

tecniche presentate nel capitolo 10. I giochi di azzardo hanno reso familiare il concetto di odds presso molte persone. In generale, un **odds** è il rapporto fra la frequenza di appartenenza a una data categoria e la frequenza di *non* appartenenza a quella categoria. Ad esempio, nel gioco del poker l'odds che la prima carta servita sia la donna di cuori è pari a 1/51 (leggi: «uno a cinquantuno»). In termini decimali, questo odds equivale a 0,01961. L'odds si distingue dalla probabilità in quanto quest'ultima è uguale al rapporto fra la frequenza associata a una data categoria di interesse e la frequenza associata a *tutte* le categorie (frequenza totale). Tornando all'esempio del poker, la probabilità che la prima carta servita sia la donna di cuori è uguale a 1/52 (leggi: «un cinquantaduesimo» oppure «una su cinquantadue»). In termini decimali, questa probabilità equivale a 0,01923, un valore più basso di quello dell'odds. La differenza fra odds e probabilità è ancora più evidente quando la proporzione di casi per categoria aumenta. Ad esempio, nel poker l'odds che la prima carta estratta sia una carta di fiori è uguale a 13/39 (cioè 1/3), mentre l'equivalente probabilità è pari a 13/52 (cioè 1/4). In termini decimali, l'odds (0,333) risulta più alto della probabilità (0,250). Un'altra importante differenza fra odds e probabilità riguarda la gamma di valori che possono assumere. Le probabilità non possono essere inferiori a zero (certezza che un dato evento *non* si verificherà) né superiori a 1 (certezza che un dato evento si verificherà). Al contrario, poiché l'odds equivale al rapporto fra la probabilità che un dato evento si verifichi e la probabilità che quell'evento *non* si verifichi, esso può assumere valori che variano fra zero e infinito positivo. Quando la probabilità che un dato evento si verifichi supera la probabilità che non si verifichi, l'odds assume valori superiori a 1. Per esempio, l'odds che la regina di cuori *non* venga estratta è uguale a 51/1, cioè 51 in termini decimali. Quando, invece, la probabilità che un dato evento si verifichi è uguale alla probabilità che non si verifichi, l'odds assume valore 1/1, cioè 1 in termini decimali. Infine, quando la probabilità che un dato evento si verifichi è minore della probabilità che non si verifichi, l'odds assume valori non negativi inferiori a 1.

Una probabilità può essere facilmente trasformata in un odds: è sufficiente dividere la proporzione associata alla categoria di interesse (p_i) per 1 meno questa proporzione:

$$\text{odds}_i = \frac{p_i}{1 - p_i}$$

Il concetto di odds semplice può essere esteso a quello di **odds condizionato**, che viene calcolato in modo analogo alle percentuali. Un odds condizionato è uguale al rapporto fra la probabilità di appartenere alla categoria di una data variabile e la probabilità di appartenere alle rimanenti categorie di quella variabile, calcolato all'in-

4. Odds e odds-ratio

Attualmente la più importante misura di associazione per le tavole di contingenza è rappresentata dall'odds ratio, che forma la base delle

terno di una specifica categoria di una seconda variabile. A titolo di esempio, torniamo alla relazione fra genere e comportamento di fumo illustrata dalla tabella 5.5. Se rivolgiamo la nostra attenzione all'ultima colonna (quella che rappresenta la distribuzione marginale della variabile di riga), possiamo vedere che l'odds (semplice) di fumare è uguale al rapporto fra il numero totale degli intervistati che fumano (319) e il numero totale degli intervistati che non fumano (697), cioè $319/697 = 0,46$. L'odds di fumare, tuttavia, cambia se teniamo sotto controllo il genere. Fra le donne, l'odds condizionato di fumare scende a $167/431 = 0,39$, mentre fra gli uomini sale a $152/266 = 0,57$. In altri termini, il genere degli intervistati *condiziona* il loro comportamento di fumo.

Quando le variabili che formano una tavola di contingenza bivariata sono correlate, i loro odds condizionati non sono uguali. Per confrontare direttamente due odds condizionati si può fare ricorso a una singola statistica descrittiva, chiamata **odds-ratio** (OR), che non è altro che il rapporto fra due odds. Utilizzando i simboli introdotti nel paragrafo precedente, la formula dell'odds-ratio per una tavola di contingenza 2×2 è la seguente:

$$OR^{xy} = \frac{b/d}{a/c} = \frac{bc}{ad}$$

Come si può vedere, l'odds-ratio equivale al rapporto fra il prodotto incrociato delle celle diagonali e il prodotto incrociato delle celle extradiagonali. Per questo l'odds-ratio è talvolta chiamato anche **rapporto fra prodotti incrociati** (*cross-product ratio*). Nel caso della tabella 5.5, l'odds-ratio è uguale a $OR^{cd} = (167)(266)/(152)(431) = 0,68$: questo valore indica che fra le donne l'odds di fumare corrisponde a circa due terzi dell'odds di fumare osservato fra gli uomini. Se le due categorie della variabile di colonna (il genere) fossero state invertite, lasciando invariata la posizione delle righe, avremmo ottenuto un odds-ratio a prima vista diverso: $OR^{cd} = (152)(431)/(167)(266) = 1,47$. Questa differenza, tuttavia, è solo apparente. Ciò che cambia, infatti, è solo l'interpretazione della relazione fra le due variabili che, come l'ordine delle colonne, deve essere ribaltata: la propensione degli uomini a fumare è 1,47 volte quella delle donne. In generale, qualsiasi combinazione delle quattro celle di una tavola 2×2 dà sempre lo stesso risultato sostanziale, purché l'interpretazione tenga conto delle categorie di riferimento.

Se due variabili non sono correlate i loro odds condizionati sono identici e, quindi, $OR = 1$. Odds-ratio maggiori di 1 indicano l'esistenza di una covariazione positiva fra le due variabili, mentre odds-ratio minori di 1 indicano l'esistenza di una covariazione negativa o inversa. Poiché l'odds è indefinito quando il suo denominatore è uguale a zero, ogni volta che una frequenza di cella è uguale a zero l'odds-ratio

non può essere calcolato. Questo problema è identico a quello già incontrato presentando la misura Q di Yule. Per inciso, si può osservare che Q è una semplice funzione dell'odds-ratio:

$$Q = \frac{OR - 1}{OR + 1} = \frac{bc/ad - 1}{bc/ad + 1} = \frac{bc - ad}{bc + ad}$$

In altri termini, Q è semplicemente un odds-ratio che è stato «normalizzato» in modo tale da essere simmetrico e variare fra -1 e $+1$. Si noti che quando $OR = 1$ (nessuna relazione), $Q = 0$ (sempre nessuna relazione).

Gli odds-ratio possono essere calcolati anche nelle tavole di contingenza più ampie di quelle 2×2 . La tabella 5.8, ad esempio, è una tavola di contingenza 2×7 che incrocia il voto espresso dagli intervistati alle elezioni presidenziali americane del 1988 (V) e il loro orientamento politico (P). La relazione fra queste due variabili risulta statisticamente significativa ($\chi^2 = 103,8$, $gl = 6$, $p < 0,001$) seppure debole ($G = 0,47$, $\tau_c = 0,36$ e $d_{xy} = 0,34$ assumendo il voto come variabile dipendente). A partire dalle frequenze di questa tavola si possono calcolare molti odds-ratio. Ad esempio, l'odds-ratio che confronta la propensione a votare Bush anziché Dukakis manifestata dagli intervistati di estrema destra con la medesima propensione manifestata dagli intervistati di destra è uguale a $(28)(28)/(110)(5) = 1,43$. Analogamente, il confronto fra gli intervistati di destra e quelli di centro-destra in termini di propensione a votare per Bush anziché per Dukakis dà luogo a un odds-ratio uguale a $(110)(43)/(28)(112) = 1,51$. A partire dalla tabella 5.8 possono essere calcolati 21 odds-ratio diversi. Chiaramente,

TAB. 5.8. Voto espresso alle elezioni presidenziali americane del 1988, secondo l'orientamento politico: frequenze assolute e percentuali (di colonna)

Candidato votato	Orientamento politico							Totale
	Estrema sinistra	Sinistra	Centro-sinistra	Centro	Centro-destra	Destra	Estrema destra	
<i>Frequenze assolute</i>								
Bush	5	29	51	221	112	110	28	551
Dukakis	13	63	83	136	43	28	5	371
Totale	18	92	134	357	155	138	33	922
<i>Percentuali di colonna</i>								
Bush	27,8	31,5	38,1	61,9	72,3	79,7	82,1	59,8
Dukakis	72,2	68,5	61,9	38,1	27,7	20,3	17,9	40,2
Totale	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
(N)	(18)	(92)	(134)	(357)	(155)	(138)	(33)	(922)

Nota: Casi con informazioni mancanti: 595.

Fonte: General Social Survey 1991.

non tutti questi odds ratio sono indipendenti. Di fatto, poiché a questa tabella sono associati solo 6 gradi di libertà, possiamo calcolare solo 6 odds-ratio indipendenti, a partire dai quali è possibile derivare tutti gli altri.