

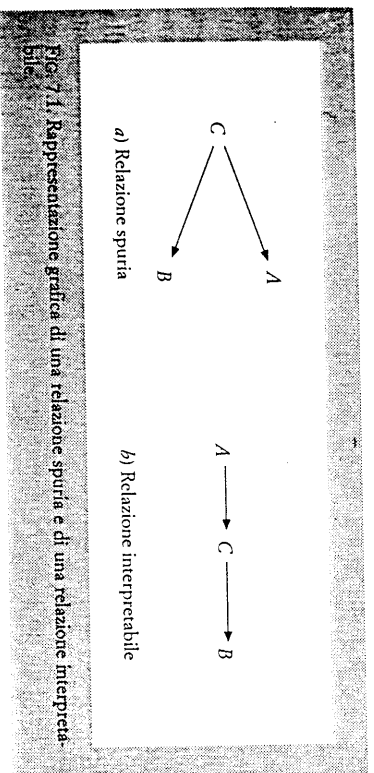


Fig. 7.1. Rappresentazione grafica di una relazione spuria e di una relazione interpretabile.

La figura 7.1 illustra la differenza concettuale esistente fra questi due tipi di relazione. Come si può vedere, in entrambi i casi la covariazione osservata fra le variabili A e B è completamente «spiegata» da una terza variabile, C . Il ruolo di C , tuttavia, varia in modo sostanziale nei due casi. Nel caso della relazione spuria, C rappresenta la causa comune di entrambe le variabili. Nel caso della relazione interpretabile, invece, C è una variabile interviniente che rende più intelligibile il meccanismo causale che lega A e B . È importante sottolineare che il ruolo specifico svolto dalla terza variabile non è mai rivelato dall'analisi statistica. Per determinare con precisione il ruolo svolto dalla terza variabile il ricercatore può solo basarsi sulla teoria o sui risultati delle ricerche precedenti.

1.3. Le cause multiple

A partire dal secolo scorso, sono veramente poche le teorie che hanno affermato che il comportamento umano è causato da un unico fattore. Al contrario, la maggior parte di esse postulano l'esistenza di processi di **causazione multipla**: in altri termini, la maggior parte delle teorie ritengono che la variazione osservata in un dato fenomeno di interesse sia dovuta all'azione congiunta di diverse cause.

Nella maggior parte delle ricerche, dunque, l'enfasi viene posta non sulla relazione fra una singola variabile indipendente e la variabile dipendente di interesse, bensì sull'identificazione delle diverse variabili indipendenti che, simultaneamente, influiscono (in misura maggiore o minore) sulla variabile dipendente. Per esempio, nell'ambito dello studio della stratificazione sociale diversi ricercatori hanno sviluppato e sottoposto a verifica modelli piuttosto elaborati del processo di conseguimento dello status sociale. Il reddito di un individuo può essere visto come la conseguenza degli effetti intrecciati esercitati dall'origine sociale, dall'istruzione, dall'occupazione attuale, dal settore di impiego,

e da altre variabili legate alla carriera lavorativa. *Controllando simultaneamente gli effetti esercitati da diversi fattori causali il ricercatore può trarne alcune inferenze sull'importanza relativa di ogni fattore nel processo di determinazione del reddito, tenendo contemporaneamente conto delle possibili relazioni spurie e dei possibili processi intervinienti.*

Nei prossimi paragrafi illustreremo le tecniche di base per controllare statisticamente l'effetto di una singola variabile supplementare.

2. Il controllo di una variabile supplementare nell'analisi delle tavole 2×2

Come abbiamo osservato nel capitolo 5, la relazione bivariata più semplice è rappresentata dall'incrocio di due variabili dicotomiche. Analogamente, per illustrare la logica dell'analisi multivariata ricorremo al caso più semplice, rappresentato da una tavola di contingenza formata dall'incrocio di tre variabili dicotomiche. Precisamente, utilizzando dati ipotetici mostreremo come può variare la relazione fra due variabili dicotomiche una volta che gli effetti di una terza variabile sono controllati statisticamente.

2.1. Un esempio: religiosità familiare e attività sessuale degli adolescenti

Nel nostro esempio assumeremo l'esistenza di una relazione bivariata fra religiosità della famiglia (X) e attività sessuale premaritimoniale degli adolescenti (Y). I dati ipotetici riportati nella tabella 7.1 classificano 192 adolescenti in base a due criteri: l'appartenenza a una famiglia «molto religiosa» (sì o no) e l'attività sessuale premaritimoniale (sì o no). La tabella 7.1 può essere definita una **tavola di ordine zero**, dove il termine *ordine* si riferisce al numero di variabili supplementari mantenute costanti.

Tab. 7.1. *Attività sessuale premaritimoniale degli adolescenti, secondo la religiosità della famiglia (dati ipotetici)*

	Famiglia molto religiosa		Totale
	No	Sì	
Esperienze sessuali premaritimoniali			
Sì	42,9% (36)	16,0% (16)	27,1% (52)
No	57,1% (56)	84,0% (84)	72,9% (140)
Totale	100,0% (92)	100,0% (100)	100,0% (192)

$$r_{xy} = -0,26$$

In questo esempio la religiosità della famiglia è la variabile indipendente, mentre l'attività sessuale prematrimoniale rappresenta la variabile dipendente. Per questo nella tabella 7.1 le percentuali sono state calcolate per colonna. Come era lecito aspettarsi, gli adolescenti appartenenti a famiglie meno (o per nulla) religiose hanno molte più probabilità di avere esperienze sessuali prematrimoniali. La correlazione fra queste due variabili è pari a $-0,26$, valore che indica l'esistenza di una relazione inversa moderata. Per calcolare questa correlazione (così come le altre presentate in questo capitolo) abbiamo assegnato valore 1 alla risposta «Sì» e valore zero alla risposta «No» per entrambe le variabili.

Osservando la tabella 7.1, ci si può chiedere se l'apparente relazione causale osservata potrebbe essere spiegata dagli effetti esercitati da altri processi sociali che intervengono fra la religiosità della famiglia e l'attività sessuale prematrimoniale degli adolescenti. Supponiamo che l'attività sessuale prematrimoniale di frequente opportunità di avere rapporti sessuali sia la disponibilità di frequenti opportunità di avere rapporti sessuali sia un fattore importante nel determinare la probabilità che un adolescente abbia esperienze di questo tipo. Forse gli adolescenti che vivono in famiglie più religiose sono sottoposti a maggiori controlli e a maggiori limitazioni di attività da parte dei loro genitori. Se questa ipotesi è vera, controllando statisticamente gli effetti delle opportunità la relazione fra religiosità della famiglia e attività sessuale prematrimoniale dovrebbe cambiare.

Per verificare questa ipotesi introdurremo nell'analisi una terza variabile, detta appunto *variabile di controllo* (W), che classifica gli adolescenti in base alla loro possibilità di utilizzare regolarmente un'automobile (sì o no). Le tabelle 7.2-7.5 illustrano quattro possibili modi in cui la relazione bivariata originale si potrebbe modificare introducendo la variabile di controllo. Ognuna di queste quattro tabelle suggerisce una differente interpretazione della relazione iniziale; ognun-

TAB. 7.2. Caso 1: effetto nullo della variabile di controllo

		Possibilità di usare regolarmente un'automobile					
		No			Sì		
		Famiglia molto religiosa			Famiglia molto religiosa		
Esperienze sessuali prematrimoniali	Sì	No	Sì	Totale	No	Sì	Totale
	No	39,7% (25)	15,9% (11)	27,3% (36)	37,9% (11)	16,1% (5)	26,7% (16)
		60,3% (38)	84,1% (58)	72,7% (96)	62,1% (18)	83,9% (26)	73,3% (44)
Totale		100,0% (63)	100,0% (69)	100,0% (132)	100,0% (29)	100,0% (31)	100,0% (60)
		$r_{XY} = -0,27$					
		$r_{XY} = -0,25$					

TAB. 7.3. Caso 2: effetto parziale della variabile di controllo

		Possibilità di usare regolarmente un'automobile					
		No			Sì		
		Famiglia molto religiosa			Famiglia molto religiosa		
Esperienze sessuali prematrimoniali	Sì	No	Sì	Totale	No	Sì	Totale
	No	25,5% (12)	12,9% (11)	17,4% (23)	53,3% (24)	33,3% (5)	48,3% (29)
		74,5% (35)	87,1% (74)	82,6% (109)	46,7% (21)	66,7% (10)	51,7% (31)
Totale		100,0% (47)	100,0% (85)	100,0% (132)	100,0% (45)	100,0% (15)	100,0% (60)
		$r_{XY} = -0,16$					
		$r_{XY} = -0,17$					

TAB. 7.4. Caso 3: spiegazione completa della variabile di controllo

		Possibilità di usare regolarmente un'automobile					
		No			Sì		
		Famiglia molto religiosa			Famiglia molto religiosa		
Esperienze sessuali prematrimoniali	Sì	No	Sì	Totale	No	Sì	Totale
	No	11,1% (4)	11,1% (8)	11,1% (12)	66,7% (32)	66,7% (8)	66,7% (40)
		88,9% (40)	88,9% (80)	88,9% (120)	33,3% (16)	33,3% (4)	33,3% (20)
Totale		100,0% (44)	100,0% (88)	100,0% (132)	100,0% (48)	100,0% (12)	100,0% (60)
		$r_{XY} = 0,00$					
		$r_{XY} = 0,00$					

TAB. 7.5. Caso 4: effetto di interazione della variabile di controllo

		Possibilità di usare regolarmente un'automobile					
		No			Sì		
		Famiglia molto religiosa			Famiglia molto religiosa		
Esperienze sessuali prematrimoniali	Sì	No	Sì	Totale	No	Sì	Totale
	No	50,0% (26)	11,8% (8)	28,3% (34)	25,0% (10)	25,0% (8)	25,0% (18)
		50,0% (26)	88,2% (60)	71,7% (86)	75,0% (30)	75,0% (24)	75,0% (54)
Totale		100,0% (52)	100,0% (68)	100,0% (120)	100,0% (40)	100,0% (32)	100,0% (72)
		$r_{XY} = -0,42$					
		$r_{XY} = 0,00$					

ria di esse, inoltre, è definita **tabella di primo ordine** in quanto include una sola variabile di controllo.

2.2. Effetto nullo della variabile di controllo

La tabella 7.2 (così come le tabb. 7.3, 7.4 e 7.5) è stata ottenuta semplicemente dividendo la tavola 2×2 originale in due *sottotavole* aventi la stessa forma ma frequenze di cella diverse. Le frequenze totali (N) delle due sottotavole mostrano che circa due terzi degli adolescenti intervistati non dispongono in modo regolare di un'automobile. Ognuna delle due sottotavole è omogenea in termini della variabile di controllo: i 132 adolescenti della prima sottotavola sono tutti privati dell'uso regolare di un'automobile, mentre i 60 adolescenti della seconda sottotavola sono tutti in grado di usare regolarmente un'auto. Dunque, ogni sottotavola «mantiene costante» la variazione nel terzo fattore, permettendoci di vedere cosa succede alla covariazione fra le due variabili principali.

Nel caso della tabella 7.2, si può constatare che dopo l'introduzione nell'analisi della variabile di controllo la relazione fra religiosità della famiglia e attività sessuale prematrimoniale rimane invariata. In entrambe le sottotavole risulta che gli adolescenti appartenenti alle famiglie più religiose hanno meno probabilità di avere rapporti sessuali prematrimoniali. Oltre alla direzione, anche la forza della relazione appare invariata: le percentuali di colonna delle due sottotavole sono sostanzialmente identiche, così come sono molto simili i corrispondenti coefficienti di correlazione ($-0,27$ fra coloro che non usano regolarmente l'auto e $-0,25$ fra coloro che la usano regolarmente). Questi ultimi sono chiamati **coefficienti di correlazione condizionati** in quanto si riferiscono a correlazioni osservate in certe condizioni della variabile di controllo.

Poiché in questo primo caso le relazioni di primo ordine sono uguali fra loro e alla relazione di ordine zero originale, possiamo concludere che l'opportunità di avere rapporti sessuali – almeno nel modo in cui è stata operativizzata – *non esercita alcun effetto* sulla covariazione bivariata. Quando si giunge a un risultato di questo tipo bisogna cercare spiegazioni differenti del legame osservato fra le due variabili principali.

Un modo alternativo di analizzare la tabella 7.2 consiste nel calcolare gli odds-ratio delle due sottotavole e di confrontarli fra loro e con l'odds-ratio calcolato per la tavola 2×2 originale. Se la disponibilità di un'automobile non esercita alcun effetto sulla relazione fra la religiosità familiare e l'attività sessuale prematrimoniale, allora gli odds-ratio relativi alle due sottotavole della tabella 7.2 risulteranno sostanzialmente uguali fra loro e all'odds-ratio della tabella 7.1. Quest'ultimo risulta pari a $(36)(84)/(16)(56) = 3,38$, mentre i due odds-ratio relativi

alla tabella 7.2 sono uguali rispettivamente a $(25)(58)/(11)(38) = 3,47$ e a $(11)(26)/(5)(18) = 3,18$. Anche questi valori confermano che, *indipendentemente dalla possibilità di usare regolarmente un'automobile*, coloro che appartengono a famiglie poco o per nulla religiose hanno una propensione ad avere rapporti sessuali prematrimoniali pari a oltre tre volte la propensione di coloro che, invece, provengono da famiglie molto religiose.

2.3. Effetto parziale della variabile di controllo

La tabella 7.3 illustra un altro possibile caso. Introducendo la variabile di controllo, l'associazione fra religiosità della famiglia e attività sessuale prematrimoniale risulta avere la stessa intensità nelle due sottotavole. Questa intensità, però, è più debole di quella osservata nella tavola originale. I coefficienti di correlazione condizionati alle due sottotavole della tabella 7.3, infatti, sono rispettivamente uguali a $-0,16$ e $-0,17$. Dunque, in questo caso possiamo concludere che l'opportunità – così come l'abbiamo operativizzata – spiega parzialmente la covariazione osservata fra le due variabili di interesse.

Per comprendere meglio il significato di una *spiegazione parziale* è utile osservare le distribuzioni percentuali di ciascuna sottotavola. Fra gli adolescenti che *non* appartengono a famiglie molto religiose, coloro che usano regolarmente un'auto hanno il doppio di probabilità di avere esperienze sessuali rispetto a coloro che non dispongono dell'auto ($53,3\%$ contro $25,5\%$). Analogamente, fra gli adolescenti che appartengono a famiglie molto religiose coloro che possono disporre di un'automobile hanno più del doppio delle probabilità di avere rapporti sessuali rispetto a coloro che non usano l'auto ($33,3\%$ contro $12,9\%$). Tuttavia, la disponibilità di un'automobile non elimina *completamente* la differenza osservata in termini di esperienza sessuale fra gli adolescenti appartenenti a famiglie molto religiose e gli adolescenti appartenenti a famiglie poco o per nulla religiose: sebbene in modo meno intenso, *a parità di possibilità di usare regolarmente un'automobile* gli adolescenti che appartengono a famiglie molto religiose continuano ad avere minori probabilità di avere rapporti sessuali rispetto ai loro coetanei provenienti da famiglie meno (o per nulla) religiose. In questo caso, dunque, potremmo concludere che 1) l'opportunità spiega una *parte* dell'associazione fra religiosità della famiglia e attività sessuale prematrimoniale; 2) per interpretare la parte di covariazione rimasta senza spiegazione bisognerebbe introdurre nell'analisi ulteriori variabili di controllo.

2.4. Spiegazione completa della variabile di controllo

La tabella 7.4 illustra il caso in cui la variabile di controllo spiega completamente l'associazione osservata nella tavola di ordine zero. Come si può vedere, in entrambe le sottotavole la correlazione fra le due variabili principali è esattamente uguale a zero, indicando che quando si mantiene costante il livello di opportunità (cioè la disponibilità di un'auto) ogni differenza fra adolescenti in termini di probabilità di avere rapporti sessuali scompare. Fra coloro che non dispongono di un'automobile, solo l'11,1% hanno avuto esperienze sessuali, contro il 66,7% di coloro che, invece, possono usare regolarmente un'auto. Il motivo per cui nella tavola di ordine zero emerge una forte relazione inversa fra religiosità della famiglia e attività sessuale prematrimoniale risulta evidente dall'esame delle distribuzioni marginali delle due sottotavole della tabella 7.4: più della metà degli adolescenti provenienti da famiglie poco o per nulla religiose usano regolarmente un'automobile, ma solo un adolescente di famiglia molto religiosa su otto dispone di un'auto.

In questo caso, dunque, possiamo concludere che l'opportunità spiega tutta la covariazione osservata nella relazione bivariata originale. In altri termini, la religiosità della famiglia determina la possibilità dell'adolescente di disporre di un'automobile in modo regolare; a sua volta, questa possibilità influisce fortemente sulla probabilità di avere rapporti sessuali prematrimoniali.

Per quale motivo l'introduzione della variabile di controllo spiega completamente la relazione fra religiosità della famiglia e attività sessuale prematrimoniale nella tabella 7.4 ma non nella tabella 7.2? La risposta è relativamente semplice: nella tabella 7.2 la variabile di controllo W è virtualmente indipendente sia dalla variabile indipendente X che dalla variabile dipendente Y . Di conseguenza, tenendo costante W la relazione fra X e Y non può subire alcuna modifica. Al contrario, nella tabella 7.4 W è fortemente correlata sia con Y ($r_{WY} = 0,60$) che con X ($r_{WX} = 0,43$).

Queste osservazioni sono molto importanti e ci portano a formulare un principio generale: una terza variabile può spiegare la covariazione fra altre due variabili solo se è correlata con entrambe. Se l'associazione bivariata originale è positiva, la variabile di controllo deve essere associata positivamente sia con la variabile indipendente che con la variabile dipendente. Se, invece, l'associazione bivariata è negativa, la variabile di controllo deve essere associata positivamente con una delle due variabili principali e negativamente con l'altra (come nell'esempio della tab. 7.4).

È importante notare che i dati illustrati nella tabella 7.4 si potrebbero osservare anche nel caso in cui la covariazione fra le due variabili principali fosse spuria. In altri termini, l'introduzione della variabile di controllo riduce a zero l'associazione di primo ordine fra le due varia-

bili di interesse sia quando la corrispondente associazione di ordine zero è interpretabile, sia quando è spuria. La decisione di considerare una data relazione come spuria o interpretabile dipende dal ruolo causale assegnato alla variabile di controllo. Nel nostro esempio abbiamo considerato la possibilità di usare regolarmente un'automobile come una variabile interveniente, cioè come una variabile che è nello stesso tempo un effetto della religiosità familiare e una causa dell'attività sessuale prematrimoniale. Se la variabile di controllo fosse stata un'altra, avremmo potuto considerare spuria la relazione di ordine zero fra le due variabili principali. Ad esempio, se i risultati illustrati dalla tabella 7.4 fossero stati ottenuti tenendo sotto controllo la struttura familiare (distinguendo fra famiglie «intatte» e famiglie «disgregate»), probabilmente avremmo preferito descrivere la relazione bivariata fra religiosità della famiglia e attività sessuale prematrimoniale come spuria. Le famiglie intatte, infatti, tendono ad essere più restrittive nei confronti di ogni forma di comportamento individuale – religioso, sessuale, e così via – di quanto lo siano le famiglie disgregate. In questo caso, dunque, la relazione bivariata fra le due variabili di interesse rifletterebbe non una relazione causale, bensì un modello di moralità più generale derivante dalla situazione familiare.

Con queste osservazioni vogliamo sottolineare il fatto che i risultati dell'analisi statistica non possono dire nulla sul carattere spurio o interpretabile di una data relazione bivariata. Il ruolo causale delle variabili di controllo può essere stabilito solo basandosi sulla teoria sociale e sui risultati delle ricerche empiriche precedenti.

2.5. Effetto di interazione della variabile di controllo

La tabella 7.5 illustra un'ultima situazione che si verifica con una certa frequenza quando si introduce una variabile di controllo nella relazione bivariata di interesse. Come possiamo vedere, la correlazione di primo ordine fra religiosità della famiglia e attività sessuale prematrimoniale varia in modo sostanziale a seconda del valore assunto dalla variabile di controllo: fra gli adolescenti che usano regolarmente un'automobile la correlazione è nulla (cioè pari a zero), mentre fra coloro che non dispongono di un'auto la correlazione è negativa ed equivale a $-0,42$. Il calcolo degli odds-ratio rende ancora più chiaro l'effetto di interazione. Fra gli adolescenti che usano l'auto l'odds-ratio è uguale a 1: ciò significa che la religiosità della famiglia di appartenenza non esercita alcun effetto sulla probabilità di avere rapporti sessuali prematrimoniali. Al contrario, fra coloro che non hanno la disponibilità di un'automobile l'odds-ratio assume un valore pari a 7,5: ciò significa che la propensione degli adolescenti di famiglie poco o per nulla religiose ad avere esperienze sessuali prematrimoniali è 7,5 volte l'analoga propensione degli adolescenti che appartengono a fami-

glic molto religiose. Dunque, tenendo costante la variabile di controllo l'associazione fra le due variabili di interesse diminuisce in una sottotavola ma aumenta nell'altra. A volte si può perfino verificare che la relazione assuma direzioni opposte.

Nei casi come questo in cui le relazioni osservate nelle diverse sottotavole risultano diverse, è presente un **effetto di interazione**. In altri termini, si ha un effetto di interazione quando *l'associazione fra due variabili varia secondo il valore della variabile di controllo*. In questi casi non è possibile descrivere l'effetto della variabile di controllo in maniera sintetica, ma bisogna fare esplicito riferimento alle singole sottotavole.

La scoperta di un'interazione spesso rappresenta l'inizio di ulteriori analisi. È compito del ricercatore spiegare perché le due variabili di interesse sono associate in modo diverso a seconda del valore assunto dalla terza variabile.

2.6. Effetti condizionati: una visione di insieme

Poiché i diagrammi offrono spesso un'immagine più chiara dei dati, i risultati ottenuti analizzando le tabelle 7.2-7.5 sono illustrati graficamente nella figura 7.2. Innanzitutto, il riquadro *a* mostra il caso in cui la variabile di controllo non esercita alcun effetto sulla relazione bivariata originale; questo effetto nullo è evidenziato con chiarezza dal fatto che le rette associate alle due categorie della variabile di controllo sono sostanzialmente identiche, cioè quasi sovrapposte. Il riquadro *b*, invece, mostra il caso in cui la propensione all'attività sessuale prematrimoniale è influenzata *sia* dalla religiosità della famiglia (le due rette sono inclinate), *sia* dalla disponibilità di un'automobile (le due rette sono piuttosto distanti). In questo caso la variabile indipendente e quella di controllo influiscono sulla variabile dipendente in modo *autonomo*, nel senso che ognuna di esse fornisce un proprio contributo, indipendente da quello fornito dall'altra, alla spiegazione della variabile dipendente. Per questo abbiamo detto che la variabile di controllo esercita solo un *effetto parziale*. Il riquadro *c* illustra il caso in cui la relazione fra variabile indipendente e variabile dipendente può essere considerata spuria, in quanto l'introduzione della variabile di controllo annulla completamente la loro associazione (rendendo orizzontali le due rette). Infine, il riquadro *d* illustra la presenza di un effetto di interazione: fra coloro che dispongono di un'auto la relazione fra religiosità della famiglia e attività sessuale prematrimoniale è nulla (la retta corrispondente è orizzontale), mentre fra gli adolescenti che non usano l'automobile tale relazione è negativa (la retta corrispondente è inclinata).

La figura 7.2 evidenzia un punto importante: *se non è presente alcun effetto di interazione, la relazione fra la variabile indipendente e quella dipendente non varia al variare del valore della variabile di con-*

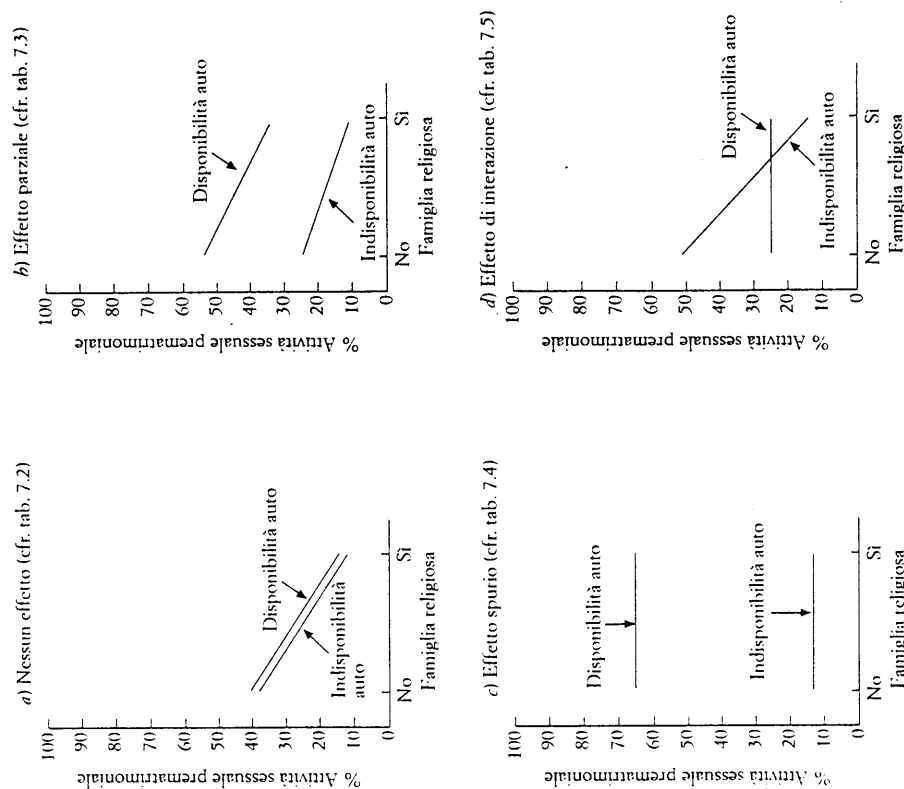


Fig. 7.2. Rappresentazione grafica degli effetti condizionati illustrati dalle tabelle 7.2-7.5.

trollo. Per questo motivo le rette associate alle due categorie della variabile di controllo sono parallele nei riquadri *a*, *b* e *c*, mentre *non* sono parallele nel riquadro *d* che, appunto, illustra la presenza di un effetto di interazione.

È importante osservare che gli esempi discussi sopra sono stati costruiti appositamente per illustrare in modo «ideale» le diverse situazioni che il ricercatore si può trovare ad affrontare. Nelle analisi dei dati reali, tuttavia, spesso i risultati sono più ambigui e meno facilmente interpretabili. Soprattutto, non si verifica mai che l'introduzione di

zione fra X e Y . Il coefficiente di correlazione parziale $r_{XY \cdot W}$, allora, sarà certamente minore del coefficiente di ordine zero r_{XY} ma anche maggiore di zero.

Il coefficiente di correlazione parziale al quadrato esprime la quantità di varianza di Y «spiegata» da X tenendo sotto controllo W . Se W non è correlata con X o con Y , allora r_{XY}^2 e $r_{XY \cdot W}^2$ saranno virtualmente identici. Se, invece, W è correlata tanto con X quanto con Y , allora $r_{XY \cdot W}^2$ sarà minore di r_{XY}^2 . In altri termini, la quantità di varianza di Y che X è in grado di spiegare diminuisce quando W viene inserita nell'analisi.

3.1. Un'applicazione pratica

Per mostrare l'uso del coefficiente di correlazione parziale prendremo in considerazione le relazioni esistenti fra tre variabili, tratte dalla General Social Survey del 1991: l'abitudine a bere in locali pubblici, l'atteggiamento nei confronti dei rapporti sessuali prematrimoniali e il genere. Il riquadro A della tabella 7.6 mostra la relazione di ordine zero fra l'abitudine a frequentare locali pubblici (come bar e taverne) in cui si bevono alcolici (X) e l'atteggiamento nei confronti dei rapporti sessuali prematrimoniali (Y). Come si può vedere, queste due variabili sono fortemente correlate. Supponiamo di fissare $\alpha = 0,05$. Il coefficiente di correlazione di ordine zero è $r_{XY} = -0,393$, con $t = -12,94$. Pertanto, possiamo rifiutare l'ipotesi nulla secondo la quale nella popolazione $\rho_{XY} = 0$, e concludere che coloro che non si recano mai nei locali pubblici per bere hanno una maggiore propensione a considerare sbagliati i rapporti sessuali prematrimoniali. Forse, però, le donne sono intolleranti rispetto a entrambe le pratiche.

Per verificare questa possibilità introduciamo nell'analisi il genere (W) come variabile di controllo. Il riquadro B della tabella 7.6 mostra che $r_{Y \cdot W} = 0,122$, indicando che le donne sono più propense degli uomini a considerare sbagliati i rapporti sessuali prematrimoniali. Osservando il riquadro C, invece, possiamo vedere che $r_{XY \cdot W} = -0,221$; ciò significa che le donne frequentano i bar meno degli uomini (entrambi i coefficienti di correlazione sono significativi al livello $\alpha = 0,05$).

Disponendo di questi dati, possiamo calcolare il coefficiente di correlazione parziale come segue:

$$r_{XY \cdot W} = \frac{-0,393 - (-0,221)(0,122)}{\sqrt{1 - (-0,221)^2} \sqrt{1 - (0,122)^2}} = -0,378$$

Dunque, tenendo sotto controllo il genere, la relazione di ordine zero fra l'abitudine a bere nei locali pubblici e l'atteggiamento nei confronti dei rapporti sessuali prematrimoniali diminuisce in modo

Tab. 7.6. Relazioni di ordine zero fra abitudine a bere in locali pubblici (X), atteggiamento nei confronti dei rapporti sessuali prematrimoniali (Y) e genere (W)

Riquadro A

Abitudine a bere in locali pubblici					
Atteggiamento nei confronti delle esperienze sessuali prematrimoniali	Sì			Totale	
	No				
	Sì				
	Sbagliate				
	Non sbagliate				
	55,5% (288)	17,4% (82)		37,4% (370)	
	44,5% (231)	82,6% (388)		62,6% (619)	
Totale	100,0% (519)	100,0% (470)		100,0% (989)	
$r_{xy} = -0,393$					

Riquadro B

Atteggiamento nei confronti delle esperienze sessuali prematrimoniali	Genere			Totale	
	Uomini				
	Donne				
	Sbagliate				
	Non sbagliate				
	30,4% (125)	42,4% (245)		37,4% (370)	
	69,6% (286)	57,6% (333)		62,6% (619)	
Totale	100,0% (411)	100,0% (578)		100,0% (989)	
$r_{xy} = -0,122$					

Riquadro C

Abitudine a bere in locali pubblici	Genere			Totale	
	Uomini				
	Donne				
	Sì				
	No				
	60,6% (249)	38,2% (221)		47,5% (470)	
	39,4% (162)	61,8% (357)		52,5% (519)	
Totale	100,0% (411)	100,0% (578)		100,0% (989)	
$r_{xy} = -0,221$					

molto limitato, passando da $-0,393$ a $-0,378$. Questo risultato ci porta a concludere che la relazione fra le nostre due variabili principali non è «spiegata» dal genere.

3.2. Test di significatività per il coefficiente di correlazione parziale

Per verificare la significatività statistica del coefficiente di correlazione parziale dobbiamo innanzitutto formulare l'ipotesi nulla H_0 : $\rho_{XY \cdot W} = 0$ e l'ipotesi alternativa H_1 : $\rho_{XY \cdot W} < 0$. Quest'ultima implica che ci aspettiamo una relazione negativa.

La statistica del test è la seguente:

glie molto religiose. Dunque, tenendo costante la variabile di controllo l'associazione fra le due variabili di interesse diminuisce in una sottotavola ma aumenta nell'altra. A volte si può perfino verificare che la relazione assuma direzioni opposte.

Nei casi come questo in cui le relazioni osservate nelle diverse sottotavole risultano diverse, è presente un **effetto di interazione**. In altri termini, si ha un effetto di interazione quando *l'associazione fra due variabili varia secondo il valore della variabile di controllo*. In questi casi non è possibile descrivere l'effetto della variabile di controllo in maniera sintetica, ma bisogna fare esplicito riferimento alle singole sottotavole.

La scoperta di un'interazione spesso rappresenta l'inizio di ulteriori analisi. È compito del ricercatore spiegare perché le due variabili di interesse sono associate in modo diverso a seconda del valore assunto dalla terza variabile.

2.6. Effetti condizionati: una visione di insieme

Poiché i diagrammi offrono spesso un'immagine più chiara dei dati, i risultati ottenuti analizzando le tabelle 7.2-7.5 sono illustrati graficamente nella figura 7.2. Innanzitutto, il riquadro *a* mostra il caso in cui la variabile di controllo non esercita alcun effetto sulla relazione bivariata originale; questo effetto nullo è evidenziato con chiarezza dal fatto che le rette associate alle due categorie della variabile di controllo sono sostanzialmente identiche, cioè quasi sovrapposte. Il riquadro *b*, invece, mostra il caso in cui la propensione all'attività sessuale prematrimoniale è influenzata *sia* dalla religiosità della famiglia (le due rette sono inclinate), *sia* dalla disponibilità di un'automobile (le due rette sono piuttosto distanti). In questo caso la variabile indipendente e quella di controllo influiscono sulla variabile dipendente in modo *autonomo*, nel senso che ognuna di esse fornisce un proprio contributo, indipendente da quello fornito dall'altra, alla spiegazione della variabile dipendente. Per questo abbiamo detto che la variabile di controllo esercita solo un *effetto parziale*. Il riquadro *c* illustra il caso in cui la relazione fra variabile indipendente e variabile dipendente può essere considerata spuria, in quanto l'introduzione della variabile di controllo annulla completamente la loro associazione (rendendo orizzontali le due rette). Infine, il riquadro *d* illustra la presenza di un effetto di interazione: fra coloro che dispongono di un'auto la relazione fra religiosità della famiglia e attività sessuale prematrimoniale è nulla (la retta corrispondente è orizzontale), mentre fra gli adolescenti che non usano l'automobile tale relazione è negativa (la retta corrispondente è inclinata).

La figura 7.2 evidenzia un punto importante: *se non è presente alcun effetto di interazione, la relazione fra la variabile indipendente e quella dipendente non varia al variare del valore della variabile di con-*

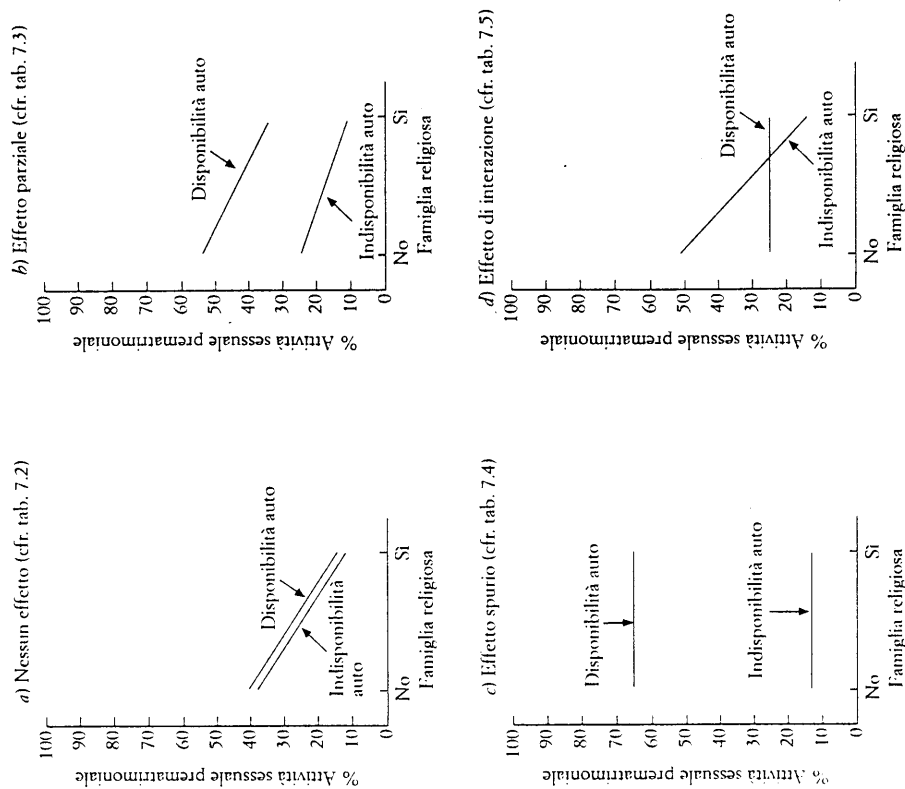


Fig. 7.2. Rappresentazione grafica degli effetti condizionati illustrati dalle tabelle 7.2-7.5.

trollo. Per questo motivo le rette associate alle due categorie della variabile di controllo sono parallele nei riquadri *a*, *b* e *c*, mentre *non* sono parallele nel riquadro *d* che, appunto, illustra la presenza di un effetto di interazione.

È importante osservare che gli esempi discussi sopra sono stati costruiti appositamente per illustrare in modo «ideale» le diverse situazioni che il ricercatore si può trovare ad affrontare. Nelle analisi dei dati reali, tuttavia, spesso i risultati sono più ambigui e meno facilmente interpretabili. Soprattutto, non si verifica mai che l'introduzione di

$$F_{1,N-3} = \frac{r_{XY,W}^2(N-3)}{1-r_{XY,W}^2}$$

In alternativa, estraendo la radice quadrata di entrambi i lati dell'equazione si ottiene:

$$t_{N-3} = \frac{r_{XY,W}\sqrt{N-3}}{\sqrt{1-r_{XY,W}^2}}$$

Utilizzando i dati dell'esempio discusso nel paragrafo precedente e ponendo $\alpha = 0,05$ otteniamo i seguenti risultati:

$$F_{1,986} = \frac{(-0,393)^2(989-3)}{1-(-0,393)^2} = 180,1$$

ovvero

$$t_{986} = -13,42$$

Poiché in questo caso il valore critico del test è uguale a $-1,65$, possiamo tranquillamente rifiutare l'ipotesi nulla e concludere che, anche tenendo sotto controllo il genere, nella popolazione esiste una correlazione negativa fra abitudine a bere nei locali pubblici e atteggiamento nei confronti dei rapporti sessuali prematrimoniali.

Problemi

1. Le seguenti ipotesi potrebbero essere formulate da un sociologo della famiglia:

- i figli maschi di famiglie disgregate sono sottoposti a un controllo meno intenso rispetto ai figli maschi di famiglie integre;
- il numero di azioni illegali commesse dai giovani maschi diminuisce all'aumentare dell'intensità del controllo familiare al quale essi sono sottoposti.

Quale tipo di relazione causale è espresso da queste ipotesi? Rappresenta tale relazione sotto forma di diagramma.

2. Considera la popolazione delle donne sposate. Suggestisci il motivo per cui, all'interno di questa popolazione, una covarianza positiva fra partecipazione alla forza lavoro retribuita e probabilità di divorziare potrebbe essere resa spuria dal numero di figli presenti in

famiglia. Rappresenta la tua ipotesi sotto forma di diagramma. Supponendo che l'ipotesi secondo la quale la relazione fra partecipazione alla forza lavoro e probabilità di divorzio è spuria sia vera, cosa pensi che succederebbe al corrispondente coefficiente di correlazione se si tenesse costante il numero di figli?

3. Suggestisci tre possibili processi intervenienti che «interpretino» la relazione causale fra grado di dipendenza delle nazioni dall'esportazione di un singolo prodotto grezzo e probabilità di essere governati da una dittatura militare. Rappresenta queste relazioni sotto forma di diagrammi causali.

4. L'incrocio fra genere, livello salariale e tipo di impiego dà luogo alla seguente tavola di contingenza $2 \times 2 \times 2$:

Salario	Alto	Basso	Tempo pieno		Tempo parziale	
			Uomini	Donne	Uomini	Donne
	50	20	10	15		
	20	20	10	25		

Calcola il coefficiente di correlazione e l'odds-ratio per la relazione di ordine zero fra genere e livello salariale. Quale tipo di relazione emerge quando si tiene sotto controllo il tipo di impiego?

5. Uno studioso della sessualità ipotizza che gli uomini guardano i film pornografici più spesso delle donne perché sono molto meno propensi a ritenere che questi film portino a un indebolimento del comportamento morale. I dati tratti dalla General Social Survey del 1991 danno luogo alla seguente tavola di contingenza:

Lei ha visto qualche film pornografico nel corso dell'ultimo anno?		Il materiale pornografico provoca un indebolimento della morale?	
		Sì	No
Uomini	Donne	Uomini	Donne
Sì	No	Sì	No
46	48	90	46
162	347	92	112

Ricostruisci la tavola di ordine zero che incrocia il genere con la domanda sulla visione dei film pornografici, calcola i coefficienti di correlazione e gli odds-ratio relativi alle due sottotavole e determina se l'ipotesi iniziale è corretta. Quale tipo di relazione riesci a vedere?

6. Nella tabella che segue, il riquadro A rappresenta la tavola di ordine zero che incrocia l'età con l'atteggiamento nei confronti della legalizzazione della marijuana, mentre il riquadro B rappresenta una delle due sottotavole ottenute inserendo nell'analisi la variabile di controllo «Livello di religiosità».

A. Tavola di ordine zero

Favorevole alla legalizzazione della marijuana?	Giovani		Anziani	
	Sì	No	Sì	No
	17	28	14	31

B. Sottotavola relativa agli individui molto religiosi

Giovani		Anziani	
Sì	No	Sì	No
10	25	7	18

Calcola il coefficiente di correlazione di ordine zero, i due coefficienti di correlazione condizionati e i corrispondenti odds-ratio. Quale tipo di effetto si osserva quando il livello di religiosità è tenuto costante?

7. Per verificare la possibilità che i contatti interrazziali siano una funzione delle opportunità mediate dagli atteggiamenti, agli intervistati bianchi della General Social Survey del 1991 è stato chiesto: *a)* se nel loro vicinato risiedono persone di colore; *b)* se negli ultimi anni qualche membro della famiglia ha invitato a casa per cena un amico di colore. La relazione di ordine zero fra queste due variabili è la seguente:

Vicini di colore?	Persone di colore a casa per cena?	
	No	Sì
Sì	95	143
No	40	133

Tenendo sotto controllo le opinioni degli intervistati sulla quantità di denaro speso dal paese per risolvere i problemi delle persone di colore (troppo o troppo poco), si ottiene la seguente tavola $2 \times 2 \times 2$:

Vicini di colore?	Spesa eccessiva		Spesa insufficiente	
	Persone di colore a casa per cena?		Persone di colore a casa per cena?	
	No	Sì	No	Sì
Sì	10	26	85	117
No	5	29	35	104

Ritieni che questi risultati sostengano l'ipotesi secondo la quale la relazione fra vicinanza e comportamento è completamente mediata da un atteggiamento interveniente?

8. Utilizzando i seguenti valori, calcola i coefficienti di correlazione parziale r_{XYZ} fra X e Y tenendo sotto controllo Z :

	r_{XY}	r_{XZ}	r_{YZ}
a)	0,30	0,40	0,70
b)	-0,45	0,20	-0,30
c)	0,15	0,20	-0,75
d)	-0,60	-0,40	-0,40

9. Il termine «omogamia» si riferisce al fatto che generalmente i matrimoni si svolgono fra persone che hanno origini sociali simili. Il grado di omogamia, tuttavia, può variare secondo il contesto sociale. Supponi di trovare la seguente relazione:

Religione della moglie	Origine urbana		Origine rurale	
	Protestante	Cattolica	Protestante	Cattolica
Protestante	30	30	10	80
Cattolica	30	20	20	5

Sulla base di questi dati:

a) Calcola: 1) la correlazione di ordine zero fra le religioni dei due sposi; 2) la correlazione fra le religioni nelle due sottotavole; 3) la correlazione parziale fra le religioni, tenendo costante l'origine; 4) la statistica t per la significatività del coefficiente di correlazione parziale ($\alpha = 0,001$).

b) L'origine urbana/rurale quale tipo di effetto esercita sull'omogamia?

10. La relazione fra la dimensione della famiglia (numero di figli) e la soddisfazione delle donne per la vita familiare può dipendere dall'età alla quale esse hanno cominciato ad avere figli. Osserva la seguente tavola di contingenza:

Soddisfazione per la vita familiare	Primi figli durante l'adolescenza		Primi figli più tardi	
	Dimensione della famiglia		Dimensione della famiglia	
	Piccola	Media	Piccola	Media
Alta	15	10	60	30
Media	5	15	20	15
Bassa	5	5	20	15

Analizza questi dati calcolando: a) la correlazione di ordine zero fra dimensione della famiglia e soddisfazione per la vita familiare; b) le due correlazioni di primo ordine; c) la correlazione parziale fra dimensione della famiglia e soddisfazione per la vita familiare, tenendo costante l'età alla quale è stato avuto il primo figlio; d) la statistica t per la significatività della correlazione parziale ($\alpha = 0,05$). Nota: nel calcolare le correlazioni assegna valore 0 alle categorie «bassa» e «piccola», valore 1 alla categoria «media» e valore 2 alle categorie «alta» e «grande».