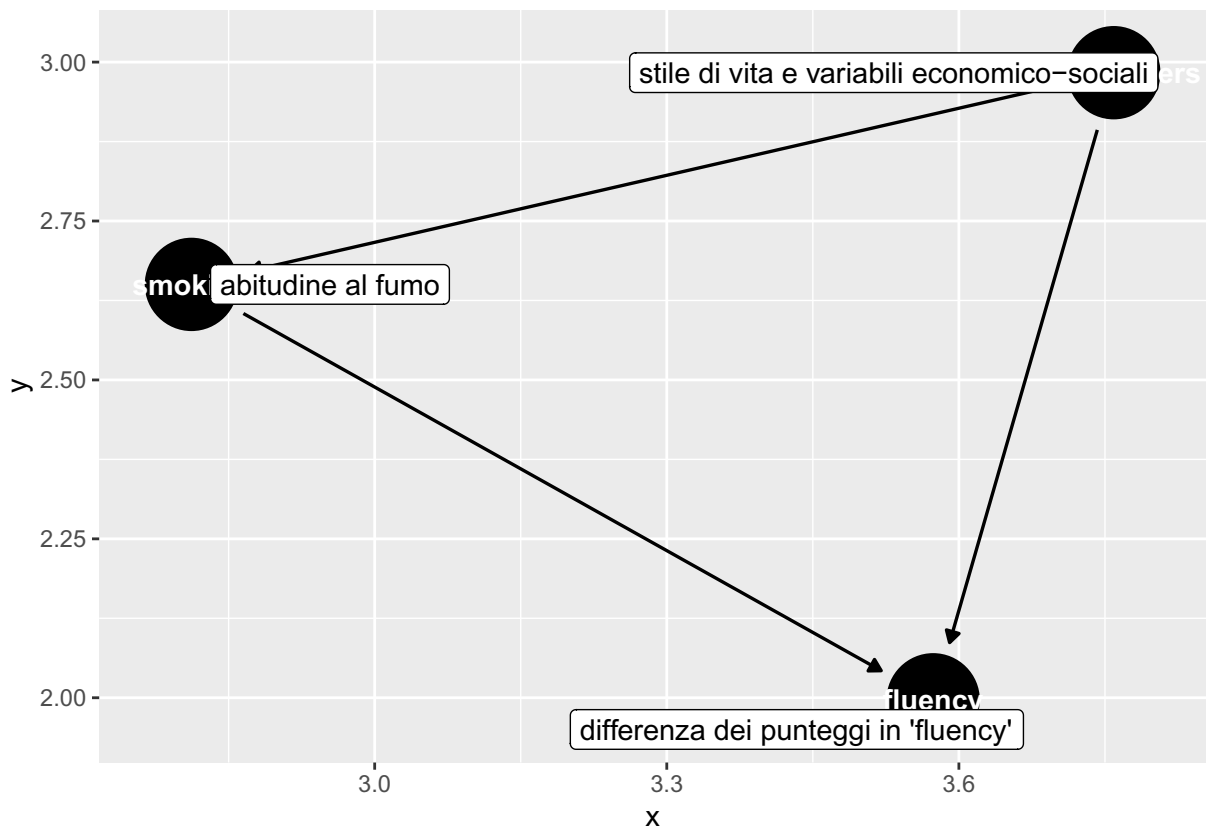


Grafico 7. Rappresentazione grafica dell'ipotesi causale riguardo l'effetto del fumo sul decadimento cognitivo influenzato da stile di vita e fattori economico-sociali

Appendice B. Codice dell'analisi riferita al capitolo 3.2

Statistiche descrittive stratificate per genere

Nella Tabella 1 vengono riportate le statistiche descrittive del possibile fattore di rischio, l'abitudine al fumo, e delle covariate di controllo:

- per le variabili continue, range, media, deviazione standard, mediana, primo e terzo quartile e test di Wilcoxon per la verifica di uguaglianza delle medie tra i due gruppi definiti dal genere;
- per le variabili categoriali, distribuzione di frequenza e test chi-quadrato di indipendenza.

Variabili	N	Overall, N = 33,525	Male, N = 14,675	Female, N = 18,850	p- value
age	33,525				<0.001
Range		39, 90	47, 90	39, 85	
Median (IQR)		62 (56, 68)	63 (57, 68)	61 (56, 67)	
Mean (SD)		62 (8)	63 (8)	62 (8)	
years_of_education	33,525				<0.001
Range		0.0, 21.0	0.0, 21.0	0.0, 21.0	
Median (IQR)		5.0 (5.0, 8.0)	5.0 (5.0, 11.0)	5.0 (5.0, 8.0)	
Mean (SD)		7.2 (4.2)	7.7 (4.2)	6.8 (4.1)	
iscd	33,525				<0.001
Isced-97 code 1-2		24,500 (73%)	10,550 (72%)	13,950 (74%)	
Isced-97 code 3-4		6,125 (18%)	2,825 (19%)	3,300 (18%)	
Isced-97 code 5-6		1,900 (5.7%)	1,000 (6.8%)	900 (4.8%)	
missing		1,000 (3.0%)	300 (2.0%)	700 (3.7%)	
home	33,525				<0.001
No home		6,225 (19%)	2,415 (16%)	3,810 (20%)	
(0,1.5e+05]		15,370 (46%)	6,790 (46%)	8,580 (46%)	
(1.5e+05,Inf]		11,930 (36%)	5,470 (37%)	6,460 (34%)	
health_status	33,525				<0.001
Excellent		2,325 (6.9%)	1,400 (9.5%)	925 (4.9%)	
Very good		4,830 (14%)	2,330 (16%)	2,500 (13%)	
Good		14,410 (43%)	6,710 (46%)	7,700 (41%)	
Fair		9,685 (29%)	3,435 (23%)	6,250 (33%)	
Poor		2,275 (6.8%)	800 (5.5%)	1,475 (7.8%)	
marital_status	33,525				<0.001

Variabili	N	Overall, N = 33,525	Male, N = 14,675	Female, N = 18,850	p- value
Married, living with spouse		27,300 (81%)	12,825 (87%)	14,475 (77%)	
Registered partnership		200 (0.6%)	125 (0.9%)	75 (0.4%)	
Married, not living with spouse		175 (0.5%)	100 (0.7%)	75 (0.4%)	
Never married		1,725 (5.1%)	850 (5.8%)	875 (4.6%)	
Divorced		650 (1.9%)	350 (2.4%)	300 (1.6%)	
Widowed		3,475 (10%)	425 (2.9%)	3,050 (16%)	
gali	33,525				<0.001
Not limited		21,420 (64%)	10,245 (70%)	11,175 (59%)	
Limited		12,105 (36%)	4,430 (30%)	7,675 (41%)	
number_of_chronic_diseases	33,525				<0.001
Range		0, 10	0, 7	0, 10	
Median (IQR)		1 (1, 2)	1 (0, 2)	1 (1, 3)	
Mean (SD)		2 (2)	1 (1)	2 (2)	
bmi	33,525				<0.001
Range		17.0, 51.1	19.5, 43.5	17.0, 51.1	
Median (IQR)		26.3 (23.9, 29.1)	26.6 (24.7, 29.1)	26.0 (23.4, 29.1)	
Mean (SD)		26.7 (4.2)	27.0 (3.5)	26.5 (4.6)	
mobility	33,525				<0.001
Range		0, 10	0, 9	0, 10	
Median (IQR)		1 (0, 2)	0 (0, 1)	1 (0, 3)	
Mean (SD)		1 (2)	1 (1)	2 (2)	
job_status	33,525				<0.001
Retired		17,345 (52%)	9,720 (66%)	7,625 (40%)	
Employed or self-employed		7,200 (21%)	4,275 (29%)	2,925 (16%)	
Permanently sick		200 (0.6%)	150 (1.0%)	50 (0.3%)	
Homemaker		7,900 (24%)	0 (0%)	7,900 (42%)	
Unemployed or other		880 (2.6%)	530 (3.6%)	350 (1.9%)	
smoking	33,525	14,720 (44%)	9,495 (65%)	5,225 (28%)	<0.001
drinking	33,525	8,125 (24%)	6,150 (42%)	1,975 (10%)	<0.001
physical_inactivity	33,525	5,200 (16%)	1,600 (11%)	3,600 (19%)	<0.001
depression_scale	33,525				<0.001
Range		0, 11	0, 9	0, 11	
Median (IQR)		2 (1, 4)	2 (1, 3)	3 (1, 5)	

Variabili	N	Overall, N = 33,525	Male, N = 14,675	Female, N = 18,850	p-value
Mean (SD)		3 (2)	2 (2)	3 (2)	
economic_status	33,525				<0.001
(0,2.37e+03]		1,305 (3.9%)	555 (3.8%)	750 (4.0%)	
(2.37e+03,8e+03]		8,450 (25%)	2,530 (17%)	5,920 (31%)	
(8e+03,1.5e+04]		8,555 (26%)	5,285 (36%)	3,270 (17%)	
(1.5e+04,2.88e+05]		8,135 (24%)	5,600 (38%)	2,535 (13%)	
missing		7,080 (21%)	705 (4.8%)	6,375 (34%)	
number_of_children	33,525				<0.001
No child		3,275 (9.8%)	1,500 (10%)	1,775 (9.4%)	
(0,1]		5,980 (18%)	2,675 (18%)	3,305 (18%)	
(1,2]		13,930 (42%)	6,170 (42%)	7,760 (41%)	
(2,3]		6,555 (20%)	2,775 (19%)	3,780 (20%)	
(3,Inf]		3,785 (11%)	1,555 (11%)	2,230 (12%)	

Tabella 1. Statistiche descrittive per la variabile 'smoking' e le variabili di controllo

Nella Tabella 2 vengono riportate le statistiche descrittive della variabile dipendente, 'fluency', rilevata in due momenti diversi:

- 'fluency.x' rappresenta l'abilità di fluency verbale dei soggetti alla Wave 1 (2004);
- 'fluency.y' rappresenta l'abilità di fluency verbale dei soggetti alla Wave 5 (2013).

Trattandosi di una variabile continua, le statistiche calcolate sono range, media, deviazione standard, mediana, primo e terzo quartile e test di Wilcoxon per la verifica di uguaglianza delle medie tra i due gruppi definiti dal genere.

Variabili	N	Overall, N = 33,525	Male, N = 14,675	Female, N = 18,850	p-value
fluency.x	33,525				<0.001
Range		0, 33	0, 33	0, 33	
Median (IQR)		14 (11, 18)	15 (11, 19)	14 (10, 18)	
Mean (SD)		15 (5)	15 (5)	14 (5)	
fluency.y	33,525				<0.001
Range		0, 100	0, 33	0, 100	

Variabili	N	Overall, N = 33,525	Male, N = 14,675	Female, N = 18,850	p-value
Median (IQR)		14 (11, 18)	14 (11, 19)	14 (10, 18)	
Mean (SD)		14 (7)	15 (6)	14 (7)	

Tabella 2. Statistiche descrittive per la variabile 'fluency'

Characteristic	OR	95% CI	p-value	OR	95% CI	p-value
drinking						
No						
Yes	0.75	0.71, 0.80	<0.001	0.72	0.69, 0.75	<0.001
physical_inactivity						
No						
Yes	0.80	0.74, 0.85	<0.001	0.87	0.82, 0.91	<0.001
age	1.02	1.01, 1.02	<0.001	1.02	1.02, 1.02	<0.001
gender						
Male						
Female	0.68	0.64, 0.72	<0.001	0.68	0.65, 0.71	<0.001
iscd						
Isced-97 code 1-2						
Isced-97 code 3-4	1.23	1.16, 1.31	<0.001	1.20	1.15, 1.25	<0.001
Isced-97 code 5-6	1.22	1.10, 1.36	<0.001	1.15	1.07, 1.24	<0.001
missing	1.06	0.92, 1.22	0.4	0.99	0.89, 1.11	>0.9
marital_status						
Married, living with spouse						
Registered partnership	2.06	1.52, 2.81	<0.001	2.01	1.57, 2.57	<0.001

Characteristic	OR	95% CI	p-value	OR	95% CI	p-value
Married, not living with spouse	0.79	0.58, 1.07	0.13	1.12	0.89, 1.40	0.3
Never married	0.97	0.85, 1.11	0.7	0.76	0.69, 0.84	<0.001
Divorced	1.15	0.98, 1.36	0.090	1.14	1.03, 1.26	0.012
Widowed	1.12	1.03, 1.21	0.009	1.22	1.14, 1.29	<0.001
number_of_children						
No child						
(0,1]	0.96	0.86, 1.07	0.5	0.78	0.72, 0.84	<0.001
(1,2]	1.09	0.99, 1.21	0.087	0.91	0.85, 0.98	0.011
(2,3]	1.22	1.10, 1.37	<0.001	0.99	0.92, 1.07	0.8
(3,Inf]	1.26	1.12, 1.42	<0.001	1.32	1.21, 1.43	<0.001
economic_status						
(0,2.37e+03]						
(2.37e+03,8e+03]	0.95	0.84, 1.08	0.4	0.96	0.88, 1.04	0.3
(8e+03,1.5e+04]	0.85	0.75, 0.96	0.011	0.88	0.81, 0.96	0.004
(1.5e+04,2.88e+05]	0.72	0.64, 0.82	<0.001	0.72	0.67, 0.79	<0.001
missing	0.46	0.40, 0.53	<0.001	0.42	0.38, 0.47	<0.001
home						
No home						
(0,1.5e+05]	0.87	0.82, 0.93	<0.001	0.92	0.88, 0.96	<0.001
(1.5e+05,Inf]	0.75	0.70, 0.80	<0.001	0.75	0.71, 0.78	<0.001
job_status						
Retired						
Employed or self-employed	0.88	0.82, 0.95	<0.001	0.84	0.80, 0.88	<0.001

Characteristic	OR	95% CI	p-value	OR	95% CI	p-value
Permanently sick	4.20	2.96, 6.11	<0.001	2.86	2.23, 3.68	<0.001
Homemaker	1.71	1.57, 1.87	<0.001	1.94	1.82, 2.07	<0.001
Unemployed or other	0.32	0.27, 0.39	<0.001	0.21	0.18, 0.24	<0.001
health_status						
Excellent						
Very good	0.60	0.54, 0.66	<0.001	0.70	0.65, 0.76	<0.001
Good	0.73	0.66, 0.80	<0.001	0.97	0.91, 1.04	0.4
Fair	0.67	0.61, 0.74	<0.001	0.79	0.73, 0.85	<0.001
Poor	0.40	0.34, 0.46	<0.001	0.50	0.45, 0.56	<0.001
number_of_chronic_diseases	1.11	1.09, 1.13	<0.001	1.09	1.07, 1.10	<0.001
bmi	1.01	1.00, 1.01	0.075	1.00	0.99, 1.00	0.2
mobility	0.98	0.97, 1.00	0.045	0.95	0.94, 0.96	<0.001
depression_scale	1.02	1.00, 1.03	0.005	1.01	1.00, 1.01	0.2
smoking						
No						
Yes				1.05	1.01, 1.08	0.004
gali						
Not limited						
Limited				1.11	1.06, 1.15	<0.001

Tabella 3. Confronto dei risultati ottenuti dal modello di regressione logistica senza e con propensity score

Statistiche inferenziali

```
## Analisi binomiale senza propensity score
fit1_full<-glm(difference_in_fluency_bin~smoking+drinking
              +physical_inactivity+age+gender+isced+marital_status
              +number_of_children+economic_status+home+job_status
              +health_status+gali+number_of_chronic_diseases+bmi
              +mobility+depression_scale, data=db_sel1,
              family=binomial) # modello di regressione logistica

# stepAIC backward: dal modello con tutte le variabili indicate,
# deseleggono quelle che non hanno un effetto sul decadimento della fluency
fit1_final<-stepAIC(fit1_full,direction = "backward")
t1<-tbl_regression(fit1_final,exponentiate = TRUE)

## Analisi binomiale con propensity score
# Stima del PS tramite metodo IPTW
mod_ps_smoking<-glm(I(smoking=="Yes")~physical_inactivity+drinking+age
                   +gender+isced+marital_status+number_of_children
                   +economic_status+home+job_status+health_status
                   +gali+number_of_chronic_diseases+bmi+mobility
                   +depression_scale,data=db_sel1,family=binomial)
mod_ps_smoking %>% tbl_regression(exponentiate = TRUE)

# Stima delle probabilità
ps_smo<-predict(mod_ps_smoking,type="response")
# Calcolo dei pesi
db_sel1$pesi_smo<-0
db_sel1$pesi_smo[db_sel1$smoking=="Yes"]<-1/ps_smo[db_sel1$smoking=="Yes"]
db_sel1$pesi_smo[db_sel1$smoking=="No"]<-1/(1-ps_smo[db_sel1$smoking=="No"])

# Modello binomiale corretto con i pesi del PS
mod_fluency_ps_smo<-glm(difference_in_fluency_bin~smoking+drinking
                       +physical_inactivity+age+gender+isced+marital_status
                       +number_of_children+economic_status+home+job_status+health_status
                       +gali+number_of_chronic_diseases+bmi+mobility+depression_scale,
                       data=db_sel1,family=binomial,weights = pesi_smo)

t2<-tbl_regression(mod_fluency_ps_smo,exponentiate = TRUE)
```


Bibliografia

- Anstey, K. J., Sanden, C. von, Salim, A., & O’Kearney, R. (2007). Smoking as a risk factor for dementia and cognitive decline: A meta-analysis of prospective studies. *American Journal of Epidemiology*, *166*(4), 367–378.
- Bleske-Rechek, A., Morrison, K. M., & Heidtke, L. D. (2015). Causal inference from descriptions of experimental and non-experimental research: Public understanding of correlation-versus-causation. *The Journal of General Psychology*, *142*(1), 48–70.
- Börsch-Supan, A. (n.d.-a). *Survey of health, ageing and retirement in europe (SHARE) wave 1. Release version: 8.0.0. SHARE-ERIC*. Retrieved February 10, 2022, from 10.6103/SHARE.w1.800
- Börsch-Supan, A. (n.d.-b). *Survey of health, ageing and retirement in europe (SHARE) wave 5. Release version: 8.0.0. SHARE-ERIC*. Retrieved February 10, 2022, from 10.6103/SHARE.w5.800
- Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., & Zuber, on behalf of the S. C. C. T., Sabrina. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, *42*(4), 992–1001.
- D’Onofrio, B. M., Sjölander, A., Lahey, B. B., Lichtenstein, P., & Öberg, A. S. (2020). Accounting for confounding in observational studies. *Annual Review of Clinical Psychology*, *16*, 25–48.
- Foster, E. M. (2010). Causal inference and developmental psychology. *Developmental Psychology*, *46*(6), 1454.
- Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, *15*(3), 234.
- Hogg, M., & Vaughan, G. M. (2016). *Psicologia sociale. Teorie e applicazioni*. Pearson Italia.
- Li, L., Kleinman, K., & Gillman, M. W. (2014). A comparison of confounding adjustment methods with an application to early life determinants of childhood obesity. *Journal of Developmental Origins of Health and Disease*, *5*(6), 435–447.
- Lilly, C. M. (2014). The ProCESS trial—a new era of sepsis management. *N Engl J Med*, *370*(18), 1750–1751.
- Rosenbaum, P. (2018). *Observation and experiment*. Harvard University Press.
- Rubin, D. B. (2006). *Matched sampling for causal effects*. Cambridge University Press.
- West, S. G., & Thoemmes, F. (2010). Campbell’s and rubin’s perspectives on causal inference. *Psychological Methods*, *15*(1), 18.
- Weuve, J., Tchetgen, E. J. T., Glymour, M. M., Beck, T. L., Aggarwal, N. T., Wilson, R. S., Evans, D. A., & Leon, C. F. M. de. (2012). Accounting for bias due to selective attrition: The example of smoking and cognitive decline. *Epidemiology (Cambridge)*

Mass.), 23(1), 119.