Il paradosso di Simpson

Cicchitelli Cap. 9 pp. 228-230

Soccorso con elicottero o con ambulanza?

Ovvero: come la lettura acritica di una tavola doppia può portare a conclusioni errate

	Elicottero	Ambulanza	*Totale*
Morti	64	260	*324*
Soprawissuti	136	840	*976*
Totale	*200*	*1100*	*1300*

Pazienti morti:

- Elicottero: 64/200 = 32%
- Ambulanza: 260/1100 = 24%

Soccorso con elicottero o con ambulanza? (cont.)

Incidenti gravi

	Elicottero	Ambulanza	*Totale*
Morti	48	60	*108*
Soprawissuti	52	40	*92*
Totale	*100*	*100*	*200*

Pazienti morti:

- Elicottero: 48/100 = 48%
- Ambulanza: 60/100 = 60%

Incidenti non gravi

	Elicottero	Ambulanza	*Totale*
Morti	16	200	*216*
Soprawissuti	84	800	*884*
Totale	*100*	*1000*	*1100*

Pazienti morti:

- Elicottero: 16/100 = 16%
- Ambulanza:200/1000= 20%

È un esempio del paradosso di Simpson!

Il paradosso di Simpson

Il paradosso di Simpson si manifesta quando la relazione tra due variabili nell'intera popolazione è molto diversa dalla relazione tra le stesse variabili nelle sotto-popolazioni definite da una terza variabile, tanto che le conclusioni sono opposte.

Nell'esempio precedente:

- X = mezzo di soccorso (elicottero/ambulanza)
- Y = esito (morto/sopravvissuto)
- Z = gravità dell'incidente (grave/non grave)

La relazione tra X e Y nell'intera popolazione (1300 persone) è ben diversa dalla relazione tra X e Y nella sottopopolazione con incidente grave (Z= 'grave', 200 persone) e in quella con incidente non grave (Z='non grave', 1100 persone)

- Simpson, E. H. 1951. The interpretation of interaction in contingency tables. *J. Roy. Statist. Soc. Ser. B* 13: 238-241.
- http://it.wikipedia.org/wiki/Paradosso_di_Simpson
- Tesi di laurea di A. Pupo (http://local.disia.unifi.it/grilli/projects.htm)

Discriminazione sessuale?

Negli anni Settanta un responsabile dell'università di Berkeley ha analizzato i dati relativi alle assunzioni nei vari dipartimenti, scoprendo che il tasso di ammissione delle femmine era sostanzialmente inferiore a quello dei maschi. L'università poteva essere accusata di discriminazione sessuale!
Per illustrare quel caso, consideriamo un esempio semplificato con 200 candidati, di cui 100 maschi e 100 femmine

	Maschi	Femmine
Assunti	55	45
Non assunti	45	55
Totale	100	100
% assunti	*55.0%*	*45.0%*

Il minore tasso di assunzione delle femmine è una prova di discriminazione sessuale?

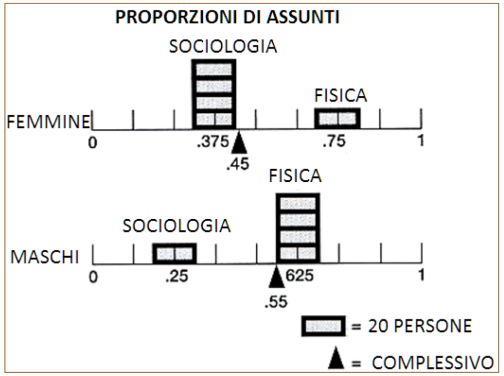
Discriminazione sessuale? /cont.

	Sociologia		Fisica		Totale	
	Maschi	Femmine	Maschi	Femmine	Maschi	Femmine
Assunti	5	30	50	15	55	45
Non assunti	15	50	30	5	45	55
Totale	20	80	80	20	100	100
% assunti	*25.0%*	*37.5%*	*62.5%*	*75.0%*	*55.0%*	*45.0%*

L'analisi dei dati ha mostrato che l'accusa era palesemente infondata: infatti, considerando i tassi di ammissione per dipartimento, la situazione era rovesciata poiché nella maggior parte dei dipartimenti le femmine facevano registrare un tasso di successo più elevato (il tasso di assunzione globale risultava inferiore perché le femmine facevano domanda soprattutto nei dipartimenti più «difficili», cioè con molti candidati per ogni posizione).

Cfr. Bickel PJ, Hjammel EA, O'Connell JW (1975) Sex Bias in Graduate Admissions: Data From Berkeley. *Science*, 187: 398–404.

Discriminazione sessuale? /cont.

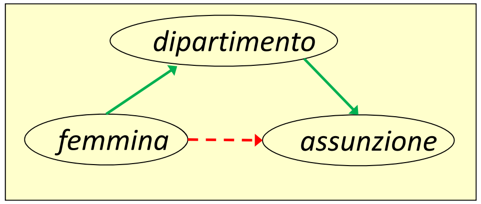


Il tasso di successo complessivo si ottiene come media pesata dei tassi di successo nei due dipartimenti, usando pesi proporzionali al numero di partecipanti.
Ad es., le femmine hanno un tasso di successo di 0.375 a sociologia (dove partecipano in 80) e 0.75 a fisica (dove partecipano in 20) → il tasso di successo complessivo è **0.375*(80/100)+0.75*(20/100)=0.45**

Il disegno mostra che in entrambi i dipartimenti il tasso di successo delle femmine è superiore a quello dei maschi, ma per il tasso complessivo accade il contrario a causa dei pesi (le femmine partecipano in prevalenza alla selezione nel dipartimento più 'difficile').

Discriminazione sessuale? /cont.

Interpretazione in termini di effetti causali



Essere femmina influenza probabilità di assunzione in due modi:
- Effetto diretto (la freccia tratteggiata)
- Effetto indiretto tramite la scelta del dipartimento (le altre frecce)

L'effetto di interesse per valutare la discriminazione è quello diretto. Tuttavia, studiando la relazione tra femmina e assunzione senza considerare il dipartimento non si ottiene l'effetto diretto, ma quello complessivo (diretto + indiretto). Per ottenere l'effetto diretto l'analisi va svolta condizionatamente al dipartimento, cioè si deve studiare la relazione tra femmina e assunzione separatamente per ogni dipartimento.

Trattamento dei calcoli renali

Uno studio clinico ha confrontato due trattamenti per i calcoli renali:

- Trattamento A: intervento chirurgico tradizionale
- Trattamento B: nefrolitotomia percutanea (inserimento di strumenti attraverso una piccola incisione nella schiena)

Lo studio ha coinvolto 700 pazienti, divisi equamente tra i due tipi di trattamento

	Tr. A	Tr. B
Riuscito	273	289
Non riuscito	77	61
Totale	350	350
% successo	*78.0%*	*82.6%*

Il maggior tasso di successo del trattamento B è una prova della sua maggiore efficacia?

Trattamento dei calcoli renali /cont.

Nei casi più difficili, cioè quelli con calcoli grandi (large stones), i medici hanno usato prevalentemente il trattamento A, e viceversa.

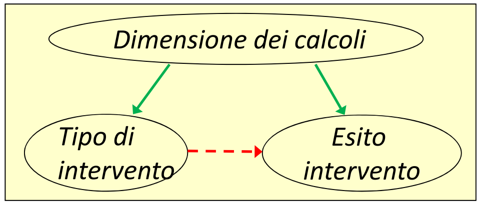
	Treatment A	Treatment B
Small Stones	*Group 1* **93% (81/87)**	*Group 2* 87% (234/270)
Large Stones	*Group 3* **73% (192/263)**	*Group 4* 69% (55/80)
Both	78% (273/350)	**83% (289/350)**

Il trattamento A ha un tasso di successo superiore in entrambe le tipologie (calcoli piccoli e calcoli grandi), anche se nel complesso dei 700 casi presi in esame appare inferiore.

Steven A. Julious and Mark A. Mullee (12/03/1994). "Confounding and Simpson's paradox" (http:/ / bmj. bmjjournals. com/ cgi/ content/full/ 309/ 6967/ 1480). BMJ 309 (6967): 1480–1481

Trattamento dei calcoli renali /cont.

Interpretazione in termini di effetti causali



Il tipo di intervento ha effetto sull'esito (freccia tratteggiata). Tuttavia la dimensione dei calcoli influenza sia il tipo di intervento che l'esito (altre frecce): nel linguaggio statistico-epidemiologico, la dimensione dei calcoli viene detta **confondente** (confounder), perché confonde la relazione di interesse principale (quella tra tipo di intervento e esito).
Ignorando il confondente, cioè studiando la relazione tra tipo di intervento e esito senza considerare la dimensione dei calcoli, si ottiene un inutile effetto misto. Per ottenere l'effetto di interesse l'analisi va svolta condizionatamente alla dimensione dei calcoli. *Nota: in questo esempio il confondente provoca un'inversione del segno dell'effetto (paradosso di Simpson), in generale questo non accade (il confondente riduce o amplifica l'effetto, che però rimane dello stesso segno).*

Razza e pena di morte

	Imputato bianco		Imputato nero	
	Vittima bianca	Vittima nera	Vittima bianca	Vittima nera
A morte	19	0	11	6
No	132	9	52	97
	12.6%	**0%**	**17.5%**	**5.8%**

Esercizio:

Costruire la tabella bivariata con "Esito (a morte vs no)" e "Imputato (bianco vs nero)". Commentare il paradosso.

Fallacia di Berkson

Paradosso di Simpson: la relazione tra due variabili di interesse X e Y si inverte quando ci si condiziona ad una terza variabile Z.

Negli esempi visti fino ad ora l'analisi che ignora Z è sbagliata, mentre l'analisi che condiziona a Z è corretta. Questo è ciò che accade di solito, ma ci sono situazioni in accade il contrario, ovvero è l'analisi condizionata a Z a produrre risultati fuorvianti.

Un caso interessante è la **fallacia di Berkson**, in cui Z denota l'appartenenza al campione osservato. Esempio (fittizio, da J. Ellenberg): X: diabete, Y: ipertensione, questa è la tabella doppia nell'intera popolazione

	Ipertensione		
Diabete	No	Sì	
No	420	180	600
Sì	280	120	400
	700	300	1000

Nella popolazione le due patologie sono indipendenti: la proporzione di ipertesi è la stessa tra i non diabetici (180/600=0.3) e diabetici (280/400=0.3)

Fallacia di Berkson /cont.

Supponiamo di non poter osservare l'intera popolazione, ma solo coloro che sono ricoverati in ospedale (Z=1) e consideriamo due scenari.

Scenario A: sono in ospedale tutti coloro che hanno almeno una delle due patologie

	Ipertensione		
Diabete	No	Sì	
No	0	180	180
Sì	280	120	400
	280	300	580

Nella sottopopolazione in ospedale le due patologie sono positivamente associate: la proporzione di ipertesi è <u>maggiore</u> tra i non diabetici (180/180=1) che tra i diabetici (120/400=0.3)

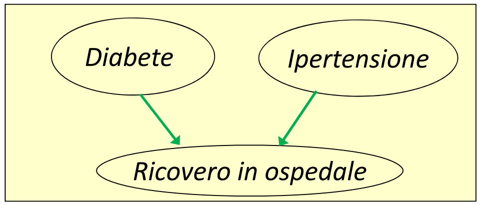
Scenario B: sono in ospedale tutti coloro che hanno entrambe le due patologie + 1 su 10 di coloro che hanno una sola patologia + altri 12 soggetti che non hanno ne l'una né l'altra

	Ipertensione		
Diabete	No	Sì	
No	12	18	30
Sì	28	120	148
	40	138	178

Nella sottopopolazione in ospedale le due patologie sono negativamente associate: la proporzione di ipertesi è <u>minore</u> tra i non diabetici (18/30=0.6) che tra i diabetici (120/148=0.81)

Fallacia di Berkson /cont.

Interpretazione in termini di effetti causali



In questo esempio non c'è associazione tra diabete e ipertensione (nessuna freccia). Tuttavia, diabete e ipertensione hanno un effetto comune (ricovero in ospedale), per cui l'analisi condizionata allo stato di ricovero in ospedale crea una associazione fittizia tra diabete e ipertensione. In questo caso l'analisi condizionata (cioè sui soggetti ricoverati in ospedale) è sbagliata, mentre quella non condizionata (cioè su tutta la popolazione) è corretta.
La fallacia di Berkson è detta anche **distorsione da selezione campionaria**: in generale, si verifica quando l'appartenere o meno al campione osservato dipende dal valore delle variabili di interesse, per cui la relazione tra le variabili nel campione è diversa dalla relazione nella popolazione.