Metodologia delle scienze umane

Fabrizio Martire

La regressione logistica e i modelli log-lineari nella ricerca sociale

FrancoAngeli

Metodologia delle scienze umane

Collana della Sezione di Metodologia dell'Associazione Italiana di Sociologia

La collana è un punto d'arrivo e allo stesso tempo un punto di partenza delle riflessioni sul metodo entro l'ampio ventaglio delle scienze umane.
Come punto d'arrivo di una tradizione complessa e ricca di solidi sedimenti, la collana intende collocarsi sul versante dell'alta divulgazione e raggiungere non solo gli studenti e i docenti universitari, ma anche il pubblico crescente delle professioni interessate alle varie forme di trattamento delle informazioni.
Come punto di partenza, essa non mancherà di presentare in modo problematico quei settori della tradizione metodologica teoricamente incerti, o fondati su presupposti discutibili, o soggetti ad abusi applicativi; né trascurerà di suggerire nuove direzioni e orientamenti.
Il piano della collana prevede ora una cinquantina di volumi, programmati su un arco di tempo di circa dieci anni e affidati a studiosi di sociologia, psicologia, statistica, storiografia, economia e altre discipline: una enciclopedia per il consolidamento e lo sviluppo delle scienze umane.

1120. *Metodologia delle scienze umane*

1. Gianni Losito, *L'analisi del contenuto nella ricerca sociale*
2. Luca Ricolfi, *Tre variabili. Un'introduzione all'analisi multivariata*
3. Alberto Marradi, *L'analisi monovariata*
4. Roberto Biorcio, *L'analisi dei gruppi*
5. Oscar Itzcovich, *L'uso del calcolatore in storiografia*
6. Giuseppe A. Micheli, Piero Manfredi, *Correlazione e regressione*
7. Francesca Zajczyk, *Fonti per le statistiche sociali*
8. Giampietro Gobo, *Le risposte e il loro contesto. Processi cognitivi e comunicativi nelle interviste standardizzate*
9. Paolo Montesperelli, *L'intervista ermeneutica*
10. Roberto Fideli, *La comparazione*
11. Antonio M. Chiesi, *L'analisi dei reticoli*
12. Cinzia Meraviglia, *Le reti neurali nella ricerca sociale*
13. Elisabetta Ruspini, *La ricerca longitudinale*
14. Juan Ignacio Piovani, *Alle origini della statistica moderna. La scuola inglese di fine Ottocento*
15. Giovanni Di Franco, *Corrispondenze multiple e altre tecniche multivariate per variabili categoriali*
16. Ivana Acocella, *Il focus group: teoria e tecnica*
17. Erika Cellini, *L'osservazione nelle scienze umane*
18. Paolo Parra Saiani, *Gli indicatori sociali*
19. Maria C. Pitrone, *Sondaggi e interviste. Lo studio dell'opinione pubblica nella ricerca sociale*
20. Giovanni Delli Zotti, *Tecniche grafiche di analisi e rappresentazione dei dati*
21. Federico Podestà, *Tecniche di analisi per la ricerca comparata trans-nazionale*
22. Fabrizio Martire, *La regressione logistica e i modelli log-lineari nella ricerca sociale*

Questo volume è stato accettato nella collana in seguito
al giudizio positivo conforme di due *referees* anonimi,
di cui uno straniero.
Per conto del Comitato Scientifico della collana
hanno seguito la redazione del volume:

Antonio De Lillo (†)
Alberto Marradi

Fabrizio Martire

La regressione logistica e i modelli log-lineari nella ricerca sociale

Metodologia delle scienze umane / 22

FrancoAngeli

Indice

Introduzione

Negli anni novanta i modelli log-lineari godevano di un credito notevole tra i ricercatori sociali. Le loro potenzialità come strumento di analisi multivariata espressamente dedicato alle variabili categoriali venivano illustrate e analizzate in numerosi saggi metodologici, e messe alla prova in diversi ambiti di interesse sociologico (comportamento elettorale, mobilità sociale, etc.). Di recente l'interesse verso i modelli log-lineari è rapidamente diminuito; nello stesso periodo, e altrettanto rapidamente, la regressione logistica si è affermata nella comunità dei ricercatori sociali come nuovo modo per analizzare le relazioni tra variabili categoriali.

Tutto ciò potrebbe far supporre che le due tecniche siano intercambiabili in riferimento alle forme di analisi che consentono, e che la regressione logistica si sia affermata perché più adatta alle esigenze specifiche dei ricercatori sociali. A mio avviso i modelli log-lineari non sono pienamente sostituibili dalla regressione logistica (per approfondimenti vedi oltre, par. 1.4). Le due tecniche infatti presuppongono due diverse rappresentazioni delle relazioni tra variabili: i modelli log-lineari sono adatti ad analizzare strutture di relazioni simmetriche tra le variabili; la regressione logistica analizza invece relazioni di dipendenza e presuppone quindi che il ricercatore distingua tra variabili dipendenti e variabili indipendenti.

A partire da questa differenza generale, nel libro illustro le specificità delle due tecniche, cercando soprattutto di mettere in luce in quali circostanze una tecnica è preferibile all'altra.

Oltre che per far emergere le loro specificità, il confronto fra i modelli log-lineari e la regressione logistica è utile anche per mostrare le potenzialità e i limiti che le accomunano. In particolare mi riferisco alle difficoltà che si pongono nell'interpretazione dei risultati dei modelli log-lineari e della regressione logistica quando le due tecniche vengono usate per analizzare

le relazioni tra variabili categoriali politomiche. Tali difficoltà non sono imputabili ad aspetti formali delle due tecniche, quanto piuttosto all'alta autonomia semantica delle variabili categoriali, che ha conseguenze rilevanti in sede di analisi dei dati indipendentemente dalla tecnica adottata[1].

Nel primo capitolo tratto alcune questioni metodologiche a mio avviso utili per inquadrare al meglio i modelli log-lineari e la regressione logistica: la natura particolare delle variabili categoriali politomiche e dicotomiche; i diversi scopi conoscitivi delle tecniche di analisi dei dati più usate nella ricerca sociale; la storia (ricostruita nelle sue fasi principali) degli strumenti di analisi ideati espressamente per le variabili categoriali; le tecniche di analisi delle relazioni fra due variabili categoriali.

Nel secondo e nel terzo capitolo introduco, rispettivamente, i modelli log-lineari e la regressione logistica. Entrambi i capitoli sono dedicati agli aspetti fondamentali delle due tecniche: i modi di rappresentazione delle relazioni tra variabili che esse presuppongono; le procedure di calcolo; l'interpretazione dei risultati che producono. Nel terzo capitolo, in relazione a ciascuno di questi aspetti, mi soffermo sulle principali differenze fra le due tecniche.

Nel secondo e nel terzo capitolo presento i modelli log-lineari e la regressione logistica mostrando alcune applicazioni all'analisi delle relazioni tra variabili dicotomiche, cioè nelle situazioni in cui le due tecniche dànno il loro meglio. Nel quarto e nel quinto capitolo mostro le principali difficoltà che le variabili politomiche pongono alle due tecniche e suggerisco alcune soluzioni (soprattutto in riferimento ai modelli log-lineari) per gestire tali difficoltà.

Nella stesura dei capitoli ho ridotto allo stretto indispensabile la trattazione degli aspetti formali e matematici. Ho invece cercato di approfondire gli assunti che le procedure di calcolo delle due tecniche implicano. Ho illustrato gli usi possibili dei modelli log-lineari e della regressione logistica attraverso una serie di esempi basati su dati effettivamente rilevati. Più che per mostrare il funzionamento delle due tecniche nelle situazioni in cui i

[1] Sul concetto di autonomia semantica vedi il par. 1.2. Questo concetto fondamentale per la raccolta e ancor più per l'analisi dei dati è stato introdotto da Alberto Marradi nel 1980, e da lui approfondito in lavori successivi, dal 1992 al 2007. Il fatto che prima il concetto fosse ignorato (e tuttora lo sia largamente) nei manuali statistici e metodologici potrebbe sorprendere chi non tenga presente che l'impostazione positivista e comportamentista che ha dominato – e continua sostanzialmente a dominare – la statistica e l'analisi dei dati condanna come soggettivo, e quindi non-scientifico, ogni accenno al significato, al fatto che le variabili e le loro categorie vanno interpretate da menti umane perché i calcolatori non possono farlo, e così via.

risultati che esse producono sono chiari, ho scelto esempi che facciano emergere anche le debolezze e gli aspetti problematici.

Ringrazio Maria Concetta Pitrone per i consigli che mi ha dato durante la progettazione del volume. Ringrazio inoltre il compianto Antonio De Lillo e Alberto Marradi per le preziose indicazioni che mi hanno dato revisionando i capitoli. Ringrazio in particolare Alberto Marradi per avermi aiutato a depurare il testo da inutili tecnicismi.

1. Le relazioni tra variabili categoriali

1.1. Proprietà e variabili

Le tecniche di analisi che illustrerò in questo testo presuppongono una matrice dei dati come forma specifica di organizzazione del materiale empirico, che a sua volta presuppone dal parte del ricercatore la rappresentazione di un fenomeno da indagare in termini di proprietà, oggetti e stati su proprietà[1].

Per riempire una matrice dei dati un ricercatore deve definire delle regole e delle procedure che gli consentano di individuare i referenti del tipo di oggetto sul quale ha deciso di condurre l'indagine, e di rilevare su essi le proprietà che gli interessano. Questa operazione di ricerca è un passaggio cruciale di qualsiasi indagine che si basa su una matrice dei dati, anche se, come lamenta Ricolfi (1995, 407), è "troppo sovente trascurato o assunto come automatico". Adottando la terminologia usata da Marradi (2007, capitolo 6) chiamerò tale operazione di ricerca 'definizione operativa'; 'variabili', 'categoria' e 'casi' i suoi esiti. Nelle matrici dei dati non abbiamo quindi le proprietà, ma le variabili; abbiamo i casi e non gli oggetti; e nella cella-incrocio tra un determinato caso e una determinata variabile abbiamo un

[1] Con il termine 'proprietà' si intende una caratteristica che il ricercatore decide di rilevare sistematicamente su tutti gli oggetti della sua ricerca. Con il termine 'oggetto' si intende l'unità sulla quale il ricercatore decide di rilevare (o alla quale decide di riferire) le proprietà. Una volta stabiliti gli oggetti e le proprietà, la logica matriciale impone che tutte le proprietà siano rilevate su (o riferite a) tutti gli oggetti. Con l'espressione 'stato su una proprietà' si intende lo stato che un dato oggetto assume in riferimento a una data proprietà; ad esempio 'in cerca di occupazione' è un possibile stato sulla proprietà 'condizione lavorativa'; 'italiana' è un possibile stato sulla proprietà 'cittadinanza'; etc. Per approfondimenti sui concetti di matrice, proprietà, oggetto e stato su una proprietà vedi Marradi (2007, capitoli 5 e 6).

dato, cioè il modo in cui la definizione operativa registra lo stato di quel caso (in riga) su quella proprietà (rappresentata dalla variabile in colonna).

Questa distinzione tra il piano concettuale (proprietà, oggetti e stati sulle proprietà) e quello della matrice (variabili, casi e dati) è particolarmente rilevante perché tra le coppie di elementi di ciascun piano non esiste una corrispondenza univoca. In altre parole, una volta che ho stabilito le proprietà da rilevare o gli oggetti su cui condurre l'indagine, la definizione operativa che posso adottare per trasformali, rispettivamente, in variabili e in casi non è determinata soltanto dalla loro natura, ma dipende anche dagli obiettivi, dalle esigenze e dai vincoli della ricerca.

La figura 1.1 rappresenta una possibile corrispondenza tra una tassonomia di proprietà e una di variabili[2].

proprietà		*variabili*
discrete categoriali	➔	categoriali
discrete ordinali	➔	ordinali
discrete cardinali	➔	cardinali naturali
continue percepibili con i sensi	➔	cardinali metriche
continue non percepibili con i sensi	➔	ordinali quasi cardinali

Fig. 1.1 - Quadro delle corrispondenze fra tipi di proprietà e tipi di variabili (adattamento da Marradi 2007, tab. 7.8).

Le corrispondenze rappresentate nella figura 1.1 non vanno tutte intese come un vincolo. Solo le proprietà categoriali *devono* essere trasformate nel tipo di variabile ad esse corrispondente; per tutte le altre proprietà esiste un margine – più o meno ampio – di scelta. In questo senso possiamo immaginare una gerarchia tra le proprietà. Quelle che possono dar luogo a variabili cardinali sono trasformabili anche in ordinali o categoriali[3]. La ragione di tutto ciò è semplice: se tra gli stati di una proprietà posso stabilire relazioni quantitative posso anche limitarmi a considerare il loro ordine (e trasformarle in variabili ordinali), o semplicemente il fatto che sono uno diverso dall'altro (trasformandole in variabili categoriali). Ovviamente, il margine di scelta per le proprietà ordinali è meno ampio: posso trasformale in variabili ordinali o categoriali.

[2] Per approfondimenti sulle due tassonomie vedi Marradi (ivi, capitolo 7).

[3] Lo stesso discorso vale per le proprietà continue i cui stati non sono percepibili con i sensi, cioè quelle che possono dar luogo a variabili quasi-cardinali.

La relativa indipendenza tra proprietà e variabili può essere una caratteristica da sfruttare sia nella fase della raccolta dei dati sia in quella dell'analisi. Si pensi al reddito, una proprietà discreta cardinale e quindi trasformabile in una variabile cardinale naturale. Nei sondaggi il declassamento di questa proprietà a varabile ordinale è una pratica consolidata. Il reddito infatti è considerato una proprietà che molti soggetti ritengono riservata: chiedere informazioni su di esso potrebbe causare risposte infedeli o, peggio, l'interruzione dell'intervista. Si preferisce quindi proporre all'intervistato diverse fasce di reddito (costruendo così una variabile ordinale) e chiedergli a quale fascia appartiene.

Il declassamento delle variabili è un'operazione comune anche nell'analisi dei dati. Si pensi all'età anagrafica; in un sondaggio, per non rinunciare a livelli di dettaglio potenzialmente rilevanti, è opportuno chiedere l'età in anni compiuti. Ma in fase di presentazione monovariata dei dati una distribuzione di frequenza con una cinquantina (o più) di categorie, ciascuna con bassa frequenza, è una maniera di disperdere l'informazione.

Nei capitoli 4 e 5 mostrerò come in alcuni casi i modelli log-lineari e la regressione logistica riescano a far emergere relazioni particolari tra due o più variabili ordinali proprio quando il ricercatore decide di trattarle come categoriali.

La lista di variabili riportata nella figura 1.1 è costruita in funzione dei diversi modi per definirle operativamente. Se invece della raccolta si assume come riferimento l'analisi dei dati, può essere opportuno proporne una diversa; nella figura 1.2 (colonna a destra) propongo una classificazione di tipi di variabili rilevante per un'adeguata illustrazione delle tecniche di analisi illustrate in questo libro.

Alcuni tecniche di analisi possono trattare solo variabili cardinali. Rispettato questo vincolo, l'opportunità di prendere in considerazione una data variabile dipende, oltre che da riflessioni teoriche, dalla sua distribuzione monovariata e/o dalle relazioni che ha con le altre variabili considerate. Si tratta di valutazioni che prescindono dalla specifica definizione operativa attraverso la quale la variabile in esame è stata costruita. Di conseguenza, in sede di analisi dei dati la distinzione tra cardinali naturali, metriche e quasi-cardinali perde rilevanza.

tipi di variabili costruiti in funzione della definizione operativa		*tipi di variabili costruiti in funzione dell'analisi dei dati*
categoriali	→	categoriali dicotomiche
	→	categoriali politomiche
ordinali	→	ordinali
cardinali naturali cardinali metriche quasi cardinali	→	cardinali

Fig. 1.2 - Due classificazioni di variabili.

Oltre ad accorpare i diversi tipi di variabili cardinali in una categoria unica, passando dalla prospettiva della raccolta dei dati a quella dell'analisi è utile, a mio avviso, specificare due sotto-tipi delle variabili categoriali: le dicotomiche e le politomiche. L'opportunità di questa distinzione è strettamente legata al concetto di autonomia semantica delle categorie. Si tratta di un concetto cruciale che chiamerò sistematicamente in causa nei capitoli che seguono; per questo gli dedicherò il prossimo paragrafo.

In relazione alla classificazione di tipi di variabili proposta nella colonna di destra della figura 1.2 è possibile individuare una prima grande differenza tra i modelli log-lineari e la regressione logistica: i primi sono tipicamente usati per analizzare variabili categoriali (sia dicotomiche, sia politomiche) e ordinali; la regressione logistica può gestire anche le variabili cardinali[4].

1.2. I concetti di autonomia semantica e grado di libertà

L'autonomia semantica è una proprietà di una categoria e consiste nella maggiore o minore possibilità di interpretarla senza far ricorso al nome della variabile o delle altre categorie della variabile. Le categorie delle variabili categoriali hanno massima autonomia semantica; quelle delle categoriali ordinate hanno un'autonomia minore mentre quelle delle variabili cardinali hanno in genere autonomia semantica minima o nulla. Il concetto è stato proposto da Marradi (1980, 57-65) per mettere in luce i problemi che diversi tipi di variabili pongono nella fase della raccolta dei dati[5]. Successiva-

[4] Per approfondimenti vedi il capitolo 5.

[5] Ad esempio l'autore (ivi, 65) sottolinea: "le scale Likert […] sono estremamente vulnerabili alle distorsioni perché le loro categorie (d'accordo, sfavorevole, etc.) mancano di autonomia semantica, cioè dipendono integralmente, per la loro interpretazione, dal testo

mente, l'autore ha analizzato le conseguenze dell'autonomia semantica in sede di analisi dei dati (Marradi 1993; 1997; 2007). In questo paragrafo riprendo questa seconda prospettiva.

Si prendano le due variabili riportate nella tabella 1.1: le categorie della variabile di sinistra (categoriale) hanno maggiore autonomia semantica di quelle della variabile di destra (cardinale).

Tab. 1.1 - Due ipotetiche distribuzioni di frequenza.

genere cinematografico preferito	*%*		*gradimento del film "Il discorso del Re" su una scala da 1 a 10*	*%*
animazione	8		1	2
avventura	20		2	14
commedia	14		3	6
drammatico	4		4	8
horror	2		5	4
musicale	6		6	7
sentimentale	9		7	12
storico	7		8	18
thriller	18		9	15
western	12		10	14
totale	1.500		totale	1.500

Il grado di autonomia semantica ha importanti conseguenze sul piano dell'analisi dei dati. Dalla distribuzione di frequenza della variabile 'genere cinematografico preferito' risulta che un intervistato su cinque ha indicato 'avventura'. Ovviamente la distribuzione si presta ad interpretazioni più sofisticate; tuttavia questo dato elementare assume pieno significato indipendentemente dalle percentuali degli altri generi cinematografici.

Al contrario nella colonna di destra per dare un pieno significato al fatto che, ad esempio, il 14% degli intervistati ha scelto la modalità 2 devo far riferimento non solo al nome della variabile, ma anche alle percentuali delle altre modalità[6]. In questo senso, l'incidenza relativamente alta della modalità 2 si configura come un'eccezione in una distribuzione di frequenza

della domanda ('d'accordo' con che cosa? 'sfavorevole' a che cosa?). È questo fatto a rendere possibile la 'curvilinearità': se la categoria è semanticamente autonoma, cioè è un'affermazione di senso compiuto, è assai più difficile che venga disapprovata da due persone per motivi opposti".

[6] Allo scopo di mantenere distinte le fasi di ricerca, userò il termine 'modalità' in riferimento alle operazioni di analisi dei dati e 'categoria' in riferimento ai procedimenti della definizione operativa e della raccolta dei dati.

sbilanciata verso i punteggi alti. Qualora le percentuali dei punteggi bassi fossero state più alte, avrei dato un'interpretazione diversa alla percentuale della modalità 2.

Il grado di autonomia semantica influenza non solo l'interpretazione delle singole modalità, ma anche l'analisi della variabile nel suo complesso. Come abbiamo appena visto, la distribuzione di frequenza nella colonna di destra si presta a questa interpretazione sommaria, ma difficilmente contestabile: tendenzialmente tra gli intervistati si registra una valutazione positiva del film *Il discorso del Re*. Se avessimo deciso di registrare il livello di gradimento con una scala da 1 a 100 e se anche in questo caso le risposte si fossero concentrate sui livelli alti della scala, avremmo potuto dare la stessa interpretazione.

Molto diverso è il caso delle variabili categoriali, in cui ogni modalità è un centro autonomo di interesse semantico. L'interpretazione della distribuzione riportata nella colonna sinistra non può che essere più articolata di quella della colonna di destra. Sintetizzandola discorsivamente si può infatti sostenere: i generi avventura e *thriller* sono i più scelti, anche il genere commedia fa registrare un discreto consenso, mentre i film *horror* decisamente non piacciono. L'interpretazione può essere più o meno sintetica, ma comunque non si può fare a meno di fare riferimento alle singole modalità. Di conseguenza, qualora avessimo deciso di scorporare, ad esempio, il genere *thriller* nei sotto-generi *noir*, giallo e poliziesco, l'interpretazione complessiva della distribuzione di frequenza sarebbe stata ulteriormente complicata dall'aumento dei centri autonomi di interesse semantico.

Queste differenze tra variabili categoriali e cardinali emergono anche nell'analisi bivariata. Se, ad esempio, su un campione di studenti universitari ho rilevato le variabili 'tempo dedicato allo preparazione dell'esame X' e 'voto conseguito all'esame X' posso analizzare la relazione tra le due variabili e trovarmi realisticamente in una di queste situazioni: a) tra le due variabili non c'è relazione; b) maggiore è il tempo dedicato più alto è il voto ottenuto; c) fino a una certa soglia di tempo dedicato le due variabili sono correlate positivamente; oltre una certa soglia si registra una relazione inversa. L'interpretazione della relazione può arricchirsi ulteriormente – ad esempio considerando la sua intensità – ma comunque quando analizzo variabili cardinali sono in grado di dare significato a una relazione facendo riferimento al nome delle variabili.

Se dalla stessa ipotetica matrice prendo due variabili categoriali come 'scuola superiore frequentata' e 'corso universitario di appartenenza' mi trovo in una situazione diversa. Dall'analisi della loro relazione potrei certamente evincere che tra le due *variabili* c'è indipendenza statistica; ma in

caso contrario non posso prescindere dal chiamare in causa le singole *modalità*. Infatti una frase come "il tipo di scuola superiore frequentata influenza il corso di laurea scelto" non ci dà molte informazioni; dire poi che tra le due variabili c'è una relazione diretta o inversa è semplicemente un *nonsense*. La natura categoriale delle variabili richiede una descrizione della loro eventuale relazione che espliciti i rapporti tra le singole modalità; ad esempio: "tendenzialmente chi proviene dalla scuola X si iscrive al corso universitario Y"; "chi ha frequentato il liceo Z tende a non iscriversi al corso V"; etc.

Ne consegue che, quando tratto variabili le cui categorie hanno alta autonomia semantica, un elevato numero di modalità può rendere lunga e complessa l'analisi e l'interpretazione dei dati. Questo fa emergere un'importante tensione tra le esigenze della raccolta dei dati e quelle dell'analisi. Per non perdere informazioni potenzialmente rilevanti, nella fase di rilevazione è consigliabile costruire variabili molto sensibili[7]; tuttavia nella fase di analisi la sensibilità delle variabili categoriali può diventare un problema per le ragioni che abbiamo appena visto. Di conseguenza per analizzare le variabili categoriali è spesso opportuno aggregare le loro categorie in un numero ridotto di modalità[8].

Alla fine del paragrafo precedente ho introdotto una distinzione tra variabili categoriali politomiche e dicotomiche chiamando in causa il concetto di autonomia semantica. In un certo senso le variabili dicotomiche possono essere assimilate alle variabili ordinali, che sono caratterizzate da un livello di autonomia semantica ridotto rispetto alle categoriali[9]. Si prenda ad esempio la variabile genere; posso ordinare le due modalità maschio e femmina se ri-concettualizzo la variabile come appartenenza al genere femminile (o maschile, in quel caso l'ordine sarebbe ovviamente inverso); in tal caso infatti posso dire che un maschio è *meno* appartenente al genere femminile di una femmina. Una simile ri-concettualizzazione è applicabile a qualsiasi variabile dicotomica (Marradi 1997, 54).

[7] Con il termine 'sensibilità' intendo il rapporto tra le categorie di una variabile e gli stati potenziali della proprietà da cui deriva (Marradi 2007, 107).

[8] Oltre che per ragioni semantiche, la riduzione del numero di modalità è necessaria anche per ragioni statistiche. Maggiore è il numero delle modalità, maggiori sono i rischi di avere distribuzioni di frequenza squilibrate o modalità con frequenza scarsa o nulla, le quali a loro volta ostacolano il corretto funzionamento delle principali tecniche di analisi dei dati (Marradi 1993; 1997; Di Franco 2006). Tornerò sul punto nel par. 1.5 e nei capitoli successivi.

[9] Infatti, le principali differenze tra le possibilità di interpretazione delle due distribuzioni illustrate nella tabella 1.1 sono largamente imputabili al fatto che tra le modalità della variabile 'gradimento del film *Il discorso del Re*' posso stabilire un ordine.