

4.3. Conclusioni su misure di associazione e di cograduazione

Malgrado la trattazione piuttosto ampia che ne abbiamo dato, dobbiamo dire che le misure di forza delle relazioni fra variabili nominali ed ordinali non sono molto utilizzate nella ricerca sociale. Ciò per almeno tre motivi. Innanzitutto la loro pluralità non ne ha facilitato la diffusione e popolarità: non c'è un unico indice, dal significato chiaro e intuitivo, da poter utilizzare in tutte le situazioni, e questo perché nessuna delle misure proposte è pienamente soddisfacente (la loro numerosità deriva appunto da questa insoddisfazione che ha generato una ricerca continua di nuove misure).

A ciò va aggiunto il fatto che quando le variabili sono nominali può avere poco senso calcolare un'unica misura sintetica di associazione, in quanto può darsi – come abbiamo già visto per il chi-quadrato – che la forza della relazione sia dovuta al comportamento specifico di alcune modalità. È abbastanza inutile per esempio misurare la forza della relazione fra orientamento religioso e voto politico quando questa associazione è dovuta al voto della maggioranza dei cattolici per un determinato partito, mentre i cattolici restanti si distribuiscono uniformemente sugli altri partiti. In tal caso sarebbe più opportuno parlare di relazione fra orientamento religioso ed una specifica modalità della variabile «partito votato» (questa obiezione non vale se le variabili considerate sono entrambe dicotomiche – tabella 2×2 – oppure se sono entrambe ordinali).

Infine queste misure sono di difficile interpretazione, soprattutto se confrontate col chiaro significato che ha per la semplice lettura delle percentuali della tabella. Prendiamo ad esempio il caso, fra quelli trattati, della relazione fra età e pratica religiosa. In tabella 3.15 avevamo visto che la percentuale di praticanti era del 24,4% fra i giovani, del 27,6% fra gli adulti e del 41,5% fra gli anziani. Questi dati sono facilmente comprensibili e mostrano l'esistenza di una relazione fra le due variabili che è positiva e che tuttavia non è lineare (lo scarto fra adulti ed anziani è ben maggiore di quello che esiste fra giovani ed adulti, e questo offre lo spunto a diverse interpretazioni). Il test del chi-quadrato ci aveva poi detto che la relazione fra le due variabili era molto significativa ($p < 0,001$). A questi elementi cosa hanno aggiunto le misure di associazione? Praticamente nulla. In tabella 3.18 ave-

vamo riportato le varie misure di forza della relazione: ma come il lettore potrà constatare si tratta di valori tutti assai piccoli e soprattutto che dicono assai meno, quando è possibile dare loro un'interpretazione, di quanto non dica la lettura delle percentuali della tabella (sapere che $gamma = 0,124$ e quindi che il fatto di conoscere l'età dei soggetti permette di fare il 12% in meno di errore nel prevederne la pratica religiosa è francamente un'informazione assai meno ricca di quella che ricaviamo dalla lettura delle percentuali di praticanti per i tre diversi livelli di titolo di studio).

Nulla quindi può sostituire l'attenta lettura della tabella e delle percentuali non solo per interpretare la relazione ma anche per valutarne la forza; e spesso semplici differenze fra percentuali sono più informative di sofisticate misure di associazione.

5. RAPPORTI DI PROBABILITÀ

Concludiamo questa parte sulle tavole di contingenza presentando un diverso modo di guardare alle frequenze delle celle al fine di interpretare la relazione fra due variabili categoriali. Il lettore potrà trovare questo paragrafo come un'inutile complicazione di quanto abbiamo già visto nei paragrafi 2 e 4. Di fatto l'interpretazione della relazione fra due variabili in termini di rapporti di probabilità (*odds*)²³ ha assunto sempre maggiore importanza negli ultimi anni, in quanto le tecniche di analisi multivariata dei dati categoriali generano, come prodotto conclusivo, proprio dei rapporti di probabilità che il ricercatore sociale (anche alle prime armi) deve saper interpretare.

Data una variabile, abbiamo visto che il metodo tradizionale per esprimere il diverso peso esistente fra le sue categorie è quello di ricorrere alle proporzioni (o al loro equivalente, le percentuali). Avendo per esempio 1.188 favorevoli e 604 contrari alla pena di morte (tab. 3.20), comunemente si esprime lo squilibrio numerico fra le due categorie dicendo che i primi sono il 66,3% degli intervistati (da cui consegue, essendo la variabile dicotomica, che i secondi ne rappresentano il 33,7%); dove il valore 66,3%, esprimibile anche con 0,663 assumendo a totale l'unità invece che il valore 100, deriva dal rapporto fra i favorevoli ed il totale degli intervistati ($1.188/1.792 = 0,663$).

Ma si potrebbe rappresentare lo squilibrio fra favorevoli e contrari an-

TAB. 3.20. Proporzioni e rapporti di probabilità (relazione fra istruzione e atteggiamento verso la pena di morte)

ATEGGIAMENTO	ISTRUZIONE		
	Inferiore	Superiore	Totale
	<i>a</i>	<i>b</i>	
Favorevoli	1.027	161	1.188
Contrari	<i>c</i>	<i>d</i>	
	397	207	604
Totale	1.424	368	1.792

Proporzione: $p = \frac{1.188}{1.792} = 0,663$ Rapporto di probabilità: $\omega = \frac{1.188}{604} = 1,97$

Proporzioni condizionate: $p_1 = \frac{1.027}{1.424} = 0,721$ $p_2 = \frac{161}{368} = 0,438$

Rapporti di probabilità condizionali: $\omega_1 = \frac{1.027}{397} = 2,59$ $\omega_2 = \frac{161}{207} = 0,77$

Rapporto di associazione (odds ratio): $\omega_1 = \frac{1.027 \cdot 207}{397 \cdot 161} = 3,3$

che facendo il rapporto fra gli appartenenti alle due categorie, ottenendo in questo caso $1.188/604 = 1,97$, esprimibile verbalmente dicendo che ci sono quasi 2 favorevoli per ogni persona contraria.

Mentre chiamavamo *proporzione* il rapporto fra la parte ed il tutto, chiamiamo questo rapporto fra la frequenza di una categoria e la frequenza della categoria alternativa (nel caso di variabili dicotomiche) **rapporto di probabilità (odds)** e lo indichiamo con la lettera greca omega (ω). Esso è anche definibile come il rapporto fra la probabilità che un individuo, estratto a caso dalla popolazione, appartenga ad una categoria della variabile considerata e la probabilità che non vi appartenga (da cui il suo nome italiano di «rapporto di probabilità»).

Rapporto di probabilità (odds):

$$\omega = \frac{f_1}{f_2} = \frac{p_1}{1 - p_1}$$

I rapporti di probabilità non vengono normalmente utilizzati nella pratica quotidiana (contro il comunissimo uso delle percentuali), salvo un'applicazione assai particolare nel campo delle scommesse ed in particolare nel le corse dei cavalli. *Odds* in inglese è il rapporto fra l'eventuale vincita e la cifra scommessa. Se un cavallo è dato 3 a 1, significa che, in caso che vinca, la cifra scommessa viene resa triplicata allo scommettitore, e il rapporto 3 : 1 è per l'appunto l'*odds*.

Il passaggio dalla proporzione al rapporto di probabilità (e viceversa) è molto semplice:

$$\omega = \frac{p}{1 - p} \quad p = \frac{\omega}{1 + \omega}$$

Il rapporto di probabilità assume il valore 1 quando le due categorie della variabile hanno lo stesso peso (equivalente alla proporzione di 0,5 per entrambe); ha come valore minimo il valore 0, ma non ha un limite superiore.

Il rapporto di probabilità che abbiamo fin qui presentato fa riferimento a due categorie di una stessa variabile dicotomica, ma può essere esteso anche al caso di *due variabili*. Data una tavola di contingenza tra due variabili dicotomiche (sempre la stessa tab. 3.20), siano queste l'istruzione (X) e l'atteggiamento verso la pena capitale (Y), abbiamo visto che il metodo più comunemente utilizzato per studiare la relazione fra le variabili fa anche in questo caso ricorso alle proporzioni, e in particolare alle *proporzioni condizionate* (le percentuali di riga o di colonna presentate nel par. 2.1).

$$\text{Meno istruiti: } p_1 = \frac{a}{a + c} = \frac{1.027}{1.424} = 0,721$$

$$\text{Più istruiti: } p_2 = \frac{b}{b + d} = \frac{161}{368} = 0,438$$

Ora, invece delle proporzioni condizionate, e cioè dei rapporti fra frequenze parziali e totali (favorevoli/totale) per le due categorie di istruzione, si considerino i **rapporti di probabilità condizionali**, e cioè i rapporti favorevoli/contrari sempre per le due categorie di istruzione:

$$\text{Meno istruiti: } \omega_1 = \frac{a}{c} = \frac{1.027}{397} = 2,59$$

$$\text{Più istruiti: } \omega_2 = \frac{b}{d} = \frac{161}{207} = 0,77$$

Abbiamo tanti rapporti di probabilità condizionati per la variabile Y (atteggiamento verso la pena di morte), quante sono le categorie della variabile X (istruzione): nel nostro caso, essendo due le categorie della variabile X , abbiamo due rapporti di probabilità, uno per i meno istruiti (ω_1) ed uno per i più istruiti (ω_2).

Nell'esempio considerato, fra i meno istruiti ci sono 2,6 favorevoli alla pena capitale per ogni contrario (che si può scrivere 2,6 : 1); mentre fra i più istruiti il rapporto è circa 0,8 a 1. Se i due rapporti di probabilità condizionati fossero eguali, allora non ci sarebbe relazione fra istruzione e atteggiamento verso la pena capitale. Invece già da questi numeri si vede come il rapporto favorevoli/contrari sia maggiore fra i meno istruiti che fra i più istruiti, per cui sembra di poter dire che esiste una relazione fra le due variabili.

Questo confronto può essere formalizzato dal *rapporto fra gli odds condizionati*, che in inglese si chiama **odds ratio** (letteralmente «rapporto fra gli odds», cioè rapporto fra i rapporti di probabilità), che traduciamo con **rapporto di associazione**, pari a:

Rapporto di associazione (*odds ratio*):

$$\frac{\omega_1}{\omega_2} = \frac{a/c}{b/d} = \frac{ad}{bc} = \frac{1.027 \cdot 207}{397 \cdot 161} = 3,33$$

Il valore ottenuto può essere interpretato nel modo che segue: posto uguale a 1 il rapporto favorevoli/contrari fra i più istruiti, esso assume il valore 3,33 fra i meno istruiti. Come a dire che passando dai più istruiti ai meno istruiti il rapporto fra favorevoli e contrari passa ad oltre il triplo²⁴.

Il rapporto di associazione (*odds ratio*) fra due variabili è stato definito anche «rapporto fra i prodotti incrociati» (*cross-product ratio*), in quanto corrisponde al rapporto fra i prodotti ad e bc delle frequenze (o delle corri-

spendenti probabilità) delle celle opposte sulle due diagonali, ed è stato in questa forma utilizzato lungamente nelle scienze sociali come indice di forza della relazione fra due variabili dicotomiche.

Il rapporto di associazione fra due variabili può assumere valore compreso fra 0 e $+\infty$ (più infinito), passando per il valore 1, il quale si verifica nel caso di indipendenza fra le due variabili. Più il valore è lontano da 1, sia sopra sia sotto ad 1, maggiore è la forza della relazione.

Valori superiori ad 1 stanno a significare una associazione «positiva» fra le variabili, mentre valori inferiori ad 1 una associazione «negativa». Per associazione positiva intendiamo che i soggetti della categoria X_1 hanno probabilità di collocarsi nella categoria Y_1 maggiore di quanto sia la probabilità dei soggetti della categoria X_2 ; e viceversa. Nel nostro esempio avevamo una associazione positiva, cioè il rapporto aveva valore superiore ad 1: infatti i meno istruiti (categoria X_1) hanno probabilità maggiori dei più istruiti di collocarsi nella categoria Y_1 (favorevoli).

Se avessimo invertito l'ordine della variabile X (mettendo nella categoria 1 i più istruiti e nella 2 i meno istruiti: tab. 3.21b) oppure l'ordine della variabile Y (mettendo nella categoria 1 i contrari e nella 2 i favorevoli: tab. 3.21c), avremmo ottenuto un rapporto di associazione pari a 0,30. Un valore quindi inferiore ad 1 e, fatto che interessa sottolineare, pari a $1/3,33$ cioè all'inverso del valore precedente. Quindi, invertendo l'ordine delle righe o delle colonne, si ottiene un valore del rapporto di associazione che è l'inverso del valore originario. Mentre invece il valore del rapporto resta immutato se si cambia l'orientamento della tavola, si scambiano cioè le righe con le colonne (tab. 3.21d).

Si noti che il valore del rapporto di associazione non risente della dimensione del campione (tab. 3.21e); né cambia se entrambe le frequenze di

TAB. 3.21. Variazioni nel valore del rapporto di associazione (*odds ratio*) col variare dell'ordine delle colonne, delle righe e del numero di casi

	a)	b)	c)	d)	e)	f)
	1.027	161	161	1.027	397	207
	207	397	1.027	161	207	397
Odds ratio	3,33	0,30	0,30	3,33	3,33	3,33

una riga o di una colonna sono moltiplicate per una costante (tab. 3.21f). Questa stabilità costituisce una caratteristica importante di questa misura, e mostra come essa sia capace di *cogliere la struttura della relazione fra le due variabili*, senza risentire delle variazioni campionarie.

Abbiamo applicato finora i rapporti di probabilità solo a variabili dicotomiche. Di fatto questa misura presuppone una logica dicotomica, in quanto il rapporto di probabilità è il rapporto fra la frequenza (o probabilità) di appartenenza ad una data categoria e la frequenza (o probabilità) di non appartenenza. Che poi questa *non* appartenenza si articoli in un'altra o in più alternative, ciò è irrilevante. Questo quando si tratta una sola variabile. Quanto alla relazione fra due variabili, cioè ai rapporti di probabilità condizionati, è la variabile dipendente che deve essere vista in un'ottica dicotomica (una modalità contro le altre). Quanto alla variabile indipendente, essa può avere anche più di due categorie: in questo caso si calcoleranno tanti rapporti di probabilità condizionati quante sono le modalità della variabile indipendente.

In tabella 3.22 abbiamo riportato la tabella originaria sulla relazione fra istruzione e pratica religiosa, nella quale l'atteggiamento verso la pena di morte era a tre modalità (favorevoli, incerti e contrari) e l'istruzione a cinque. Abbiamo calcolato i rapporti di probabilità condizionati dei favorevoli/non favorevoli (che includono incerti e contrari) per ognuna delle cinque modalità della variabile istruzione.

TAB. 3.22. Rapporti di probabilità condizionati in una tabella con variabili non dicotomiche (relazione fra atteggiamento verso la pena di morte e titolo di studio)

	SENZA TITOLO	ELEMEN- TARI	MEDIA	DIPLOMA	LAUREA
Favorevoli	262	457	309	135	26
Incerti	37	90	86	56	7
Contrari	75	145	177	156	51
$\omega = \frac{\text{favorevoli}}{\text{incerti} + \text{contrari}}$	2,33	1,94	1,17	0,64	0,45
$p = \frac{\text{favorevoli}}{\text{favorevoli} + \text{incerti} + \text{contrari}} = \frac{\omega}{1 + \omega}$	0,70	0,66	0,54	0,39	0,31

Fonte: Elaborazioni su dati tratti da CORBETTA e PARISI [1983].

Da questa tabella dovrebbe apparire con chiarezza come i rapporti di probabilità (condizionati) non sono che un modo per guardare alla relazione fra due variabili alternativo a quello che si serve invece delle proporzioni (percentuali) condizionate; ma alla fin fine i due approcci sono molto simili ed è facile passare dall'uno all'altro. Guardando alla tabella 3.22 vediamo infatti che fra le persone senza titolo di studio ci sono 2,3 favorevoli per ogni contrario; mentre fra i laureati c'è meno di un favorevole (per la precisione 0,45) per ogni contrario. Possiamo anche leggere la tabella dicendo che la probabilità di essere favorevoli rispetto alla probabilità di non esserlo decresce regolarmente passando dai titoli di studio inferiori a quelli superiori. Possiamo anche dire, per esempio, che passando da persone con licenza media a persone diplomate tale probabilità circa si dimezza (passa da 1,17 a 0,64).

Ma possiamo anche trasformare i rapporti di probabilità in proporzioni secondo la formula già vista [$p = \omega/(1 + \omega)$] e quindi dire che i favorevoli sono il 70% fra le persone senza titolo di studio e il 31% fra i laureati, e così via.

Quanto ai rapporti di associazione (*odds ratio*), essi possono essere solo calcolati su tabelle 2×2 : non possono quindi essere utilizzati per sintetizzare con un'unica misura di associazione una tabella di ordine superiore, quale quella trattata in questo esempio.

6. ANALISI DELLA VARIANZA

6.1. Principi e calcoli

Come abbiamo detto la tecnica dell'analisi della varianza (detta anche **Anova**) serve per studiare la relazione fra una variabile nominale ed una cardinale. La presentiamo con un esempio che ha come oggetto la geografia politico-elettorale dell'Italia. Come è noto il nostro paese, per via delle vicende storiche che hanno diversamente caratterizzato le varie aree geografiche fino all'unificazione e per via della differente struttura economica, presenta forti differenziazioni territoriali di cultura politica. Gli studiosi hanno proposto diverse mappe politiche, la più nota è quella proposta negli anni '60 dai ricercatori dell'Istituto Cattaneo [Galli *et al.* 1968] che, nella sua for-