

STATISTICA

INTRODUZIONE ALLE RELAZIONI MULTIVARIATE

Andrea Giommi

Dipartimento di Statistica, Informatica, Applicazioni (DiSIA)
Università degli Studi di Firenze

Scuola di Psicologia
Corso di Studio in Scienze e Tecniche Psicologiche

- Nel contesto dell'associazione tra due variabili (Cap. 8 - variabili qualitative; Cap. 9 - variabili quantitative) hanno un ruolo importante il concetto di *causalità* e di *relazione causale*
- Per interpretare l'associazione tra due variabili è normale assegnare ad una il significato di variabile risposta e all'altra quello di variabile esplicativa.
- Questo porta inevitabilmente a pensare che l'esplicativa abbia un'influenza sulla variabile risposta e talvolta che ne sia la *causa*
- La presenza di associazione tra due variabili **non** implica un rapporto di **causa-effetto** tra le stesse

Controllo dell'esistenza di un nesso di causalità

- Quando si ritiene che una variabile eserciti un'influenza su di un'altra variabile si utilizza la seguente notazione:

$$X \longrightarrow Y$$

- Per poter stabilire che esista un rapporto di causa effetto, devono essere soddisfatte tre condizioni:
 - ✓ associazione tra variabili
 - ✓ un appropriato ordine cronologico
 - ✓ assenza di spiegazioni alternative
- Solo se tutte e tre le condizioni sono soddisfatte si può associare ad una relazione osservata il carattere di causalità

Esclusione di spiegazione alternativa

- E' abbastanza facile nelle ricerche in ambito sociale rilevare associazioni tra variabili unitamente a successioni cronologiche che facciano pensare a nessi di causalità
- E' tuttavia necessario sgombrare il campo da possibili spiegazioni alternative
- Eliminare possibili alternative non è facile perché spesso sono molti i fattori che possono influenzare la relazione tra variabili
- Non è mai possibile *provare* in modo definitivo un nesso di causa effetto
- E' tuttavia possibile confutare un nesso di causalità mostrando che l'evidenza empirica contraddice una delle tre condizioni necessarie

Controllo per altre variabili

- Per stabilire un nesso causale è necessario poter eliminare spiegazioni alternative a quella ipotizzata
- Per far questo valutiamo se l'associazione tra due variabili, X e Y permane anche quando rimuoviamo l'influenza di altre variabili.
- Nell'associazione tra variabili diciamo che una variabile è controllata quando la sua influenza è rimossa
- Nell'ambito della ricerca sociale, controllare una variabile significa normalmente raggruppare le osservazioni che per quella variabile presentano la stessa modalità
- Quindi, verificare se all'interno dei gruppi così ottenuti è ancora presente l'associazione tra le variabili che stiamo studiando

Esempio di controllo statistico

Analizziamo l'associazione tra l'appartenere ai *boy-scout* (Si, No) e essere *delinquente* (Si, No), controllando successivamente per una terza variabile: *frequenza alla chiesa* (Bassa, Media, Alta)

		<i>Delinquente</i>		
		<i>Si</i>	<i>No</i>	<i>Totale</i>
<i>Scout</i>	<i>Si</i>	36 (9%)	364 (91%)	400
	<i>No</i>	60 (15%)	340 (85%)	400

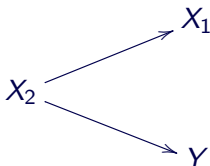
		<i>Frequenza chiesa</i>					
		<i>Bassa</i>		<i>Media</i>		<i>Alta</i>	
<i>Delinquente</i>		<i>Si</i>	<i>No</i>	<i>Si</i>	<i>No</i>	<i>Si</i>	<i>No</i>
<i>Scout</i>	<i>Si</i>	10 (20%)	40 (80%)	18 (12%)	132 (88%)	8(4%)	192 (96%)
	<i>No</i>	40 (20%)	160 (80%)	18 (12%)	132 (88%)	2 (4%)	48 (96%)

- Non sempre è facile effettuare il controllo delle variabili. Talvolta nella ricerca sociale, si corre il rischio di non considerare per nulla variabili esplicative anche molto importanti. Per fortuna il rischio è ridotto dalle esperienze di ricerca acquisite nel tempo
- Una variabile la cui influenza è provata ma non misurata, viene definita **variabile in agguato** o **variabile confondente**
- Se studio l'influenza di $X \rightarrow Y$, ma so che Z ha molta influenza su Y , ma che ha anche una qualche associazione con X , appare evidente che l'influenza di X su Y può essere in parte (o anche del tutto) spiegata da Z
- Questa è una situazione molto frequente nella ricerca medica, dove è doveroso tenere conto dell'età nello studio dei fattori di rischio per determinate patologie

Tipi di relazioni multivariate

- Nella ricerca sociale è normale spiegare una variabile risposta in funzione di diverse esplicative
- le relazioni multivariate che possono presentarsi appartengono alle seguenti tipologie:
 - ✓ **Associazioni (correlazioni) spurie**
 - ✓ **Relazioni concatenate**
 - ✓ **Cause multiple**
 - ✓ **Variabili sopprimenti**
 - ✓ **Interazioni**

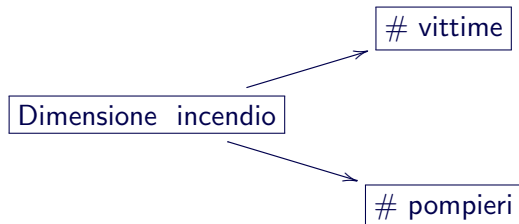
- Associazione spuria (tra X_1 e Y)



- L'associazione spuria tra X_1 e Y è dovuta all'effetto che una terza variabile X_2 ha sia sulla variabile "esplicativa" X_1 , sia sulla variabile "risposta" Y .
- **Controllando per X_2 l'associazione scompare**

Esempio di associazione spuria

- Si possono fare molti esempi di associazione spuria.
In una serie numerosa di incendi si è rilevata una correlazione positiva tra il numero delle vittime e il numero di pompieri impegnati nello spegnimento dell'incendio:



- Si potrebbe paradossalmente pensare che per ridurre il numero delle vittime fosse sufficiente ridurre il numero dei pompieri. In realtà vittime e pompieri sono correlati con la dimensione dell'incendio

Relazioni concatenate

- Un altro tipo di associazione spuria è dato dalle relazioni concatenate, nelle quali X_1 è solo causa indiretta di Y attraverso la sua influenza su X_2 , detta **variabile interveniente** o **mediatrice** che, a sua volta, influenza Y

$$X_1 \longrightarrow X_2 \longrightarrow Y$$

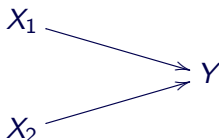
- Studi sulla longevità umana hanno rilevato una associazione positiva tra lunghezza della vita e livello di istruzione
- Si è propensi a pensare che nella maggior parte dei casi la variabile reddito *intervenga* a determinare questa associazione:



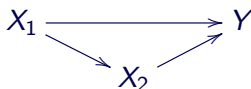
- controllando per *Reddito*, l'associazione **scompare**

Cause multiple

- Una variabile Y può avere più di una causa:



- Se le variabili X_1 e X_2 sono statisticamente indipendenti, il controllo per X_2 lascia inalterata la relazione X_1, Y
- Nella ricerca sociale, le variabili esplicative sono spesso associate. Ad esempio, X_1 può avere sulla Y anche un'influenza indiretta:



- il controllo per X_2 può alterare la relazione X_1, Y

Variabili sopprimenti

- In alcune situazioni due variabili possono risultare non associate fino a quando non si inserisce nella relazione una terza variabile.
- Nell'esempio che segue, la relazione tra Istruzione e Reddito non emerge finché non si controlla per Età

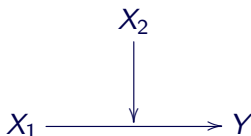
		Età=bassa			Età=alta			
		Reddito	Alto	Basso	%Alto	Alto	Basso	%Alto
Istruzione	Alta		125	225	35,7%	125	25	83,3%
	Bassa		25	125	16,7%	225	125	64,3%

- L'età è associata positivamente con il reddito e negativamente con l'istruzione

		Reddito		Reddito			Istruzione		
Istruzione		Alto	Basso	Età	Alto	Basso	Età	Alta	Bassa
Alta		250	250	Alta	350	150	Alta	150	350
Bassa		250	250	Bassa	150	350	Bassa	350	150

Interazione

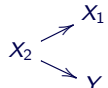
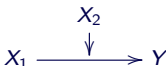

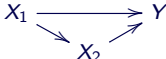
- Si dice che c'è interazione tra la variabile X_1 e la variabile X_2 se la relazione tra X_1 e Y varia al variare dei possibili livelli di X_2
- Il grafo che segue mostra chiaramente che la variabile X_2 influisce direttamente sulla relazione tra X_1 e Y



- Un esempio è rappresentato dagli effetti che l'età può avere sulla relazione tra fattori di rischio e patologie: tra numero di sigarette fumate e insorgenza di tumore polmonare, la relazione è nulla tra i giovani e molto forte tra gli anziani

- Per le relazioni spurie (X_2 influenza sia X_1 che Y) e per le relazioni a catena, (X_2 interviene tra X_1 e Y) l'associazione tra X_1 e Y scompare quando controlliamo per X_2 .
- In presenza di cause multiple, un'associazione può cambiare in presenza di un controllo ma non scomparire.
- Quando c'è una variabile sopprimente un'associazione compare soltanto in presenza di un controllo.
- Quando c'è un'interazione statistica, un'associazione ha differente forza e/o direzione in relazione a diversi livelli della variabile di controllo

Tabella riassuntiva delle relazioni multivariate

Grafico	Nome della relazione	Cosa accade dopo il controllo per X_2
 <pre>graph LR; X2 --> X1; X2 --> Y;</pre>	Associazione spuria $X_1 Y$	L'associazione tra X_1 e Y scompare
$X_1 \longrightarrow X_2 \longrightarrow Y$	Relazione a catena; X_2 interviene e X_1 causa indirettamente Y	L'associazione tra X_1 e Y scompare
 <pre>graph LR; X1 --> Y; X2 --> X1_Y; style X1_Y stroke-width:0px;</pre>	Interazione	L'associazione tra X_1 e Y varia secondo il livello di X_2
 <pre>graph LR; X1 --> Y; X2 --> Y;</pre>	Cause multiple (indipendenti)	L'associazione tra X_1 e Y non cambia
 <pre>graph LR; X1 --> Y; X1 --> X2; X2 --> Y;</pre>	Effetti sia diretti sia indiretti di X_1 su Y	L'associazione tra X_1 e Y cambia ma non scompare