

## **Uno studio sulla percezione delle emergenze metropolitane: un approccio modellistico**

Angela D'Elia, Domenico Piccolo

*Dipartimento di Scienze Statistiche, Università di Napoli Federico II*

*E-mail: angela.delia@unina.it; domenico.piccolo@unina.it*

*Summary:* In this work we propose a statistical approach for studying the perception of the more serious problems that arise in a large metropolitan area. We classify some of them as emergencies and ask to a sample of young people living in Naples to rank them with respect to their relevance. Then, we apply a mixture model for explaining the expressed evaluation; therein, the uncertainty of the answer and the seriousness of the problem are jointly examined and compared. Some final considerations about further developments conclude the paper.

*Keywords:* Ordinal data, Main emergencies, Heterogeneity measures, MUB model.

### ***1. Introduzione***

Le problematiche che coinvolgono le grandi città sono spesso comuni e vengono periodicamente poste alla ribalta dell'attenzione pubblica in funzione dell'aggravarsi di una di esse. Il complesso di tali questioni – che riguardano la vita quotidiana dei cittadini e dovrebbero costituire priorità assolute per i governanti – viene qui definito, per semplicità, *emergenze metropolitane*.

È ben noto che tali emergenze variano nel tempo e nello spazio, sia in funzione di fattori esterni (evoluzione climatica, sconvolgimenti naturali, etc.) che di fatti squisitamente umani (di natura politica, economica, sociale, etc.). La loro periodica visibilità è accentuata dai mass media che, in tempi recenti, attualizzano l'emergenza in atto diffondendone la

percezione in quasi tutti gli strati della popolazione; questo fenomeno ne amplifica la rilevanza ma, spesso, la circoscrive nel tempo. Ciò può giungere sino a rendere poco visibile un problema grave e permanente che, invece, si ripresenta all'attenzione pubblica solo quando punte eccezionali (magari localizzate) rendono nuovamente prioritaria ed attuale quella emergenza.

In tale contesto, è sembrato utile misurare come una particolare categoria sociale (studenti universitari) percepisca le emergenze metropolitane, anche in relazione ad alcune variabili connesse alle loro caratteristiche personali (sesso, età, residenza, etc.).

Per questo, dopo aver descritto il piano sperimentale dell'indagine e discusso le risultanze delle principali analisi esplorative, il lavoro si sofferma sulla variabilità delle risposte fornite. Più precisamente, vengono esaminate alcune misure di dispersione, intese quali indicatori di incertezza dei giudizi espressi dai rispondenti. Quindi, viene presentato – nelle sue linee essenziali – un modello statistico che si è dimostrato particolarmente efficace per la descrizione e l'interpretazione di risposte che si esprimono mediante variabili ordinali. Successivamente, il modello viene generalizzato mediante la introduzione di covariate del soggetto per spiegare eventuali atteggiamenti differenziati nella percezione della gravità di una emergenza. Infine, alcune considerazioni finali concludono il lavoro.

## ***2. Piano sperimentale dell'indagine***

Riteniamo che, preliminarmente all'intero studio, vada chiarita ed argomentata una scelta di tipo generale che presiede all'indagine e che concerne il modo stesso con il quale viene posto all'intervistato il quesito cui deve rispondere.

In effetti, in questa tipologia di ricerche empiriche, è possibile elencare le singole problematiche di interesse e porre su ciascuna di esse un quesito specifico circa il suo grado di rilevanza nella percezione dell'intervistato. Per contro, è anche possibile chiedere all'intervistato di ordinare, in modo esaustivo, una serie di problematiche, dalla più grave alla meno grave, escludendo la possibilità di collocazioni paritarie.

Siamo convinti che la seconda opzione conferisca maggiore rigore alle risposte che si ricevono, per almeno due ordini di motivi:

- l'ordinamento richiede una analisi comparata, il che trasforma la percezione vaga e diffusa che ciascuno possiede (la *sensazione* ...) in una misurazione ordinale che permette di individuare una graduatoria di rilevanza. Ciò riduce (o addirittura elimina) il rischio di banali uniformità di valutazioni, perché la formulazione di una graduatoria impone un confronto meditato fra le problematiche;
- l'ordinamento evita il frequente ricorso a collocazioni intermedie che derivano da fattori esterni alle problematiche sottoposte (imprecisione, stanchezza, ignoranza, superficialità, etc.).

Ovviamente, utilizzando un ordinamento si induce un vincolo tra le risposte di ciascun intervistato, ma questo non appare rilevante se si è interessati alla costruzione di modelli univariati per ciascuna emergenza, perché essi si stimano per variabili marginali rispetto alla distribuzione multivariata.

Per ciò che concerne l'indagine sperimentale, nel mese di dicembre 2004 è stato chiesto ad un campione di studenti universitari della Facoltà di Scienze Politiche della Università degli Studi di Napoli Federico II (frequentanti un insegnamento del I anno) di rispondere ad un questionario nel quale veniva chiesto – tra l'altro – di graduare  $m = 9$  emergenze metropolitane, elencate in ordine alfabetico.

Nell'esplicitare la graduatoria, veniva suggerito agli intervistati di rispondere *in modo sequenziale*, partendo dalla problematica ritenuta più grave e poi via via sino a quella ritenuta meno grave, evitando di attribuire lo stesso numero a due differenti emergenze.

Il questionario conteneva altre domande a risposta chiusa, concernenti la caratterizzazione del soggetto e della sua vita universitaria. Va aggiunto che gli intervistati avevano appena concluso un corso di Statistica di base, per cui erano stati motivati ed educati alla consapevolezza di fornire risposte complete e tra loro coerenti. A posteriori, la codifica ed il controllo dei dati hanno confermato che gli ordinamenti forniti erano non contraddittori, al punto che solo in pochissimi casi si è dovuto escludere il singolo questionario dall'indagine.

Specificamente, la domanda relativa alle emergenze metropolitane era così articolata:

- Esprimi una tua personale graduatoria di importanza (mediante un numero da 1 a 9), mettendo in ordine le seguenti *emergenze metropolitane di Napoli* in modo da attribuire il numero 1 a quella che ritieni la più grave, il numero 2 a quella successiva in ordine di gravità, e così via, sino ad attribuire il numero 9 al problema ritenuto da te meno grave rispetto a quelli elencati.
- Per favore, non dare MAI lo stesso numero a due problemi diversi.

<input type="checkbox"/>	<i>Clientelismo e corruzione</i>
<input type="checkbox"/>	<i>Criminalità organizzata</i>
<input type="checkbox"/>	<i>Disoccupazione e mancanza di lavoro</i>
<input type="checkbox"/>	<i>Inquinamento ambientale e carenza di verde</i>
<input type="checkbox"/>	<i>Mala-sanità</i>
<input type="checkbox"/>	<i>Micro-criminalità</i>
<input type="checkbox"/>	<i>Presenza di immigrati</i>
<input type="checkbox"/>	<i>Pulizia ed igiene del territorio</i>
<input type="checkbox"/>	<i>Trasporti-traffico-mobilità</i>

La ricerca si è avvalsa di  $n = 354$  questionari completi; solo per

talune analisi di modelli con covariate si è operato con un sottoinsieme di numerosità inferiore (di sole poche unità).

### 3. Analisi esplorative dei dati

La successiva Tabella 1 evidenzia la distribuzione di frequenza delle risposte fornite a ciascuna delle nove emergenze prima elencate, e costituisce il punto di partenza di tutte le analisi successive. In essa abbiamo evidenziato, in grassetto, la frequenza massima che si è registrata per ciascuna variabile.

Tabella 1. Distribuzioni di frequenza delle emergenze metropolitane.

<i>Emergenze</i>	1	2	3	4	5	6	7	8	9
<i>Clientelismo e corruzione</i>	21	43	<b>88</b>	78	55	26	19	14	10
<i>Criminalità organizzata</i>	<b>212</b>	94	26	12	5	4	0	0	1
<i>Disoccupazione</i>	<b>88</b>	84	84	44	26	15	4	6	3
<i>Inquinamento e carenza verde</i>	4	5	9	21	49	64	<b>95</b>	70	37
<i>Mala-sanità</i>	2	12	38	78	<b>101</b>	57	31	24	11
<i>Micro-criminalità</i>	20	<b>97</b>	85	72	39	20	12	8	1
<i>Presenza di immigrati</i>	3	4	3	11	11	29	29	70	<b>194</b>
<i>Pulizia e igiene territorio</i>	2	7	7	27	40	92	<b>96</b>	62	21
<i>Trasporti-traffico-mobilità</i>	2	8	14	11	28	47	68	<b>100</b>	76

La corrispondente Figura 1 mostra come l'andamento delle differenti distribuzioni sia abbastanza eterogeneo modificandosi la forma, l'asimmetria, la posizione della moda, il peso relativo delle code, etc.

Poiché l'ordinamento implica solo un confronto relativo di posizione e non fornisce di per sé una valutazione numerica, non appare corretto calcolare i più comuni indici descrittivi definiti per variabili quantitative (media, scarto quadratico medio, etc.). Come è noto, per variabili ordinali, solo la mediana e la moda assumono un significato pieno. E, in effetti, la successiva modellistica (di cui al paragrafo 5) è in linea con tale esigenza poiché si limita a proporre una struttura probabilistica coerente per variabili ordinali.

Tuttavia, reputiamo utile addivenire anche a sintesi numeriche della percezione di gravità delle emergenze, per i motivi seguenti:

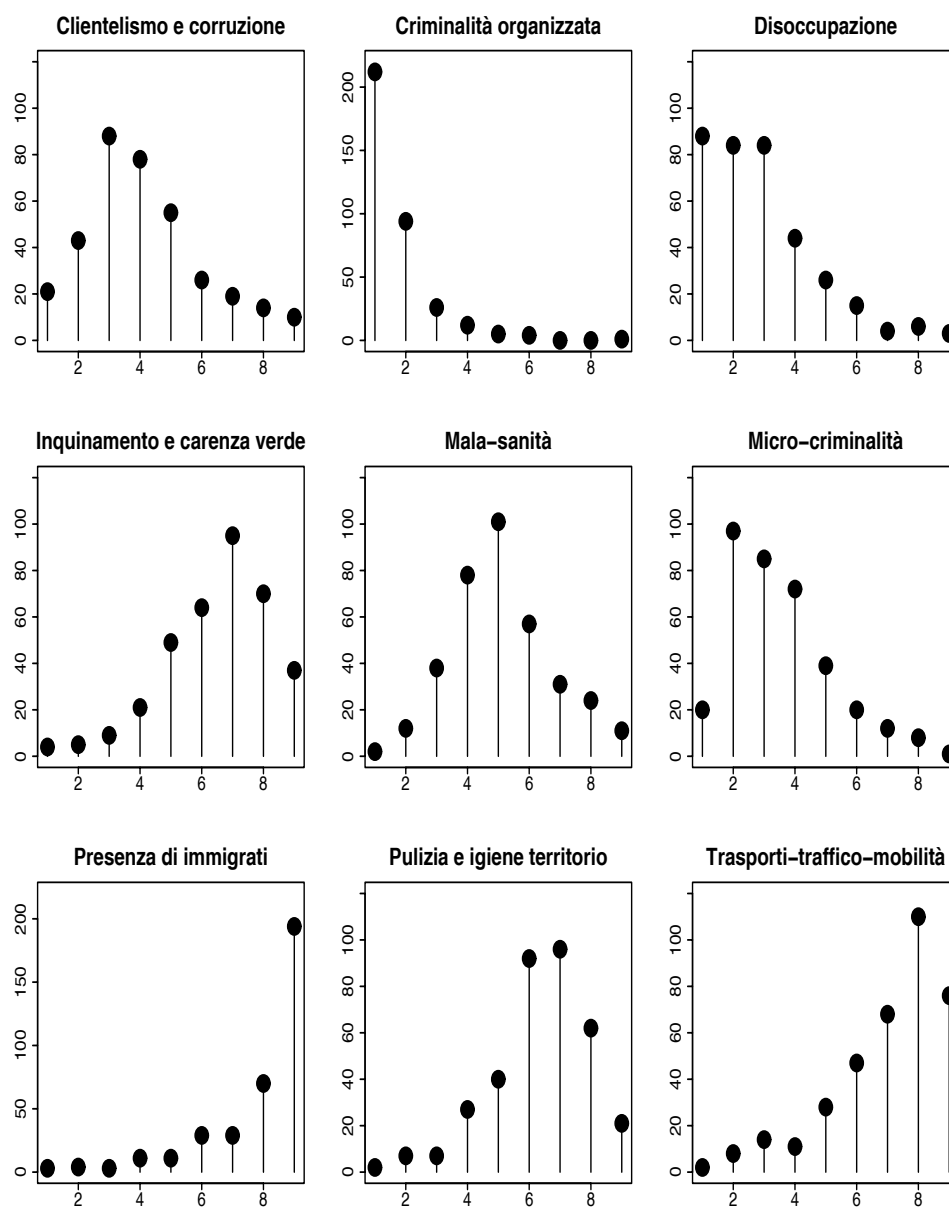


Figura 1. Distribuzioni di frequenza degli ordinamenti espressi.

- la percezione di una emergenza è sostanzialmente una variabile continua che viene opportunamente discretizzata mediante l'ordinamento che si sottopone all'intervistato. Quindi, il valore medio del rango espresso può essere inteso come una *proxy* del valore medio della variabile continua "percezione della gravità del problema";
- né la moda né la mediana sono capaci di discriminare tra maggiore o minore gravità percepita quando le distribuzioni sono abbastanza simili. Infatti, può avvenire che due variabili presentino la stessa mediana (ed anche la stessa moda, talora) pur essendo posizionate in modo differente, come risulterebbe dall'esame delle rispettive distribuzioni di frequenza.

Queste considerazioni sono evidenti se esaminiamo i principali indicatori caratteristici per le emergenze nella Tabella 2.

Tabella 2. Indicatori di posizione delle variabili ordinali.

<b>Emergenze</b>	<b>Moda</b>	<b>Mediana</b>	<b>Media</b>
<i>Clientelismo e corruzione</i>	3	4	4.093
<i>Criminalità organizzata</i>	1	1	1.650
<i>Disoccupazione</i>	1	3	2.845
<i>Inquinamento e carenza verde</i>	7	7	6.531
<i>Mala-sanità</i>	5	5	5.105
<i>Micro-criminalità</i>	2	3	3.472
<i>Presenza di immigrati</i>	9	9	7.915
<i>Pulizia e igiene territorio</i>	7	7	6.367
<i>Trasporti-traffico-mobilità</i>	8	7	7.023

Specificamente, la moda delle variabili "Pulizia e igiene territorio" e "Inquinamento e carenza verde" risulta essere pari a 7, per cui queste due problematiche (correttamente giudicate simili dai rispondenti perché connesse a questioni ambientali) richiedono di essere graduate anche sulla base dei valori medi. Emerge così che la prima è ritenuta leggermente prevalente rispetto alla seconda.

Similmente, la mediana delle variabili “Pulizia e igiene territorio”, “Inquinamento e carenza verde” e “Trasporti-traffico-mobilità” è sempre pari a 7, ma è ben evidente – dall’esame delle rispettive distribuzioni di frequenza (Tabella 1 e Figura 1) – che l’ultima variabile viene reputata comparativamente meno grave.

Sul piano metodologico, va precisato che la media aritmetica di una variabile ordinale – le cui modalità si fanno corrispondere alla successione naturale – è sempre esprimibile in funzione delle sole frequenze relative, e non dei singoli valori presupposti per le modalità.

Infatti, indicando con  $f_i$ , e  $F_i$ ,  $i = 1, 2, \dots, m$ , rispettivamente, le frequenze relative e la funzione di ripartizione empirica della variabile di interesse, e con  $\mu$  la corrispondente media aritmetica, essendo:  $f_i = F_i - F_{i-1}$ ,  $i = 1, 2, \dots, m$ ;  $F_0 = 0$ ;  $F_m = 1$ , si ottiene:

$$\mu = \sum_{i=1}^m i f_i = \sum_{i=1}^m i (F_i - F_{i-1}) = m - \sum_{i=1}^{m-1} F_i.$$

Quindi, la media aritmetica *non* dipende dall’arbitrario riferimento ai primi numeri interi.

Inoltre, si osservi che, per la natura propria dell’ordinamento utilizzato, la media aritmetica non è alterata da valori eccezionali e, quindi, in tali casi è un indicatore robusto di posizione.

Esaminando le medie della percezione di gravità nella Tabella 2, si rileva anzitutto che la variabile “Criminalità organizzata” precede in modo sensibile la “Disoccupazione”; il che può sorprendere essendo il questionario compilato per la maggior parte da giovani potenzialmente in cerca di occupazione e nella regione italiana con il più elevato tasso di disoccupazione giovanile. In effetti, questo risultato deriva dalla circostanza per cui, nelle settimane precedenti l’intervista, si è assistito ad una violenta recrudescenza nella lotta tra le famiglie camorristiche di Napoli e periferie, che ha fatto registrare parecchi morti, incluso qualche innocente. Ciò ha avuto un effetto immediato sulle risposte, confermando come l’insistenza dei mass-media sulla problematica della criminalità organizzata nel napoletano abbia prodotto una percezione di gravità (diffusa ma temporanea) che è decisamente superiore rispetto al problema della mancanza di lavoro (che è individuale e permanente).



#### 4. Dispersione dei risultati dell'indagine

Per valutare l'accuratezza delle risposte fornite, e quindi l'affidabilità che si intende attribuire all'indagine, occorre individuare misure di variabilità che siano adeguate per fenomeni ordinali<sup>1</sup>.

La problematica è stata affrontata con particolare impegno dalla scuola statistica italiana, a partire da Gini (1912), ed ha prodotto numerose proposte, spesso riscoperte negli anni più recenti e diffuse con denominazioni discordi<sup>2</sup>.

In effetti, sin dai suoi primi studi sulla variabilità e sulla concentrazione, Gini ha coerentemente affrontato la misura della variabilità per dati qualitativi come una media di tutte le *distanze*  $d(a_r, a_s)$  fra le modalità  $a_r$ ,  $r = 1, 2, \dots, m$ , a due a due considerate, mediante l'espressione:

$$\Delta = \sum_{r=1}^m \sum_{s=1}^m d(a_r, a_s) f_r f_s,$$

ove  $f_r$ ,  $r = 1, 2, \dots, m$  sono, al solito, le corrispondenti frequenze relative.

---

<sup>1</sup> Pur essendo quasi sempre utilizzati solo per variabili qualitative, tali indici sono in effetti definiti per ogni variabile a valori reali. Va anche rilevato che nella letteratura statistica essi sono indicati con denominazioni differenti: dispersione, mutabilità, eterogeneità, etc. Nel seguito del lavoro, escludiamo l'entropia che peraltro risulta connessa ai vari indici che studieremo.

Va aggiunto che, nello specifico dei nostri dati, lo scarto quadratico medio avrebbe consentito di operare una buona discriminazione circa la qualità delle risposte fornite. Tuttavia, in linea di principio, esso andrebbe sconsigliato perché influenzato dal campo di variazione della variabile, quindi da  $m$ .

<sup>2</sup> Per esempio, l'indice di Gini (attribuito a Simpson) viene chiamato *Indice di variazione qualitativa, di frammentazione (politica), di Rae e Taylor, di diversità linguistica* di Greenberg (1956), etc. Il logaritmo dell'indice di Laakso e Taagepera (proposto nel 1979) è noto dal 1949 come indice di Simpson; quello di Ray e Singer è il complemento ad 1 dell'indice di Frosini (1981). Infine, la quantità  $S(f)$ , che definiremo più avanti, è chiamata *indice di concentrazione* di Herfindahl (e Hirschman). Per alcuni di questi riferimenti, si vedano: Agresti and Agresti (1978), Allison (1981), Coulter (1984), Haberman (1982), Hill (1973), Lieberman (1969), Light and Margolin (1971), Kempton (1979), Kokko *et al.* (1999), Patil and Taillie (1982), Peet (1974).

Ora, se si pone (come generalmente avviene per variabili nominali):

$$d(a_r, a_s) = \begin{cases} 1, & \text{se } a_r \neq a_s; \\ 0, & \text{se } a_r = a_s; \end{cases}$$

la precedente espressione produce l'*indice di eterogeneità di Gini*.

Nella sua forma normalizzata, esso è definito da:

$$\mathcal{G} = \frac{m}{m-1} \left( 1 - \sum_{i=1}^m f_i^2 \right).$$

Se, invece, si pone (come generalmente avviene per variabili ordinali):

$$d(a_r, a_s) = \begin{cases} |r - s|, & \text{se } a_r \neq a_s; \\ 0, & \text{se } a_r = a_s; \end{cases}$$

si perviene ad un indicatore di dispersione per variabili ordinali, studiato da Leti (1983), 290-297.

Nella sua forma normalizzata<sup>3</sup> in  $[0, 1]$ , tale misura è definita da:

$$\mathcal{L} = \frac{4}{m-1} \sum_{i=1}^{m-1} F_i (1 - F_i) = \frac{4}{m-1} \left( m - \mu - \sum_{i=1}^{m-1} F_i^2 \right).$$

L'indice  $\mathcal{L}$  è fortemente correlato allo scarto quadratico medio<sup>4</sup> ma presenta il vantaggio di una agevole decomponibilità della dispersione all'interno dei gruppi e fra i differenti gruppi, quando si è in presenza di dati che provengono da strati diversi (come spesso avviene nelle indagini di valutazione e/o nella espressione di graduatorie): Rampichini *et al.* (2004).

---

<sup>3</sup> In effetti, la normalizzazione indicata nella formula è solo approssimata quando  $n$  è pari. Tuttavia, stante l'elevata numerosità dei dati, si può ritenere che essa sia valida in ogni caso: Leti (1983), 290-297; Gusco (1997), 58-66. In particolare, per uno studio ed una discussione di tale misura, si vedano: Picarreta (2001); Grilli e Rampichini (2002).

<sup>4</sup> Il lavoro di Scagni (2005) mostra un andamento molto simile tra i due indicatori sulla base di uno studio di simulazione per numerose distribuzioni multinomiali.

In tale ambito, merita di essere segnalato anche un indice proposto da Frosini (1981) che confronta, mediante una metrica Euclidea, le frequenze relative della distribuzione con quelle della massima eterogeneità. Dopo qualche sviluppo, la sua forma normalizzata è definita da:

$$\mathcal{F} = 1 - \sqrt{\frac{1}{m-1} \left( m \sum_{i=1}^m f_i^2 - 1 \right)}.$$

Più in generale, è noto che l'incertezza di un sistema è connessa alla sua variabilità nel senso che un sistema a risposte univoche non produce incertezza nei risultati; in effetti, l'incertezza di una distribuzione cresce sempre più mano a mano che questa si avvicina alla distribuzione uniforme. Limitandoci a variabili ordinali (espresse con modalità numeriche, come i ranghi, per esempio), è chiaro che la concentrazione di una variabile in una prefissata modalità  $i$ -esima cresce con  $f_i$  e, quindi, una misura globale del suo accentramento è espresso da una media delle frequenze relative; per esempio, dalla quantità  $S(\mathbf{f}) = \sum_i f_i f_i = \sum_i f_i^2$ .

Ora, poiché l'incertezza è funzione diretta della variabilità, che è invece funzione inversa dell'accentramento della variabile in una o poche modalità, sembra ragionevole misurare l'incertezza<sup>5</sup> (e, quindi, la variabilità, la eterogeneità, etc.) di una variabile ordinale mediante funzioni inverse di  $S(\mathbf{f})$ .

A tale proposito un indice proposto in ambito socio-politologico è quello attribuito a Laakso e Taagepera (1979) e definito come l'inverso di  $S(\mathbf{f})$ . Tale misura varia tra 1 ed  $m$  ed esprime il "numero di partiti rilevanti in un'assemblea elettiva" costituita da  $m$  partiti, se ciascun partito raccoglie una proporzione di consensi (voti, seggi, delegati, etc.) misurata da  $f_i$ ,  $i = 1, 2, \dots, m$ .

Nella sua formulazione normalizzata in  $[0, 1]$ , l'indice di Laakso e

---

<sup>5</sup> Esiste anche una interpretazione probabilistica della quantità  $S(\mathbf{f})$ , molto diffusa in un contesto ecologico. Infatti, date  $m$  differenti specie in un territorio ben definito, essa misura la probabilità che due individui scelti a caso appartengano alla stessa specie.

Taagepera è definito da<sup>6</sup>:

$$\mathcal{A} = \frac{1}{m-1} \left\{ \left( \sum_{i=1}^m f_i^2 \right)^{-1} - 1 \right\}.$$

Alla fine, è agevole concludere che queste misure sono tutte trasformazioni monotone dell'indice (normalizzato)  $\mathcal{G}$  di Gini:

$$\mathcal{F} = 1 - \sqrt{1 - \mathcal{G}}; \quad \mathcal{A} = \frac{\mathcal{G}}{m - \mathcal{G}(m-1)}.$$

Inoltre, si dimostra che  $\mathcal{F} \leq \mathcal{G}$  uniformemente, mentre  $\mathcal{A}(\mathcal{G}) \leq \mathcal{F}(\mathcal{G})$  solo per  $\mathcal{G} \in [0, m(m-2)/(m-1)^2]$ . Quindi, per i valori usuali di  $m$  e di  $\mathcal{G}$ , l'indice  $\mathcal{A}$  è sostanzialmente inferiore agli altri due.

Sulla base di tali relazioni, l'uso di differenti indici di diversità è giustificato solo per valutare se uno di essi è maggiormente selettivo rispetto agli altri. Ora, per il nostro data set, le differenti misure (presentate nella Tabella 3) producono lo stesso ordinamento; tuttavia, l'indice  $\mathcal{A}$  possiede il campo di variazione più ampio, per cui risulta maggiormente adeguato per discriminare tra situazioni eterogenee. La Figura 2 evidenzia tali risultati rispetto all'indice  $\mathcal{A}$ , ordinatamente per le 9 emergenze.

Tabella 3. Indici di eterogeneità.

<b>Emergenze</b>	$\mathcal{L}$	$\mathcal{G}$	$\mathcal{F}$	$\mathcal{A}$
<i>Clientelismo e corruzione</i>	0.51540	0.94118	0.75747	0.64002
<i>Criminalità organizzata</i>	0.23277	0.63446	0.39540	0.16168
<i>Disoccupazione</i>	0.45373	0.90278	0.68819	0.50781
<i>Inquinamento e carenza verde</i>	0.45927	0.92432	0.72490	0.57575
<i>Mala-sanità</i>	0.44945	0.92046	0.71797	0.56252
<i>Micro-criminalità</i>	0.44016	0.90642	0.69409	0.51836
<i>Presenza di immigrati</i>	0.38659	0.72556	0.47613	0.22706
<i>Pulizia e igiene territorio</i>	0.42251	0.90599	0.69339	0.51709
<i>Trasporti-traffico-mobilità</i>	0.48221	0.91154	0.70257	0.53378
<b>Campo di variazione degli indici</b>	<b>0.28263</b>	<b>0.30672</b>	<b>0.36207</b>	<b>0.47834</b>

<sup>6</sup> L'indice  $\mathcal{A}$  è stato proposto, indipendentemente, da Piccolo (1998, 153), sulla base del principio che tutti i più comuni indici di eterogeneità sono inversamente connessi alla quantità  $S(f)$ . Infatti, è agevole mostrare che gli indici di Gini, di Simpson e di Laakso e Taagepera sono tutti trasformazioni (del tipo Box e Cox) di tale quantità.

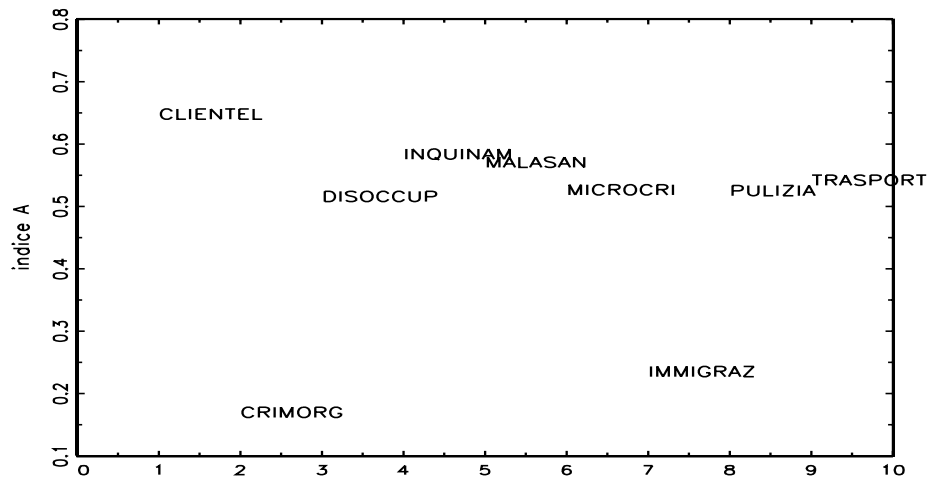


Figura 2. Ordinamento delle emergenze rispetto all'indice A.

Come si vede, la eterogeneità è contenuta per le variabili “Criminalità organizzata” e “Presenza di immigrati” mentre è più elevata per la variabile “Clientelismo e corruzione” rispetto alla quale le risposte, evidentemente, sono più difformi. Queste evidenze mostrano anche che i rispondenti hanno espresso valutazioni più concordi circa il posizionamento delle emergenze “Criminalità organizzata” e “Presenza di immigrati”, anche se – in termini di livello della percezione di gravità – esse si collocano in posizione ben diversa, come sarà confermato da altri indicatori.

### 5. Un modello per la percezione della gravità delle emergenze

Nella letteratura statistica, il trattamento di dati ordinali è da tempo oggetto di studio mediante numerose metodologie che includono analisi esplorative, studi inferenziali e modelli di comportamento: Agresti (2002), Johnson and Albert (1999), Simonoff (2003).

Qui di seguito, ci limitiamo a verificare sull'insieme dei dati che stiamo esaminando l'applicabilità e la utilità interpretativa di un modello mi-

stura che esplicita e quantifica la natura composita delle scelte umane nel formulare le graduatorie.

Rinviando alla bibliografia per una discussione critica dell'approccio<sup>7</sup>, ci limitiamo ad un breve richiamo del modello *MUB*, orientando la presentazione allo studio della percezione delle emergenze metropolitane.

In tale modello, si interpreta la risposta fornita dall'intervistato come la realizzazione di una variabile casuale (v.c.) discreta  $R$  che possiede distribuzione di probabilità:

$$P_r(R = r) = \pi b_r(\xi) + (1 - \pi) u_r(m), \quad r = 1, 2, \dots, m.$$

Abbiamo indicato con:

$$\begin{cases} b_r(\xi) &= \binom{m-1}{r-1} \xi^{r-1} (1-\xi)^{m-r}; \\ u_r(m) &= \frac{1}{m}; \end{cases}$$

rispettivamente, la distribuzione di una v.c. *Binomiale traslata* e quella di una v.c. *Uniforme discreta*, entrambe definite sul supporto  $\{1, 2, \dots, m\}$ .

Quindi, la variabile casuale  $R$  viene indicata con  $R \sim MUB(\pi, \xi)$  perché essa è una **Mistura** tra una v.c. *Uniforme discreta* ed una v.c. *Binomiale traslata*.

La prima componente è la risultante del processo decisionale che produce un giudizio di maggiore o minore gravità (ed è parametrizzata da  $\xi$ ); la seconda esplicita il grado di incertezza che è associato a tale scelta e che si ripercuote uniformemente su tutte le modalità (con una intensità parametrizzata da  $\pi$ ).

Se, come sempre avviene, si ipotizza che  $m$  sia noto e fissato in anticipo (per esempio, nel nostro data set,  $m = 9$ ), la distribuzione di pro-

---

<sup>7</sup> Il modello *MUB* è stato introdotto da D'Elia (2003b) e Piccolo (2003) ed ampiamente discusso in D'Elia (2004) e D'Elia e Piccolo (2005), sia con riferimento alle problematiche statistiche che a quelle computazionali. Tale modello è stato applicato in numerosi contesti, anche mediante la introduzione di covariate del soggetto. Tra questi studi ricordiamo le analisi sui giudizi di preferenze, le indagini di marketing, la valutazione della didattica, le analisi sensoriali, la percezione del dolore negli studi clinici, la percezione della sinonimia nella linguistica, etc.

babilità è specificata dal vettore dei parametri  $\theta = (\pi, \xi)$ . Essi sono entrambi definiti su  $[0, 1]$  ed una loro disamina consente una immediata interpretazione nei termini del giudizio di gravità che si sta esprimendo. Più specificamente, il parametro  $\pi$  è inversamente legato alla quota di incertezza insita nel giudizio espresso, mentre – a parità di altre circostanze – il parametro  $\xi$  cresce in misura diretta con la gravità espressa<sup>8</sup>.

Tale distribuzione è estremamente malleabile, capace quindi di adattarsi alle più varie situazioni reali, in termini di asimmetria, curtosi e peso relativo delle code. In particolare, essa possiede una moda che può variare su tutto il campo di definizione.

Inoltre, la struttura probabilistica può essere collegata alle caratteristiche del rispondente facendo in modo che la probabilità sia una funzione delle covariate del soggetto; questo avviene esplicitando un legame tra i parametri del modello *MUB* ed i valori delle variabili che caratterizzano il rispondente, come discuteremo nel paragrafo 8.

A tale riguardo, utilizzeremo la seguente terminologia e la corrispondente simbologia:

**MUB-00:** modello *senza covariate*, con parametri  $\theta = (\pi, \xi)$ ;

**MUB-10:** modello *con covariate per  $\pi$* , con parametri  $\theta = (\beta, \xi)$ ;

**MUB-01:** modello *con covariate per  $\xi$* , con parametri  $\theta = (\pi, \gamma)$ ;

**MUB-11:** modello *con covariate per  $\pi, \xi$* , con parametri  $\theta = (\beta, \gamma)$ .

Nel seguito, presenteremo e commenteremo prima i risultati ottenuti mediante i modelli *MUB-00*, mentre nei paragrafi successivi discuteremo quelli derivanti dalla introduzione di covariate nel modello.

---

<sup>8</sup> Il modello *MUB* è *reversibile*, cioè se la graduatoria viene posta in ordine inversa il parametro  $\pi$  non si modifica mentre  $\xi$  diventa  $1 - \xi$ . Cioè, formalmente:  $R \sim MUB(\pi, \xi) \implies (m - R + 1) \sim MUB(\pi, 1 - \xi)$ .

In effetti, il criterio di ordinamento rispetto alla gravità del problema che si intende studiare è il risultato di una convenzione (ineliminabile). Operativamente, se si chiede di ordinare le emergenze da quella ritenuta più grave a quella meno grave, allora quando la distribuzione dei ranghi risulta accentrata sui valori bassi la gravità del problema che si studia è elevata così come il valore del parametro  $\xi$ . Il viceversa avviene nei casi in cui si conviene che un rango elevato corrisponda ad un'emergenza ritenuta grave.

## 6. Un modello statistico per gli ordinamenti

L'adattamento tra osservazioni ordinali e modello *MUB* avviene mediante il metodo della massima verosimiglianza, utilizzando in modo efficiente l'algoritmo E-M, come è tipico nell'inferenza sui modelli mistura (McLachlan and Krishnan, 1997; McLachlan and Peel, 2000).

Le stime dei parametri ottenute sul nostro data set sono riportate nella Tabella 4, assieme al corrispondente errore standard asintotico (*es*); come si vede, esse sono tutte significative.

Tabella 4. Stima dei modelli *MUB* per le emergenze metropolitane.

<b>Emergenze</b>	$\hat{\pi}$	$es(\hat{\pi})$	$\hat{\xi}$	$es(\hat{\xi})$
<i>Clientelismo e corruzione</i>	0.702	0.047	0.663	0.017
<i>Criminalità organizzata</i>	0.937	0.017	0.940	0.005
<i>Disoccupazione</i>	0.810	0.034	0.820	0.009
<i>Inquinamento e carenza verde</i>	0.838	0.036	0.279	0.011
<i>Mala-sanità</i>	0.849	0.041	0.502	0.012
<i>Micro-criminalità</i>	0.848	0.035	0.724	0.010
<i>Presenza di immigrati</i>	0.756	0.030	0.051	0.006
<i>Pulizia e igiene territorio</i>	0.892	0.033	0.310	0.010
<i>Trasporti-traffico-mobilità</i>	0.751	0.037	0.178	0.010

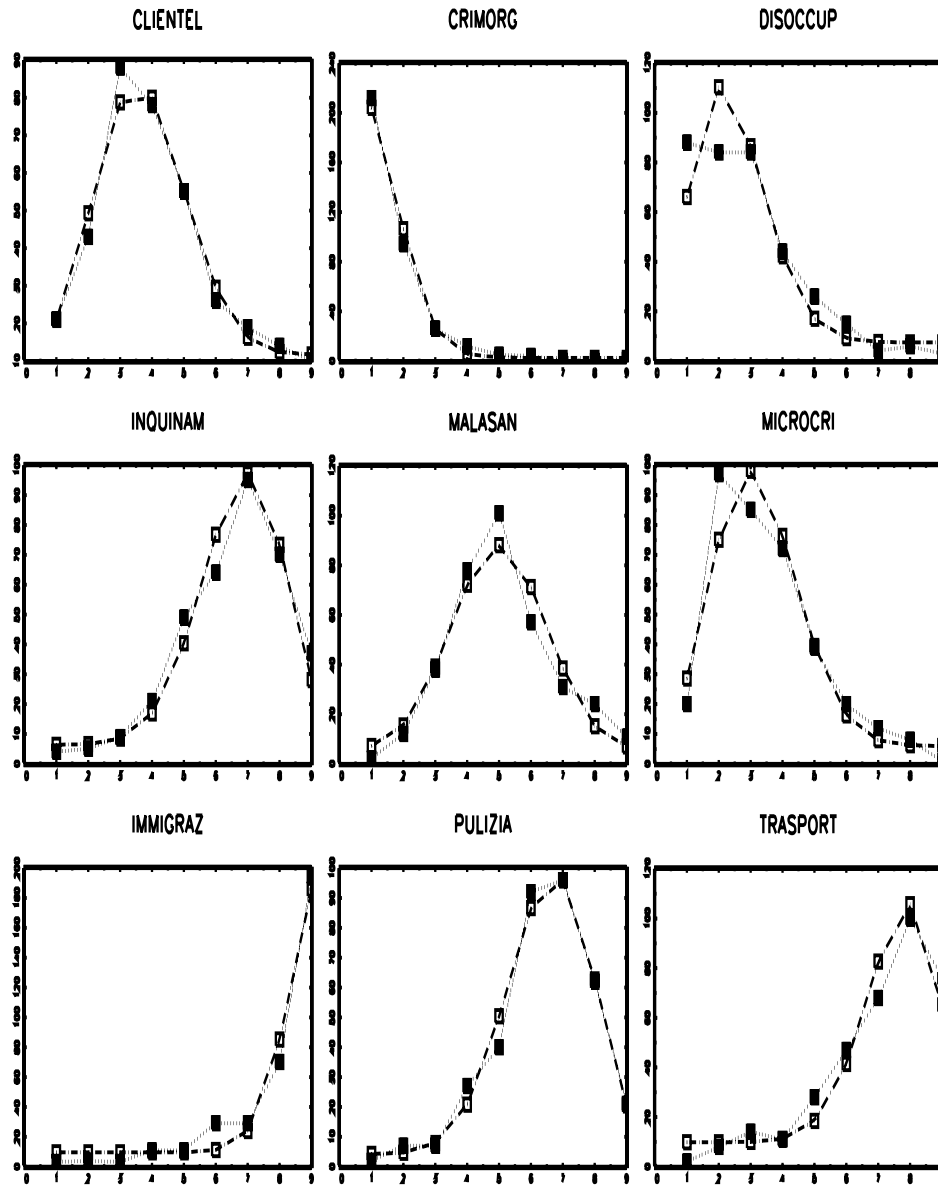
La Figura 3 mostra che l'adattamento tra frequenze osservate e frequenze attese (calcolate sulla base dei modelli *MUB* stimati) è generalmente molto buono e conferma, in particolare, la notevole capacità del modello di adattarsi a distribuzioni empiriche dalle forme molto varie<sup>9</sup>.

D'altra parte, è noto che per dati qualitativi le tradizionali misure di adattamento, come l'indice  $X^2$  di Pearson, non sono adeguate, soprattutto quando la numerosità delle osservazioni è elevata; è quindi opportuno riferirsi ad una pluralità di indicatori per una valutazione più articolata.

---

<sup>9</sup> In effetti, per le due variabili "Criminalità organizzata" e "Presenza di immigrati" si potrebbe adattare un modello *IHG*, proposto da D'Elia (1999; 2003a), e discusso in D'Elia e Piccolo (2002), essendo in presenza di valutazioni concentrate su modalità estreme. Tuttavia, abbiamo preferito discutere le emergenze all'interno di una sola classe di modelli, anche per uniformare i confronti e sottolineare la semplicità di interpretazione del modello *MUB*.





\*

Figura 3. Frequenze osservate (-■-■-■-) ed attese (-□-□-□-) secondo il modello MUB.

Se indichiamo con  $f_i$ ,  $e_i(\boldsymbol{\theta})$ ,  $i = 1, 2, \dots, m$ , rispettivamente, le frequenze relative osservate e le frequenze relative attese nell'ipotesi che il modello sia corretto, l'indice di Pearson è definito da:

$$X^2 = \sum_{i=1}^m \frac{(f_i - e_i(\boldsymbol{\theta}))^2}{e_i(\boldsymbol{\theta})},$$

mentre l'indice di dissimilarità normalizzato<sup>10</sup> è definito da:

$$Diss = \frac{1}{2} \sum_{i=1}^m |f_i - e_i(\boldsymbol{\theta})|.$$

Questo secondo indice (detto anche di *dissomiglianza*) ha un'agevole interpretazione perché esprime la quota di unità statistiche che dovrebbe essere trasferita tra le categorie ordinali per ottenere un adattamento perfetto.

Nella Tabella 5, oltre agli indici  $X^2$  e  $Diss$ , è riportato il valore della log-verosimiglianza perché, nei modelli considerati, sono costanti sia la famiglia parametrica da adattare che la numerosità campionaria. Quindi, la log-verosimiglianza può essere interpretata come una misura inversa della bontà di adattamento.

Tabella 5. Misure di accostamento dei modelli MUB.

<b>Emergenze</b>	$X^2$	(p-value)	$Diss$	$\log - lik$
<i>Clientelismo e corruzione</i>	3.299	(0.771)	0.039	-698.851
<i>Criminalità organizzata</i>	18.205	(0.006)	0.053	-397.386
<i>Disoccupazione</i>	26.969	(< 1 * 10 <sup>-4</sup> )	0.109	-641.296
<i>Inquinamento e carenza verde</i>	9.329	(0.156)	0.063	-664.530
<i>Mala-sanità</i>	18.531	(0.005)	0.089	-672.382
<i>Micro-criminalità</i>	18.709	(0.005)	0.089	-649.472
<i>Presenza di immigrati</i>	45.583	(< 4 * 10 <sup>-8</sup> )	0.096	-518.747
<i>Pulizia e igiene territorio</i>	6.775	(0.342)	0.041	-642.026
<i>Trasporti-traffico-mobilità</i>	18.740	(0.005)	0.085	-656.020

<sup>10</sup> Leti (1979); (1983), 528-530; Girone e Salvemini (1986), 15-16; Piccolo (2000), 71-73; Simonoff (2003), 77-78. Va aggiunto che, alla maggiore semplicità interpretativa di  $Diss$  rispetto a  $X^2$ , si oppone il fatto che la teoria asintotica del test di Pearson è più consolidata. Nel  $MUB - 00$ , il valore di  $X^2$  va confrontato con il quantile della v.c.  $\chi^2_{(g)}$ , con  $g = 6$  gradi di libertà, essendo  $m = 9$  e 2 i parametri stimati.

Si conferma così che l'adattamento si può considerare eccellente (l'indice di dissomiglianza si mantiene al di sotto della soglia<sup>11</sup> del 9%), eccetto per le variabili "Disoccupazione" e "Presenza di immigrati" per le quali va segnalata qualche difformità. Ora, mentre per la seconda è stata già segnalata un notevole accentramento nelle risposte (una sovradisersione che potrebbe essere spiegata da un modello differente, come accennato nella nota 8), per l'emergenza "Disoccupazione" si rileva che la difformità concerne solo le prime 3 modalità di risposta (di fatto, quasi costanti): quest'ultimo comportamento deriva verosimilmente da strati diversi fra i rispondenti le cui risposte oscillano fra le prime 3 modalità in quote più o meno costanti. Tali questioni saranno riprese nel paragrafo 10 in connessione alla particolare distribuzione dell'età del campione dei rispondenti.

### ***7. Interpretazione dei parametri stimati***

Un'analisi comparata dei vari modelli stimati può essere eseguita se indichiamo sullo spazio parametrico i punti corrispondenti alle stime dei parametri, pervenendo così allo schema della Figura 4 (per semplicità, è stato disegnato solo il sottoinsieme dello spazio parametrico di interesse).

L'ordinamento nello spazio parametrico conferma in modo ancor più immediato quanto poteva dedursi dal semplice esame dei valori numerici dei parametri. In particolare, l'incertezza nel giudizio diminuisce in modo netto per la variabile "Criminalità organizzata", per la quale vi è una quasi concorde valutazione da parte del collettivo esaminato.

Tale incertezza è piuttosto contenuta per quasi tutte le altre variabili (a conferma di una valutazione piuttosto accurata nelle risposte) mentre appare più presente per le variabili "Clientelismo e corruzione", "Trasporti-traffico-mobilità" e "Presenza di immigrati". Queste ultime poi sono anche caratterizzate da una preoccupazione piuttosto modesta (misurata dal-

---

<sup>11</sup> Il senso di questa affermazione deriva dal fatto che, in studi preliminari di simulazione, il livello critico di 0.09 è risultato essere un buon criterio per valutare la significatività di modelli *MUB* analoghi a quelli ottenuti nel testo e con dimensioni campionarie comparabili.

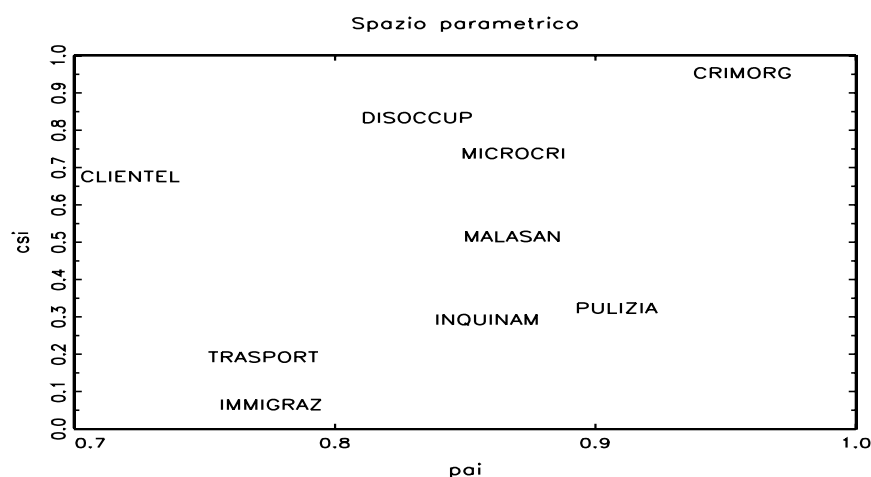


Figura 4. Spazio parametrico derivato dai modelli *MUB* stimati

le stime di  $\xi$ ). Sembrerebbe, cioè, che queste due tematiche non siano percepite né come rilevanti né come concordemente gravi, se paragonate a tutte le altre.

La Figura 4 mostra che in alto (sulla destra dello spazio parametrico) si collocano emergenze collettive (“Criminalità organizzata”, “Disoccupazione”, “Micro-criminalità”) mentre verso il centro troviamo problematiche che riguardano l’individuo ed il suo ambiente (“Mala-sanità”, “Pulizia e igiene territorio”<sup>12</sup>, “Inquinamento e carenza di verde”).

Al confronto, rilievo minore ricevono “Trasporti-traffico-mobilità” e “Presenza di immigrati”: in effetti, la prima emergenza si è notevolmente attenuata negli ultimi anni nelle aree della metropoli napoletana e l’immigrazione non ha mai costituito un problema di dimensioni collettive

<sup>12</sup> In verità, ci si poteva attendere una percezione ben più grave su tale variabile essendo essa la conseguenza visibile di una grave e tuttora irrisolta emergenza regionale costituita dallo smaltimento dei rifiuti urbani. La questione, che coinvolge ambiti più vasti rispetto alla sola area metropolitana, è oggetto di ricorrenti inconvenienti che investono anche l’ordine pubblico e la sicurezza sanitaria per ampi strati della popolazione. Tuttavia, rispetto alle altre emergenze, l’attenzione verso questo aspetto della vita quotidiana si accentua solo in periodi molto concentrati, per cui il cittadino non esprime una percezione permanente di disagio.

in tale area. Infine, va segnalato come la variabile “Clientelismo” (che in passato ha assunto ben altre dimensioni nella percezione collettiva e che, indirettamente, si interseca anche con le questioni connesse alla ricerca di una occupazione) non appaia più un problema molto sentito, essendo accompagnata anche da maggiore incertezza circa la sua precisa collocazione.

In effetti, ripensando al periodo nel quale si è svolta l’indagine (dicembre 2004), l’analisi conferma come la scarsa attenzione dei mass-media verso una certa problematica, nei tempi ravvicinati all’intervista, induce una minore sensibilità verso la sua connotazione di emergenza, allentando la corrispondente percezione di gravità.

Con riferimento alla interpretazione del modello *MUB*, è interessante collegare i valori (ed il significato) dei due parametri agli indicatori statistici più tradizionali. Specificamente, appare ragionevole ipotizzare che l’incertezza misurata da  $\pi$  sia connessa alla dispersione evidenziata dalla variabile ordinale; similmente, il grado di gravità percepita dovrebbe essere collegato alla media aritmetica dei ranghi espressi dal rispondente.

Questa interpretazione è confermata dall’esame dei dati che, per comodità di raffronto, presentiamo<sup>13</sup> nella Tabella 6.

Tabella 6. Stime dei parametri dei modelli *MUB* e indici statistici sui ranghi.

<b>Emergenze</b>	$\hat{\pi}$	Indice $\mathcal{L}$	$\hat{\xi}$	Media $\mu$
<i>Clientelismo e corruzione</i>	0.702	0.515	0.663	4.093
<i>Criminalità organizzata</i>	0.937	0.233	0.940	1.650
<i>Disoccupazione</i>	0.810	0.454	0.820	2.845
<i>Inquinamento e carenza verde</i>	0.838	0.459	0.279	6.531
<i>Mala-sanità</i>	0.849	0.449	0.502	5.105
<i>Micro-criminalità</i>	0.848	0.440	0.724	3.472
<i>Presenza di immigrati</i>	0.756	0.387	0.051	7.915
<i>Pulizia e igiene territorio</i>	0.892	0.423	0.310	6.367
<i>Trasporti-traffico-mobilità</i>	0.751	0.482	0.178	7.023

<sup>13</sup> La scelta dell’indice  $\mathcal{L}$  di Leti deriva dal fatto che tra gli indici di eterogeneità è quello risultato massimamente correlato con le stime di  $\pi$ . D’altra parte, poiché la Tabella 3 evidenzia lo stesso ordinamento per tutti gli indici di eterogeneità ivi elencati, le considerazioni del testo non sarebbero mutate se si fosse scelta un’altra misura.

La Figura 5 conferma che la relazione tra i parametri stimati dai modelli *MUB* ed i corrispondenti indicatori descrittivi ottenuti sulle risposte è ben evidente.

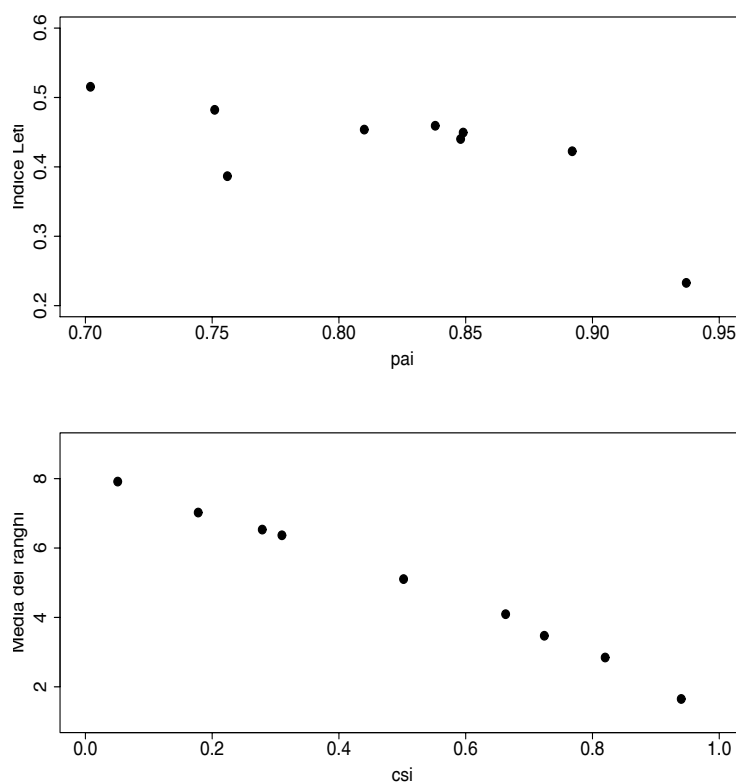


Figura 5. Relazioni tra indici descrittivi e stime dei parametri.

In particolare, da tali grafici si evince che:

- per il parametro  $\pi$ , la relazione con l'indice  $\mathcal{L}$  sembra non lineare;
- per il parametro  $\xi$ , la relazione lineare con la media dei ranghi è molto evidente<sup>14</sup>;

<sup>14</sup> Va osservato che tale adattamento quasi perfetto con una linea retta è anche do-

- la variabilità è inversamente connessa al parametro  $\pi$  mentre la percezione della gravità è direttamente connessa al parametro  $\xi$  che, a sua volta, è inversamente connesso alla media dei ranghi.

La discussione, qui impostata da un punto di vista interpretativo, ha anche un utile risvolto inferenziale. Infatti, nella stima di massima verosimiglianza dei parametri, la individuazione di valori iniziali è un problema rilevante per accelerare la convergenza dell'algoritmo E-M (che è efficiente ma lento, soprattutto nella fase iniziale).

Le considerazioni qui svolte inducono a ritenere che sia possibile individuare come stima iniziale per  $\xi$  quella derivabile dalla media aritmetica dei ranghi, mentre per  $\pi$  occorre cercare una relazione non lineare con un indice di eterogeneità (per esempio, l'indice  $\mathcal{L}$  di Leti).

Allo stadio attuale, l'algoritmo E-M che utilizziamo per la stima di massima verosimiglianza dei parametri del modello *MUB*, pone come valore iniziale per  $\xi$  la quantità  $\hat{\xi}_0 = \frac{m - \mu}{m - 1}$ , essendo  $\mu$  la media aritmetica dei ranghi osservati, mentre per il parametro  $\pi$  si è tuttora alla ricerca di una appropriata determinazione generale<sup>15</sup>.

### 8. La valutazione delle emergenze e le covariate del soggetto

Il modello *MUB-00*, sin qui esaminato, si configura come una struttura probabilistica che consente un efficace adattamento alla distribuzione osservata della percezione di gravità espressa dai rispondenti, mediante l'uso di due soli parametri, connessi rispettivamente all'incertezza decisionale del soggetto ed alla intensità della sua percezione.

E' allora interessante cercare di valutare se ed in che misura le caratteristiche proprie del rispondente siano state rilevanti nello spiegare un

---

vuto al fatto che, per i modelli qui stimati, il valore di  $\pi$  è risultato sempre piuttosto elevato, indicando quindi un maggior peso relativo della componente "Binomiale traslata" rispetto alla componente "Uniforme discreta", il che ha indubbiamente accentuato la forma della relazione lineare tra media osservata dei ranghi e stime di  $\xi$ ; infatti, se fosse esattamente  $\pi = 1$  tale relazione sarebbe di tipo matematico.

<sup>15</sup> Per il nostro data set, una buona approssimazione è risultata essere quella definita dalla relazione:  $\hat{\pi}_0 = 1.06 - 0.08 (\mathcal{L} - 1)^{-2}$ .

comportamento differenziato, sia in termini di incertezza che di gravità percepita. Si tratta, cioè, secondo un'ottica propria dei modelli lineari generalizzati (*GLM*)<sup>16</sup>, di inserire in modo esplicito nel modello statistico le covariate del soggetto in modo da stimarne il peso e la significatività.

Una differenza tuttavia caratterizza la nostra classe di modelli e va esplicitata. Mentre, nei *GLM* le covariate sono funzioni lineari di una trasformazione del valore medio della variabile risposta tramite una funzione *link*, nel nostro caso il legame è definito tramite i singoli parametri e le covariate.

Tralasciando gli aspetti tecnici di tale problematica, precisiamo che:

- i) la funzione legame da noi prescelta è di tipo logistico poiché è necessario introdurre una relazione tra i parametri definiti sull'intervallo unitario e le covariate del soggetto le quali, almeno in linea di principio, sono definite sull'asse reale;
- ii) nel modello *MUB* il valore medio non è interpretabile *separatamente* per i due parametri, per cui non appare immediata una soluzione assimilabile a quella generalmente presupposta nei *GLM*.

In tal modo, se indichiamo<sup>17</sup> con  $\mathbf{y}_i$  le covariate concernenti il parametro  $\pi$  e con  $\mathbf{w}_i$  quelle concernenti il parametro  $\xi$ , per ciascun soggetto  $j$ -esimo, per  $j = 1, 2, \dots, n$ , tali legami sono così formalizzabili:

$$(\pi | \mathbf{y}_j) = \frac{1}{1 + e^{-\beta' \mathbf{y}_j}}; \quad (\xi | \mathbf{w}_j) = \frac{1}{1 + e^{-\gamma' \mathbf{w}_j}}.$$

In particolare, è possibile esprimere il valore medio dei ranghi condizionati ai valori delle covariate. Infatti, se supponiamo, per semplicità, che la covariata sia una sola (per ciascun parametro), si derivano i valori medi riportati nel riquadro seguente. Va chiarito che l'ultima colonna indica i valori medi *al netto del valore centrale*, costante nei vari modelli.

---

<sup>16</sup> Per una disamina approfondita di tale classe di modelli, si rinvia a McCullagh e Nelder (1998). Va precisato che il modello *MUB* non appartiene alla famiglia esponenziale, che invece è quella generalmente presupposta per i *GLM*.

<sup>17</sup> Maggiori dettagli sulla notazione che stiamo utilizzando si trovano in Piccolo (2003), 30-31; una discussione sull'interpretazione del segno dei parametri stimati sulle covariate in funzione del valore medio dei ranghi è svolta da D'Elia (2003), 4-6; 10-11.



<b>Modelli</b>	<b>Parametri</b>	$\mathbb{E}(R \mid \text{covariate}) - (m + 1)/2$
<i>MUB-00</i>	$\pi, \xi$	$\pi(m - 1) \left( \frac{1}{2} - \xi \right)$
<i>MUB-10</i>	$\pi = \pi(\mathbf{y}), \xi$	$\frac{m - 1}{1 + e^{-\beta_0 - \beta_1 \mathbf{y}_j}} \left( \frac{1}{2} - \xi \right)$
<i>MUB-01</i>	$\pi, \xi = \xi(\mathbf{w})$	$\pi(m - 1) \left( \frac{1}{2} - \frac{1}{1 + e^{-\gamma_0 - \gamma_1 \mathbf{w}_j}} \right)$
<i>MUB-11</i>	$\pi = \pi(\mathbf{y}), \xi = \xi(\mathbf{w})$	$\frac{m - 1}{1 + e^{-\beta_0 - \beta_1 \mathbf{y}_j}} \left( \frac{1}{2} - \frac{1}{1 + e^{-\gamma_0 - \gamma_1 \mathbf{w}_j}} \right)$

Appare evidente, allora, che per esplicitare l'effetto delle covariate sulla risposta ordinale si possano adottare soluzioni differenti in funzione della natura della covariate. Più precisamente:

**Covariata dicotomica:** In tal caso, è preferibile disegnare le distribuzioni di probabilità nei due valori assunti dalla covariata, deducendo così differenti interpretazioni in termini di posizione e incertezza.

**Covariata politomica:** Si può seguire la procedura del caso dicotomico, sviluppando così le distribuzioni di probabilità da confrontare. In alternativa, si possono calcolare i differenti profili che i valori della covariata inducono sui parametri e sul valore medio.

**Covariata continua:** Anche se possono essere confrontate alcune specifiche distribuzioni di probabilità, per prefissati valori della covariata, appare più comune disegnare come si modifica nel continuo il valore medio della variabile ordinale al variare della covariata. Spesso, nei modelli *MUB-11*, una delle covariate è continua e l'altra è dicotomica/politomica, per cui appare efficace mostrare

l'evoluzione del valore medio dei ranghi al variare della covariata (continua), condizionatamente al valore assunto dalla covariata dicotomica/politomica.

Tali profili sono generalmente molto efficaci e saranno utilizzati frequentemente nelle pagine seguenti; essi però insistono sulla distribuzione nel suo complesso e/o sul valore medio della variabile ordinale.

In effetti, va sottolineato come sinora si sia posto scarso rilievo al ruolo che la covariata può avere sulla modifica dell'incertezza decisionale del soggetto; invece, questo elemento è significativo, come il nostro data set conferma per più variabili. Infatti, la nostra proposta è quella di sviluppare analisi omologhe a quelle precedenti, utilizzando un indicatore di eterogeneità per variabili ordinali ed esaminare il suo comportamento al variare della covariata.

Poiché (come è emerso nel paragrafo 4), tali misure descrittive sono tutte funzioni di  $\sum f_i^2$  e, quindi, – a livello di popolazione – di  $\sum p_i^2$ , essendo  $p_i = P_r(R = i | \theta)$ ,  $i = 1, 2, \dots, m$ , le probabilità implicate dal modello in esame, occorre studiare le variazioni di una di tali misure in funzione della covariata di interesse.

Se si sceglie l'indice  $\mathcal{A}$ , che è risultato maggiormente selettivo nelle analisi esplorative, si tratta di studiare – in funzione della covariata  $y$ , definita in  $y_{min} \leq y \leq y_{max}$  – la seguente funzione:

$$\mathcal{A}(y) = \frac{1}{m-1} \left\{ \left( \sum_{i=1}^m [P_r(R = i | \pi(y))]^2 \right)^{-1} - 1 \right\}.$$

Il grafico di  $\mathcal{A}(y)$  in funzione di  $y \in [y_{min}, y_{max}]$  consente una efficace interpretazione dell'effetto della covariata sulla incertezza che accompagna la percezione della gravità dell'emergenza. Una esemplificazione di tale proposta sarà sviluppata nel paragrafo 11.

## 9. Il genere nella percezione delle emergenze

L'analisi statistica del data set ha richiesto che dal complesso delle covariate disponibili sui soggetti rispondenti si scegliessero quelle signi-

ficative. Tale fase si è svolta mediante l'interazione tra analisi esplorative per sottogruppi di soggetti simili e test preliminari<sup>18</sup>.

Dei vari modelli possibili, ci limiteremo ad enucleare i risultati che ci sembrano più significativi ed interpretabili ai fini dello studio della percezione delle emergenze metropolitane. In particolare, la scelta mira ad evidenziare le potenzialità di un approccio modellistico nella disamina di variabili ordinali.

Esaminiamo, anzitutto, i modelli *MUB-10* e *MUB-01* che esplicitano il ruolo delle covariate nel differenziare, rispettivamente, l'incertezza nella risposta oppure la gravità dell'emergenza, così come espresse dall'intervistato.

La tabella successiva mostra i risultati ottenuti quando la covariata utilizzata è il *genere* del rispondente (che nel data set è stato codificato con 0 e 1, rispettivamente, per i maschi e le femmine).

Tabella 7. Stime dei modelli *MUB* in funzione del genere.

<b>Emergenze</b>	Modelli MUB-10		Modelli MUB-01	
	Stime di $\pi$		Stime di $\xi$	
	<i>Maschi</i>	<i>Femmine</i>	<i>Maschi</i>	<i>Femmine</i>
<i>Clientelismo e corruzione</i>	0.661	0.761	<i>0.658</i>	<i>0.668</i>
<i>Criminalità organizzata</i>	<i>0.933</i>	<i>0.943</i>	0.929	0.955
<i>Disoccupazione</i>	0.854	0.729	0.834	0.792
<i>Inquinamento e carenza verde</i>	<i>0.799</i>	<i>0.880</i>	0.318	0.230
<i>Mala-sanità</i>	0.777	0.943	0.479	0.528
<i>Micro-criminalità</i>	0.800	0.909	<i>0.728</i>	<i>0.721</i>
<i>Presenza di immigrati</i>	0.820	0.627	0.039	0.098
<i>Pulizia e igiene territorio</i>	0.928	0.844	0.329	0.281
<i>Trasporti-traffico-mobilità</i>	<i>0.773</i>	<i>0.721</i>	<i>0.190</i>	<i>0.160</i>

Nella tabella, abbiamo evidenziato in corsivo i confronti che *non* sono risultati significativamente differenti. In tali casi, il genere del rispondente

<sup>18</sup> Questi aspetti, sui quali non ci soffermiamo, sono oggetto di ricerca corrente ed impongono la individuazione di indicatori differenti rispetto a quelli tradizionali; per esempio, è tuttora irrisolto il problema della selezione preliminare delle variabili più idonee per un modello *MUB* a partire dalla totalità delle variabili disponibili in un data set.

non costituisce una covariata rilevante per differenziare la tipologia della risposta fornita, con riferimento all'incertezza (se riferita alla stima di  $\pi$ ) oppure all'intensità della gravità percepita (se riferita alla stima di  $\xi$ ).

Per l'incertezza, si nota come essa sia indipendente dal genere per le variabili "Criminalità organizzata", "Inquinamento e carenza verde" e "Trasporti-traffico-mobilità", mentre è un fattore rilevante per spiegare la mutevolezza nelle risposte per le altre 6 emergenze. Invece, per quanto riguarda la percezione di gravità, si rileva come il genere sia significativo per tutte le emergenze eccetto che per le variabili "Clientelismo e corruzione", "Micro-criminalità" e "Trasporti-traffico-mobilità".

In sintesi, i modelli stimati evidenziano che il genere non costituisce una covariata utile per spiegare la differente tipologia delle risposte fornite *solo* per quanto concerne le problematiche della mobilità territoriale (le quali comunque, all'epoca, non sono state giudicate particolarmente rilevanti). Invece, il genere è una covariata importante per giustificare la differente incertezza manifestata e/o la differente gravità percepita per la maggior parte delle emergenze.

In particolare, studiamo come si manifesta l'effetto del genere sulla percezione di gravità dell'emergenza "Inquinamento e carenza verde", a parità di incertezza; infatti, come abbiamo visto, la stima di  $\pi$  non è risultata significativamente alterata dal genere.

Il grafico delle due distribuzioni di probabilità (Figura 6) evidenzia nettamente che per i maschi esiste una preoccupazione maggiore mentre per le donne tale problematica assume per la gran parte valori elevati nella scala ordinale. A conferma, il valore medio condizionato è piuttosto differente tra i due generi (essendo pari a 6.292 per i maschi e 6.876 per le femmine, rispettivamente) mentre la moda (qui pari a 7 per i due generi) non è selettiva. Il grafico manifesta anche come la presenza di una covariata così netta possa migliorare la capacità di adattamento del modello, e quindi la conseguente previsione.

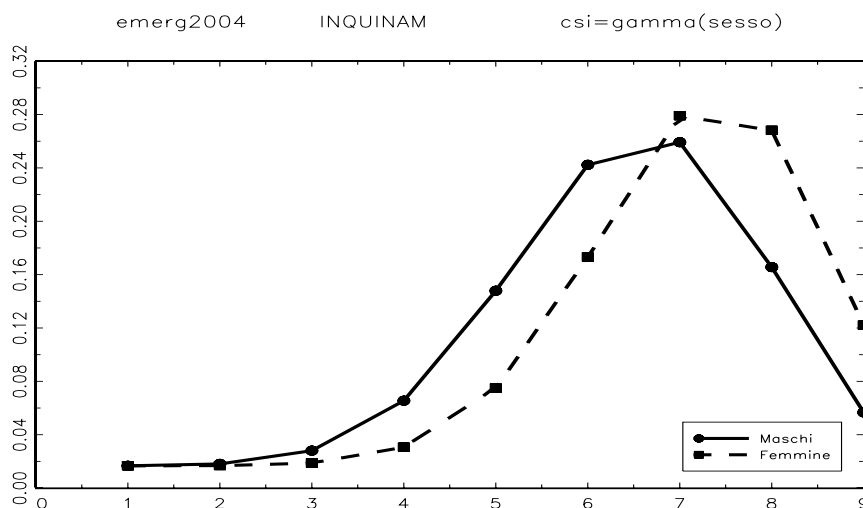


Figura 6. Percezione della gravità, condizionata al genere.

### 10. L'età nella percezione delle emergenze

Il particolare campione che stiamo esaminando consiste di un numero elevato di persone adulte, dipendenti di Enti pubblici e privati, i quali frequentano i corsi allo scopo di pervenire al conseguimento di un titolo di studio universitario<sup>19</sup>.

Questo aspetto rende il campione abbastanza diverso da una popolazione usuale di studenti universitari (anzitutto rispetto alla età) e l'interpretazione dei risultati non può prescindere da questa caratterizzazione. Per esempio, il reddito medio dei rispondenti e l'autonomia nelle scelte e nei comportamenti sono notevolmente superiori a quelli che, generalmente, ci si attende da giovani universitari.

Tuttavia, tale circostanza produce anche un vantaggio statistico perché

<sup>19</sup> Tali studenti usufruiscono quasi sempre di disposizioni legislative che consentono dei permessi retribuiti per la formazione; inoltre, nel periodo considerato, un significativo numero di questi si è giovato di apposite convenzioni che prevedono il riconoscimento di crediti universitari formativi maturati in percorsi di studio progressi rispetto alla iscrizione all'Università.

consente di dare rilievo al ruolo della covariata *età* nello spiegare la diversità delle risposte, senza ricercare altre componenti esterne. In sostanza, mentre le risposte fornite non possono essere omologate a quelle di altri giovani universitari, la maggiore variabilità delle età del nostro campione consente di esplicitare l'effetto aggiuntivo di questa variabile a parità di altre circostanze: genere, ambiente di provenienza, livello culturale, etc.

Specificamente, la Tabella 8 evidenzia, rispetto al genere e complessivamente, gli indicatori più significativi per dimensionare la componente *età* nel campione disponibile.

Tabella 8. Indicatori della covariata *età*, per genere.

<b>Genere</b>	<i>n</i>	Età dichiarata ( <i>in anni</i> )				
		minima	massima	mediana	media	≤ 30 anni
<i>Maschi</i>	209	18.167	54.500	22.583	28.448	61.2%
<i>Femmine</i>	145	18.083	44.417	19.583	22.741	84.8%
<i>Complesso</i>	354	18.083	54.500	20.500	26.111	70.9%

Le donne sono sistematicamente più giovani dei loro colleghi maschi e la dicotomia di composizione per età fra i due generi emerge in modo ancor più netto se esaminiamo i due istogrammi perequati (Figura 7). Infatti, dai grafici appare netta la presenza di due sotto-popolazioni per i due generi, e ciò in modo più massiccio per il genere maschile.

In particolare, dalle distribuzioni sopra individuate, si può affermare che l'elemento di separazione tra le due sotto-popolazioni è costituito, per entrambi i generi, approssimativamente, dall'età di 30 anni<sup>20</sup>.

Esaminiamo, allora, con maggiore dettaglio, come la percezione della gravità di alcune emergenze possa dipendere dalla covariata *età* del soggetto.

---

<sup>20</sup> In effetti, definendo una variabile dicotomica che assume valore 1 per i rispondenti che superano i 30 anni e 0 nel caso contrario, è possibile mostrare che i corrispondenti modelli *MUB-10* e *MUB-01* presentano effetti significativi rispetto all'età, per quasi tutte le emergenze. Tuttavia, disponendo dell'età sotto forma di una variabile continua, riteniamo preferibile valutare il suo effetto in modo diretto.

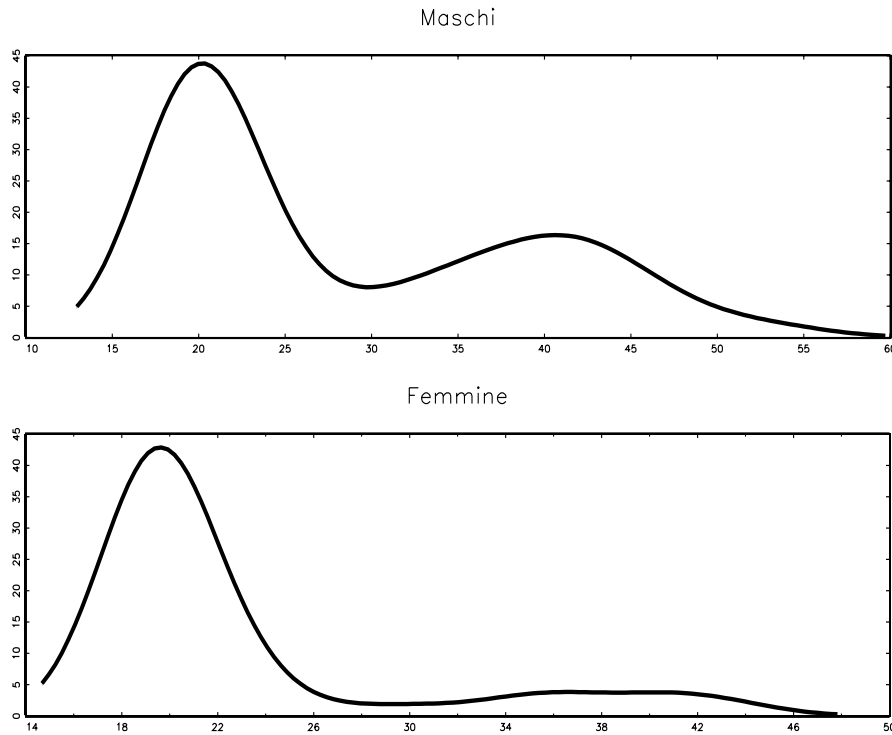


Figura 7. Distribuzioni perequate delle età, per genere.

- **Disoccupazione**

Tale emergenza consente di introdurre il modello *MUB-11*, nel quale entrambi i parametri sono funzioni di covariate del soggetto. In particolare, risulta che la quota di incertezza della risposta dipende dal *genere* del rispondente mentre la percezione della gravità è funzione della sua *età*.

Utilizzando la parametrizzazione del paragrafo 8, la Tabella 9 riporta le stime ottenute (con il rispettivo errore standard, in parentesi); inoltre, per ciascun modello, sequenzialmente, è indicato il valore della log-verosimiglianza nel punto di massimo.

Tabella 9. Stime dei modelli MUB, in funzione del genere e dell'età.

<b>Modelli stimati</b>	$\hat{\pi} = \pi(\text{genere})$	$\hat{\xi} = \xi(\text{età})$	<i>log-lik</i>
<i>MUB-00</i>	$\hat{\pi} = 0.810 (0.034)$	$\hat{\xi} = 0.820 (0.009)$	-641.29588
<i>MUB-10</i>	$\hat{\beta}_0 = 1.769 (0.196)$ $\hat{\beta}_1 = -0.781 (0.271)$	$\hat{\xi} = 0.822 (0.008)$	-639.89134
<i>MUB-01</i>	$\hat{\pi} = 0.813 (0.021)$	$\hat{\gamma}_0 = 1.207 (0.161)$ $\hat{\gamma}_1 = 0.012 (0.006)$	-639.96223
<i>MUB-11</i>	$\hat{\beta}_0 = 1.759 (0.195)$ $\hat{\beta}_1 = -0.734 (0.271)$	$\hat{\gamma}_0 = 1.243 (0.162)$ $\hat{\gamma}_1 = 0.011 (0.011)$	-638.75537

Per valutare la significatività del modello più complesso (con covariate) rispetto a quello iniziale, si confrontano le due log-verosimiglianze mediante la differenza tra le *deviance*<sup>21</sup>:

$$\begin{aligned} \Delta dev &= dev_{00} - dev_{11} = 2[\log-lik_{11} - \log-lik_{00}] \\ &= 2 * (-638.75537 - (-641.29588)) = 5.08102. \end{aligned}$$

Poiché  $\Delta dev \stackrel{a}{\sim} \chi^2_{(g)}$ , ove  $g$  è la differenza tra il numero di parametri tra il modello esteso e quello ridotto, e tenuto conto che  $\chi^2_{(0.05;2)} = 5.991$ , la riduzione apportata dalle variabili *genere* ed *età* appare debolmente significativa (il suo *p-value* è 0.079).

Una rappresentazione efficace di tali risultati può essere ottenuta se si esaminano i valori medi dei ranghi espressi, al variare dell'età e condizionati al genere dell'intervistato (Figura 8). Da tale rappresentazione emerge che la preoccupazione per la disoccupazione – che è la seconda tra le emergenze percepite dal campione intervistato – cresce con l'età (cioè la percezione della sua gravità diventa più elevata con gli anni<sup>22</sup>) e questo comportamento è omogeneo a prescindere dal genere; piuttosto, il livello della gravità percepita è sin dall'inizio superiore per i maschi (essendo essi, in gran maggioranza, i lavoratori nelle famiglie mono-reddito).

<sup>21</sup> Per l'uso della *deviance* e di  $\Delta dev$  nella scelta del modello si veda Dobson (1990).

<sup>22</sup> Si ricordi che nella nostra indagine valori bassi della graduatoria indicano una percezione di gravità elevata.



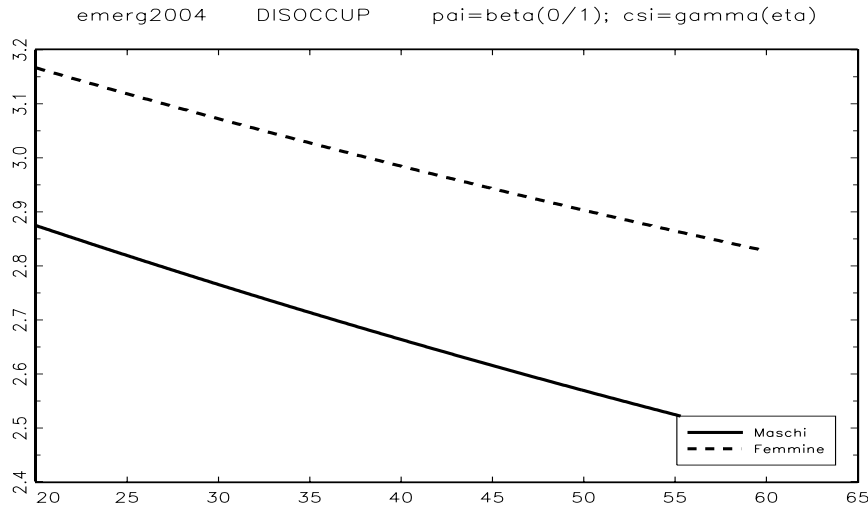


Figura 8. Valore atteso dei ranghi in funzione dell'età, per genere.

Si noti come il genere influenzi in modo netto l'incertezza nel senso che questa appare più elevata per le donne. Infatti, la corrispondente stima di  $\beta_1$  ha segno negativo: ciò comporta una diminuzione di  $\hat{\pi}$  per il genere femminile e, quindi, un aumento dell'incertezza. Complessivamente, quindi, si è in presenza di una maggiore consapevolezza di gravità nei riguardi del problema “Disoccupazione” da parte degli uomini rispetto alle donne.

- **Mala-sanità**

Esaminiamo, ora, come l'emergenza “Mala-sanità” sia percepita dal campione esaminato. Per semplificare la discussione ci limitiamo a presentare nella Tabella 10 solo il modello *MUB-11* nel quale le covariate *genere* (per spiegare l'incertezza nelle risposte) ed *età* (per spiegare la differente percezione di gravità) sono risultate ampiamente significative. Utilizzando la  $\Delta dev$  come misura di bontà del risultato rispetto al modello *MUB-00*, si ottiene:

$$\Delta dev = 2 * (-665.83656 - (-672.38235)) = 13.09158,$$

che risulta decisamente significativa (il suo  $p$ -value è 0.00144).

Tabella 10. Stime del modello MUB, in funzione del genere e dell'età.

<b>Modello stimato</b>	$\hat{\pi} = \pi(\text{genere})$	$\hat{\xi} = \xi(\text{età})$
MUB-11	$\hat{\beta}_0 = 1.414 (0.174)$ $\hat{\beta}_1 = 2.106 (0.527)$	$\hat{\gamma}_0 = 0.415 (0.118)$ $\hat{\gamma}_1 = -0.016 (0.004)$

La rappresentazione grafica del valore medio dei ranghi espressi (Figura 9), al variare dell'età e condizionato al genere dell'intervistato, mostra come la gravità del problema si riduca con l'età ed è sostanzialmente concorde tra i due generi. Forse, va segnalato che le donne in età molto giovane sembrano più preoccupate dei loro colleghi verso tale emergenza, mentre col passare dell'età le posizioni relative si invertono. Se si pone mente alle differenti problematiche medico-sanitarie che incontrano uomini e donne nel corso della loro vita, il modello che ne risulta appare ragionevole.

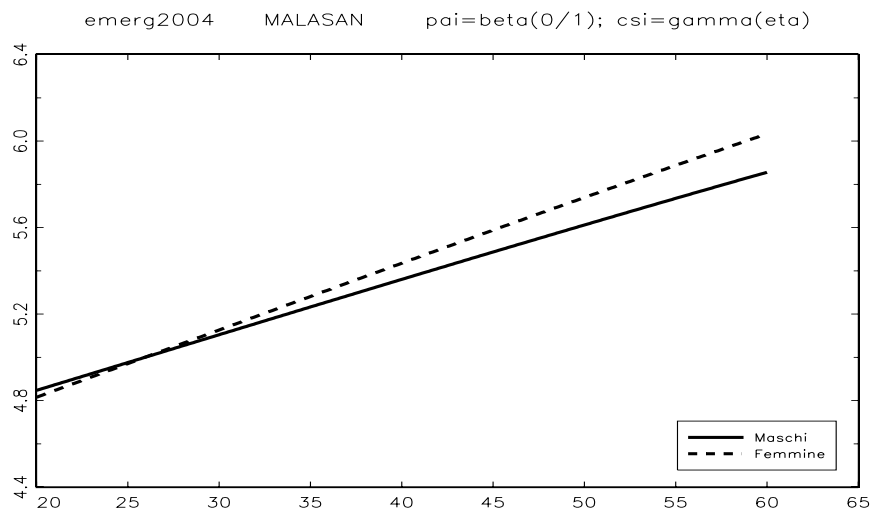


Figura 9. Valore atteso dei ranghi in funzione dell'età, per genere.

• **Immigrazione**

Le problematiche dell'immigrazione non ricevono una soglia di attenzione elevata da parte del nostro campione, ma i modelli stimati indicano maggiore incertezza e maggiore preoccupazione da parte delle donne; per entrambi i generi la percezione della gravità del problema diminuisce regolarmente con l'aumentare dell'età raggiungendo una soglia praticamente costante dopo i 35-40 anni.

Tabella 11. Stime del modello MUB, in funzione del genere e dell'età.

Modello stimato	$\hat{\pi} = \pi(\text{genere})$	$\hat{\xi} = \xi(\text{età})$
MUB-11	$\hat{\beta}_0 = 1.599 (0.185)$	$\hat{\gamma}_0 = -1.468 (0.313)$
	$\hat{\beta}_1 = -0.801 (0.258)$	$\hat{\gamma}_1 = -0.055 (0.013)$

La Tabella 11 mostra le stime dei parametri del modello MUB-11, mentre la Figura 10 evidenzia il valore medio dei ranghi al variare dell'età e condizionatamente al genere.

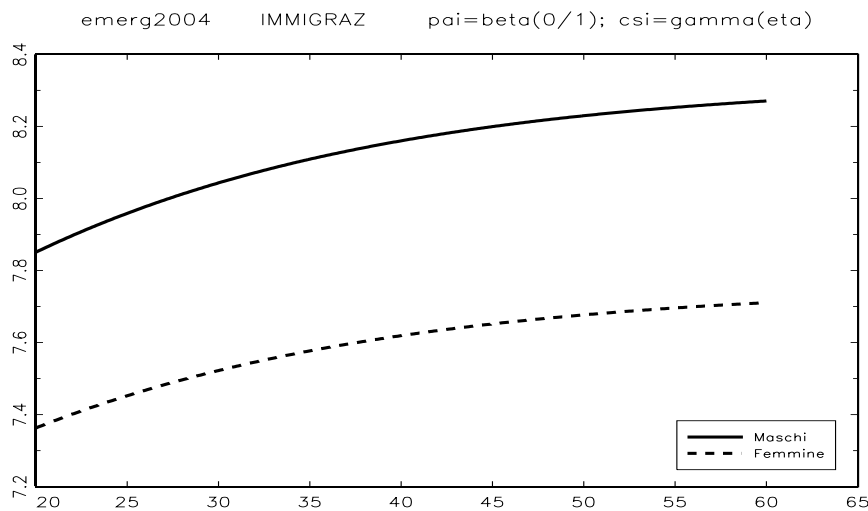


Figura 10. Valore atteso dei ranghi in funzione dell'età, per genere.

Va segnalato la notevole significatività di tutti i parametri stimati (ampiamente al di sotto dello 0.1%), come è confermato dal valore della

differenza tra le *deviance*, rispetto al modello senza covariate, che vale:

$$\Delta dev = 2 * (-506.60625 - (-518.74719)) = 24.28188,$$

e risulta molto significativa (il suo *p-value* è  $0.534 * 10^{-5}$ ).

### ***11. Età ed incertezza nella percezione delle emergenze***

Studiamo, ora, in breve, come si possa collegare l'incertezza alla età del rispondente mediante la misura  $\mathcal{A}(y)$  introdotta nel paragrafo 8. A tal fine, si tratta di stimare dei modelli *MUB-10* ove il parametro  $\pi$  è espresso in funzione della covariata *età* ( $= y$ ), calcolando poi la modifica sulla misura di incertezza  $\mathcal{A}(y)$ , al variare di  $y$ , e quindi della probabilità  $P_r(R = r \mid \pi = \pi(y))$ ,  $r = 1, 2, \dots, m$ .

Limitandoci ai soli modelli nei quali l'impatto è risultato nettamente significativo<sup>23</sup>, mostriamo nella Figura 11 la funzione  $\mathcal{A}(y)$  quando l'età  $y$  del rispondente varia tra 18 e 60 anni.

I grafici indicano che l'incertezza aumenta (sino a raddoppiare, in termini relativi) passando dai giovanissimi ai più anziani per le emergenze "Criminalità organizzata", "Mala-sanità" e "Inquinamento"; invece, decresce sino ad un terzo per quanto concerne la "Presenza di immigrati".

Poiché tale comportamento è uniforme anche per le variabili qui non discusse, si può concludere che i più giovani hanno manifestato generalmente una minore indecisione nelle risposte (per cui al loro interno l'atteggiamento e la consapevolezza sono più omogenee), mentre i più anziani hanno espresso risposte tra loro più eterogenee. Fanno eccezione (perché il comportamento dell'incertezza rispetto all'età si capovolge) le emergenze "Disoccupazione" e "Presenza di immigrati".

---

<sup>23</sup> Per sottolineare l'efficacia della interpretazione, i risultati vengono qui presentati per le sole quattro emergenze nei cui modelli la significatività del parametro stimato  $\hat{\beta}_1$  è risultata inferiore al livello 0.001. In effetti, per tutte le variabili si riscontra un'incertezza decisionale che varia con l'età del rispondente, eccetto per "Clientelismo e corruzione", "Pulizia e igiene territorio" e "Trasporti-traffico-mobilità".

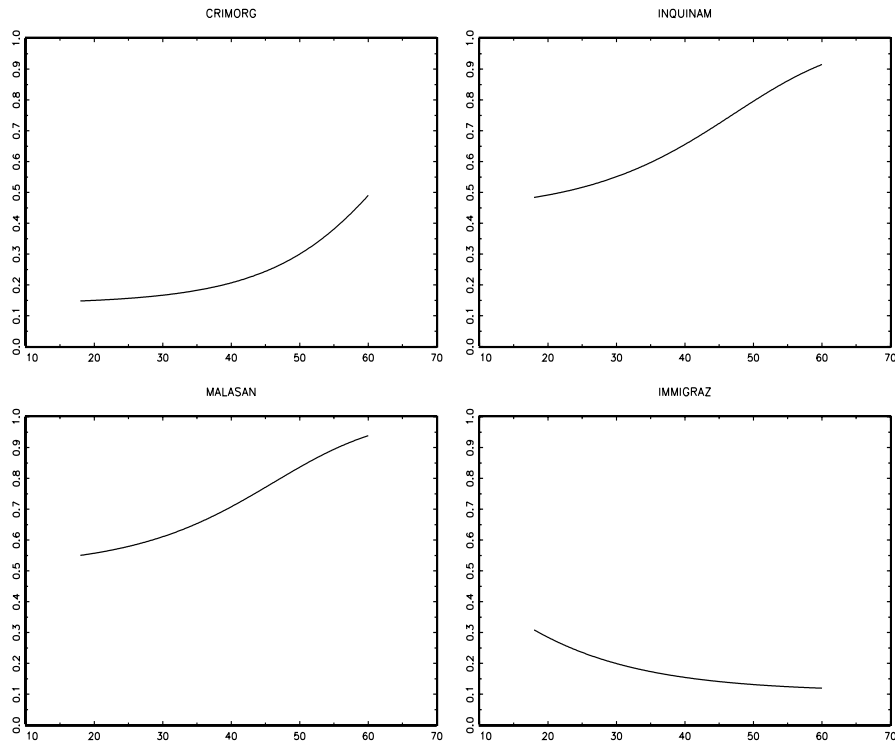


Figura 11. Incertezza  $A(y)$  in funzione dell'età.

## 12. Considerazioni finali

Questo lavoro ha esaminato un data set specifico la cui particolare composizione ha permesso di evidenziare, tra l'altro, il ruolo della variabile età nello spiegare alcune delle risposte. Trattasi di risposte che riflettono una notevole accuratezza dei rispondenti nel formulare le graduatorie, il che ha consentito di verificare su dati reali l'utilità di analisi statistiche e di modelli interpretativi per variabili ordinali. Per questo, possiamo riflettere su due aspetti complementari che emergono da questo studio e che possono essere generalizzati ad altri contesti.

Sul piano dell'interpretazione dei fenomeni analizzati, è emerso che la percezione della gravità di un problema viene alterata dalle dimensioni

spaziale e temporale entro le quali l'intervista si svolge. Questo risultato appare scontato in un contesto pesantemente dominato dai mass media ma la introduzione di modelli statistici ne consentono una quantificazione ed una comparazione degli effetti in entrambe le dimensioni. Infatti, ripetendo l'intervista in zone diverse e in anni successivi, si potrebbero studiare modifica ed evoluzione delle percezioni.

In effetti, i modelli *MUB* sin qui discussi, oltre ad esplicitare il ruolo delle covariate del soggetto, potrebbero anche includere variabili connesse all'item considerato. In tal caso, queste ultime possono essere studiate per misurare l'effettivo impatto dei messaggi mediatici su tali percezioni. Per esempio, si può studiare la modifica delle risposte al variare dello spazio (temporale o fisico) che la televisione ed i giornali concedono ad una specifica emergenza nella settimana precedente l'intervista, e così via.

Come avviene per tutte le ipotesi del mondo reale, il modello è una valenza aggiuntiva che permette una valutazione statistica dell'evidenza empirica e conferisce all'incertezza decisionale, sempre presente, un connotato probabilistico.

Sul piano metodologico, questo lavoro –seppur circoscritto alla disamina di uno specifico data set– ha mostrato che analisi esplorative condotte mediante opportuni indicatori possono essere utili per una collocazione preliminare dei parametri di un modello probabilistico. Tale aspetto migliora l'interpretazione delle stime che si ottengono dal metodo della massima verosimiglianza, ma consente altresì di pervenire a valori iniziali più idonei da inserire nelle procedure numeriche di convergenza. Inoltre, si è anche mostrato come sia possibile misurare la modifica dell'incertezza decisionale tramite lo studio di un opportuno indice di eterogeneità che espliciti l'effetto di una covariata.

Infine, il lavoro ha anche discusso il ruolo dell'indice di dissomiglianza quale alternativa efficace rispetto ai più tradizionali test di adattamento, allo scopo di evitare un rifiuto troppo frequente dell'ipotesi nulla, anche nei casi in cui appare evidente che il modello è coerente con la distribuzione empirica.

*Ringraziamenti:* Questo lavoro è stato svolto nell'ambito dei progetti di ricerca del Dipartimento di Scienze Statistiche, Università di Napoli Federico II, grazie anche alle strutture di ricerca del CFEPSR, Portici.

Si ringraziano l'editor ed i referees per i suggerimenti critici che hanno consentito di migliorare il lavoro. Si segnala altresì il contributo preliminare svolto dalla dott.ssa Nunzia Fagnani nel predisporre la sua tesi di laurea sulla problematica delle emergenze metropolitane.

### **Riferimenti bibliografici**

Agresti A. (2002), *Categorical data analysis*, Wiley-Interscience, Hoboken, NJ.

Agresti A., Agresti B.F. (1978), Statistical analysis of qualitative variation, *Sociological Methodology*, 9, 204-237.

Allison P. D. (1981), Inequality measure for nominal data, *American Sociological Review*, 46, 371-373.

Coulter P. B. (1984), Distinguishing inequality and concentration: the exponentiation principle, *Political Methodology*, 3, 323-355.

D'Elia A. (1999), A proposal for ranks statistical modelling, in: *Statistical modelling - Proceedings of the 14th International Workshop on Statistical Modelling*, (Friedl, H., Berghold, A., Kauermann, G. eds.), Graz - Austria, 468 - 471.

D'Elia A. (2003a), Modelling ranks using the Inverse Hypergeometric distribution, *Statistical Modelling: an International Journal*, 3, 65-78.

D'Elia A. (2003b), A mixture model with covariates for ranks data: some inferential developments, *Quaderni di Statistica*, 5, 1-25.

D'Elia, A. (2004), New developments in ranks data modelling with covariates, *Atti della XLII Riunione Scientifica SIS*, CLEUP, Padova, 233-244.

D'Elia, A., Piccolo, D. (2002), Analisi statistica delle preferenze: metodi e modelli a confronto, *Studi in onore di Angelo Zanella*, Vita e Pensiero, Milano, 167-187.

D'Elia A., Piccolo D. (2005), A mixture model for preference data analysis, *Computational Statistics & Data Analysis*, 49, 917-934.

Dobson A. J. (1990), *An introduction to generalized linear models*, Chapman & Hall, London.

Frosini B. V. (1981), Heterogeneity indices and distances between distributions, *Metron*, XXXIX, 3-4.

- Gini C. (1912), *Variabilità e mutabilità. Contributo allo studio delle distribuzioni e relazioni statistiche*, Studi Economici-Giuridici della Reale Università di Cagliari (ristampa Cuppini, Bologna).
- Girone G., Salvemini T. (1986), *Lezioni di Statistica*, vol.II, Cacucci editore, Bari.
- Greenberg J. H. (1956), The measurement of linguistic diversity, *Language*, 32, 109-115.
- Grilli L., Rampichini C. (2002), Scomposizione della dispersione per variabili statistiche ordinali, *Statistica*, LXII, 111-116.
- Guseo R. (1997), *Istituzioni di Statistica. Lezioni*, CEDAM, Milano.
- Haberman S. J. (1982), Analysis of dispersion of multinomial responses, *Journal of the American Statistical Association*, 77, 568-579.
- Hill M. O. (1973), Diversity and evenness: a unifying notation and its consequences, *Ecology*, 54, 427-437.
- Johnson V. E., Albert J. H. (1999), *Ordinal data modeling*, Springer, New York.
- Kempton R. A. (1979), The structure of species abundance and measurement of diversity, *Biometrics*, 35, 307-321.
- Kokko H., Mackenzie A., Reynolds, J. D., Lindström J. and Sutherland, W. J. (1999), Measure of inequality are not equal, *The American Naturalist*, 72, 358-382.
- Laakso M., Taagepera R. (1979), Effective number of Parties: a measure with application to West Europe, *Comparative Political Studies*, 12, 3-27.
- Leti G. (1979), *Distanze e indici statistici*, La Goliardica editrice, Roma.
- Leti G. (1983), *Statistica descrittiva*, Edizioni il Mulino, Bologna.
- Liebertson S. (1969), Measuring population diversity, *American Sociological Review*, 34, 850-862.
- Light R. J., Margolin B.H. (1971), An analysis of variance for categorical data, *Journal of the American Statistical Association*, 66, 534-544.
- McCullagh P., Nelder J.A. (1998), *Generalized linear models*, 2<sup>nd</sup> edition, Chapman & Hall, London.
- McLachlan G., Krishnan G. J. (1997), *The EM algorithm and extensions*, J. Wiley & Sons, New York.
- McLachlan G., Peel G. J. (2000), *Finite mixture models*, J. Wiley & Sons, New York.
- Patil G. P., Taillie C. (1982) Diversity as a concept and its measurement (with discussion), *Journal of the American Statistical Association*, 77, 548-567.
- Peet R.K. (1974), The measurement of species diversity, *Annual Review of Ecology and Systematics*, 5, 285-307.



Picarreta R. (2001), A new measure of nominal-ordinal association, *Journal of Applied Statistics*, 28, 107-120.

Piccolo D. (1998), *Statistica*, 2<sup>a</sup> edizione (2000), Edizioni il Mulino, Bologna.

Piccolo D. (2003), Computational issues in the E-M algorithm for ranks model estimation with covariates, *Quaderni di Statistica*, 5, 27-48. Corrections in: 6, 199.

Rampichini C., Grilli L., A. Petrucci (2004), Analysis of university course evaluations: from descriptive measures to multilevel models, *Statistical methods & applications*, 13, 357-373.

Scagni A. (2005), Le opinioni degli studenti sulla didattica universitaria: sono attendibili le differenze nei risultati aggregati di Facoltà?, in: C. Crocetta (a cura di), *Modelli statistici per l'analisi della transizione Università-lavoro*, CLEUP, Padova, 23-38.

Simonoff J. S. (2003), *Analyzing categorical data*, Springer, New York.