

## **Indicatori locali di povertà delle famiglie**

# **LA STIMA DI PARAMETRI DI POVERTÀ E DISUGUAGLIANZA RIFERITI A DIVERSE TIPOLOGIE FAMILIARI PER REGIONE**

**Maria Rosaria Ferrante**

*Dipartimento di Scienze Statistiche - Università degli Studi di Bologna*

*18 febbraio 2013*

---

**Obiettivo:** stimare parametri di povertà e disuguaglianza riferiti a diverse tipologie familiari in ciascuna regione sulla base dei dati EU-SILC al fine di

- valutare la situazione della povertà in Emilia-Romagna in termini relativi rispetto al quadro nazionale ed a quello delle regioni limitrofe
- identificare i segmenti della società maggiormente esposti al rischio di povertà ed esclusione sociale
- definire meglio il target per le politiche redistributive, cioè evidenziare quali sono le aree sociodemografiche di maggior bisogno di intervento e sostegno da parte dello Stato

**Problema statistico:** insufficiente numerosità campionaria in classi (domini) definiti da tipologie familiari e regioni

I parametri di interesse sono stati stimati al seguente **dettaglio:**

- 1) per l'intera Italia;
- 2) per ciascuna Regione (20 regioni);
- 3) per ciascuna tipologia familiare (9 gruppi);
- 4) per ciascuna tipologia familiare all'interno delle regioni (180 domini).

*Tipologie familiari considerate (Eurostat)*

Tipologia	Descrizione
A	Monocomponenti
B	2 adulti, senza figli dipendenti, entrambi età inferiore a 65 anni
C	2 adulti, senza figli dipendenti, almeno uno età superiore a 65 anni
D	Altra tipologia, senza figli dipendenti
E	Monogenitore, con uno o più figli dipendenti
F	2 adulti, 1 figlio dipendente
G	2 adulti, 2 figli dipendenti
H	2 adulti, 3 o più figli dipendenti
I	Altra tipologia con figli dipendenti

*figlio a carico (o dipendente)*: minore di età inferiore ai 16 anni o figlio di età compresa tra i 16 e i 26 anni che sia inattivo e dunque *dipendente* in termini economici dalla famiglia di appartenenza

Parametri di interesse, scelti tra quelli previsti dall'UE per valutare la povertà e la disuguaglianza reddituale:

**sintesi della distribuzione dei redditi**

- il reddito disponibile equivalente mediano

**incidenza di povertà**

- i tassi di povertà basati sulle soglie di povertà nazionali (SPN)
- i tassi di povertà basati sulle soglie di povertà regionali (SPR)

**disuguaglianza nella distribuzione dei redditi**

- l'indice di concentrazione di Gini

**intensità di povertà**

- il *poverty gap* mediano basato su SPN
- il *poverty gap* mediano basato su SPR

Le **stime pubblicate da ISTAT** per l'indagine EU-SILC, sono basate su stimatori “da disegno”.

- ottenuti impiegando le informazioni sul disegno campionario probabilistico (la fonte di casualità è rappresentata dalla selezione casuale del campione dalla popolazione), cioè i coefficienti di riporto
- in gergo questi stimatori vengono detti “diretti” e dirette le stime ottenute sulla loro base

I 180 domini ottenuti dall'incrocio regione x tipologia si configurano, dal punto di vista della stima come **“piccole aree”**, ossia come **domini per cui i dati campionari – e quindi gli stimatori “diretti” non sono sufficienti a produrre stime di sufficiente affidabilità statistica**: sono richieste metodologie di analisi complesse, cosiddette “da modello”.

La **strategia generale di stima che proponiamo** per i parametri di interesse:

1. produrre le stime dirette dei parametri per tutti i domini di interesse;
2. valutare l'affidabilità statistica degli stimatori diretti;
3. sostituire le stime dirette con stime ottenute con metodologie più complesse (“da modello”) per i parametri le cui stime dirette non risultano sufficientemente affidabili.

⇒ **ognuno dei passi sopra delineati presenta problemi statistici peculiari**

Campione IT-SILC 2009. *Famiglie* rispondenti per regione e tipologia familiare

Regione	Tipologia									Totale
	A	B	C	D	E	F	G	H	I	
Piemonte	445	166	224	192	56	166	136	19	32	1436
Valle D'Aosta	124	44	63	36	21	35	44	9	8	384
Lombardia	585	264	374	277	72	252	214	46	106	2190
Trentino A.A.	238	80	101	99	34	71	96	39	52	810
Veneto	370	210	251	228	44	181	186	26	85	1581
Friuli V.G.	273	100	156	107	34	75	100	17	46	908
Liguria	358	114	161	112	31	96	72	7	20	971
<b>Emilia-Romagna</b>	427	214	263	168	52	156	126	26	65	1497
Toscana	382	164	232	208	46	193	130	18	75	1448
Umbria	212	93	159	134	25	94	81	14	71	883
Marche	238	93	179	138	28	128	85	24	65	978
Lazio	444	167	218	179	57	162	181	38	92	1538
Abruzzo	121	40	66	70	9	46	63	10	20	445
Molise	97	32	50	75	10	30	43	15	23	375
Campania	240	103	161	202	35	138	194	67	151	1291
Puglia	215	88	165	133	36	95	141	38	93	1004
Basilicata	109	45	68	57	16	53	57	14	41	460
Calabria	163	47	80	90	17	64	86	25	48	620
Sicilia	283	96	168	169	41	106	135	57	90	1145
Sardegna	101	62	60	100	12	69	64	17	43	528
Totale	5425	2222	3199	2774	676	2210	2234	526	1226	20492

## PRODUZIONE DELLE STIME DIRETTE

Abbiamo già detto che l'Istat impiega stimatori “da disegno”: si basano sui coefficienti di riporto all'universo, producono “stime dirette”.

Al fine di migliorare le proprietà degli stimatori “diretti” l'Istat adotta un processo di calibrazione: intuitivamente la calibrazione consiste nel fare in modo che le stime del totale per variabili il cui totale a livello di popolazione è noto riproducano esattamente questo valore.

### **Problema statistico:**

non è opportuno usare i coeff. di riporto pubblicati da ISTAT per ottenere stime riferite a domini quali tipologia e tipologia nelle regioni: tali coefficienti “riportano” ai domini cosiddetti “pianificati” (Italia, regioni), se li usassimo per ottenere stime relative a domini non pianificati otterremmo a stime distorte

**Soluzione:** abbiamo modificato il processo di calibrazione dei pesi per adattarlo ai domini di interesse

## LA VALUTAZIONE DELL’AFFIDABILITÀ DELLE STIME DIRETTE

### **Problema statistico:**

La valutazione dell’affidabilità degli stimatori (stima delle loro varianze), non può essere effettuata con i metodi usuali: il disegno campionario è complesso (a due stadi) e i parametri oggetto di stima sono funzioni non lineari dei dati.

### **Soluzione adottata:**

Abbiamo implementato un algoritmo *bootstrap* ha hoc di stima della varianza degli stimatori diretti

### **Risultati:**

Le stime dei parametri descrittivi della povertà e della disuguaglianza distributiva del reddito

- a livello regionale
- per tipologia familiare a livello nazionale

sono in generale caratterizzate da un livello di affidabilità sufficiente

## Coefficienti di variazione

Per ciascun parametro, valori minimo, medio e massimo del *Coefficiente di Variazione*, CV, delle stime dirette per i 9 domini dell'ER definiti da *regione e tipologia familiare*:

	Reddito medio equivalente	Reddito mediano equivalente	Tasso povertà severa, SN	Tasso povertà, SN	Tasso di rischio di povertà, SN	Tasso povertà severa, SR	Tasso povertà, SR	Tasso di rischio di povertà, SR	Gap mediano, SN	Gap mediano, SR	Indice di Gini
min cv	3.14	2.47	19.40	12.71	10.24	12.86	10.31	7.38	14.96	10.75	4.65
avg cv	5.76	6.30	58.50	34.11	26.94	42.17	26.44	20.91	41.91	40.39	8.47
max cv	13.23	20.09	151.14	54.72	45.94	78.78	45.94	35.14	56.54	94.35	13.69

Per reddito pro-capite, reddito pro-capite mediano, indice di concentrazione di Gini la stima è sufficientemente affidabile anche con riferimento ai domini

Per tutti gli altri parametri, i CV medi sono molto più elevati e le stime dirette non possono essere considerate basi accettabili per l'analisi dei dati: è necessario far ricorso a stime “da modello” per domini di studio (o, in gergo, per “piccole aree”).

## IN COSA CONSISTE (A GRANDI LINEE) LA METODOLOGIA DI STIMA “PER PICCOLE AREE”

### Alcuni aspetti generali

- Piccola area (o dominio): sub–popolazione per la quale sono presenti poche osservazioni campionarie
- Non è in genere possibile ottenere stime affidabili di caratteristiche della sub–popolazione della piccola area con le sole informazioni campionarie
- L’affidabilità delle stime può essere rafforzata mediante una strategia “da modello”: si specifica un modello probabilistico descrittivo della realtà che consente di “trarre forza” (borrow strength) da informazione ausiliaria (censimenti, fonti amministrative, ecc.)

### I modelli per “piccole aree”

- due grandi famiglie: “*area level*” e “*unit level*”: in questa ricerca ci siamo basati su modelli “di area”
- il modello ‘di area’ consiste di due parti, un cosiddetto ‘*sampling model*’ in cui si definiscono le ipotesi sugli stimatori diretti e le loro relazioni con i parametri di area, e un ‘*linking model*’ che mette in relazione questi parametri mediante informazioni ausiliarie specifiche di area
- univariati e multivariati

## La stima dei tre tassi

- tre diversi tassi di povertà per i domini (quota di persone molto povere, di poveri e di coloro a rischio di povertà) sono in ordine crescente in quanto ottenuti per soglie di povertà crescenti
- poiché i tre tassi sono naturalmente correlati è opportuno prendere in considerazione la possibilità di impiegare modelli multivariati
- è necessario che gli stimatori del tasso di povertà soddisfino le seguenti proprietà di coerenza:
  - in una data area, i tassi associati a soglie sempre più elevate devono essere crescenti,
  - gli stimatori intervallari dovrebbero avere i limiti inferiore e superiore nell'intervallo  $(0,1)$ ,
  - quando si dispone della stima per domini di grandi dimensioni, la stima di dominio dovrebbe essere ragionevolmente vicino alla stima diretta
  - la stima del tasso dovrebbe essere prodotta per i domini per i quali non si dispone di unità campionate o quando non ci sono famiglie povere campionate in una data area

Alcune di queste proprietà non sono soddisfatte dagli usuali modelli di piccola area basati sull'ipotesi di normalità: proponiamo un modello in cui le ***differenze tra i tassi di povertà successivi hanno condizionalmente una distribuzione Beta ed un modello Multivariate Logistic-Normal multivariato per i valori attesi delle distribuzioni Beta.***

## La stima del gap

per la stima del gap è stato impiegato un modello univariato Normale-Normale

## Informazione ausiliaria

Sono state considerate varie fonti (Contabilità Regionale, Rilevazione Continua sulle Forze di Lavoro e altri fonti amministrative) e la scelta è caduta sulla stima del Prodotto Interno Lordo pro-capite regionale ottenuta dalla Contabilità Regionale.

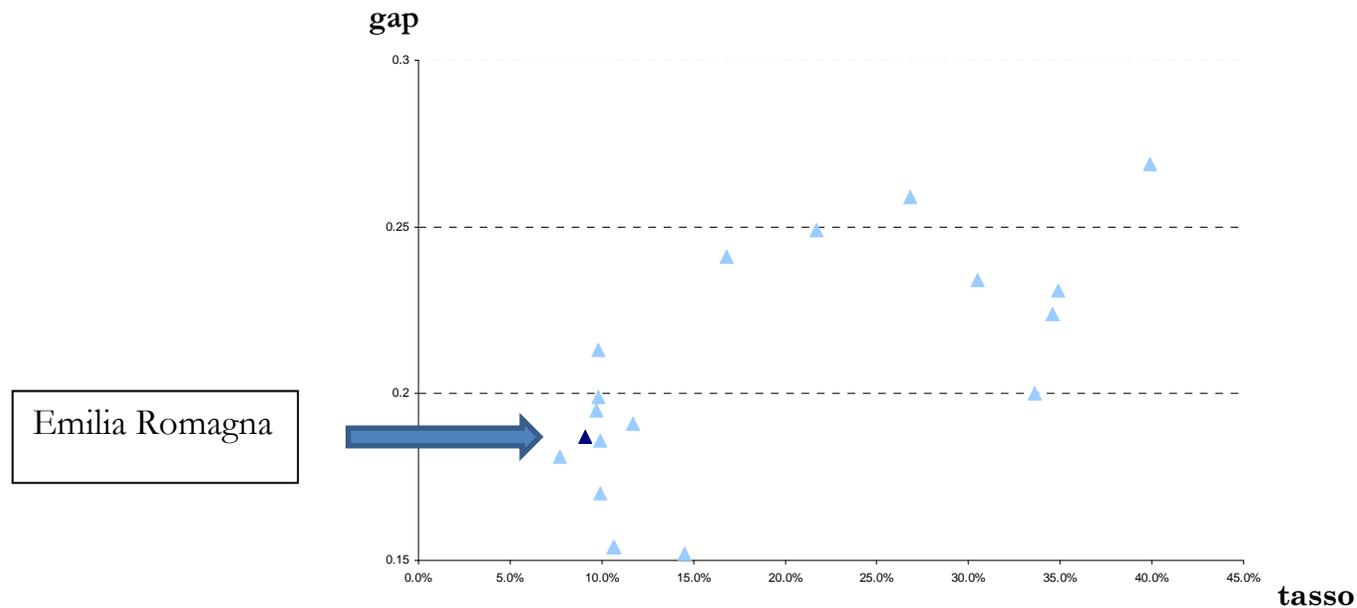
## Metodo di stima

L'impostazione seguita nella stima è quella Bayesiana: rende più agevole la specificazione di modelli che non facciano ricorso all'ipotesi di normalità come in questa ricerca

## La posizione dell'Emilia Romagna nel confronto tra stime regionali

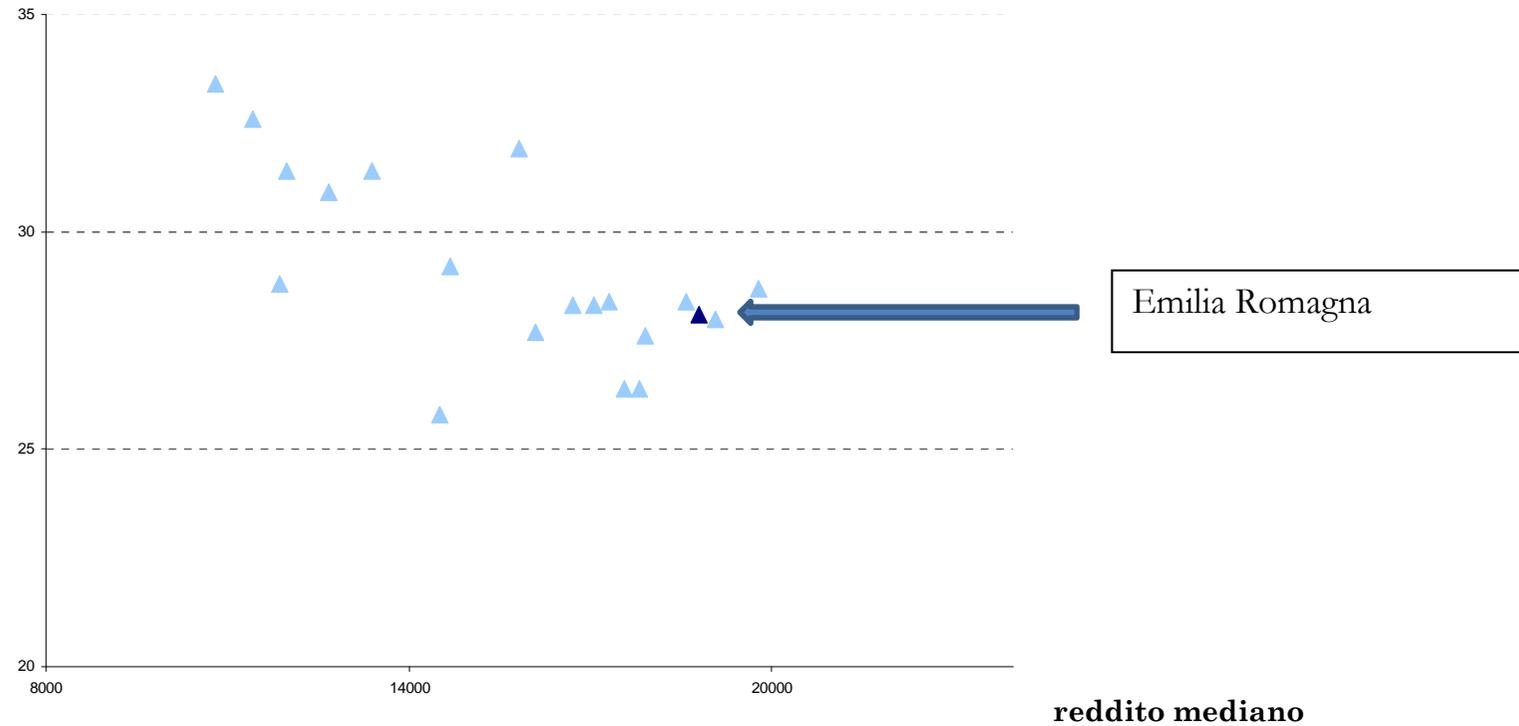
- il reddito mediano è tra i più elevati a livello nazionale
- il tasso di povertà calcolato con soglia nazionale è tra i più bassi
- in termini di gap occupa una posizione intermedia,
- in termini di disuguaglianza si colloca in posizione medio-bassa

*Tasso di povertà e poverty gap mediano (soglia di povertà 60% reddito mediano naz.)*



*Reddito disponibile equivalente mediano e indice di Gini. In colore più scuro l'Emilia-Romagna*

**indice di Gini**



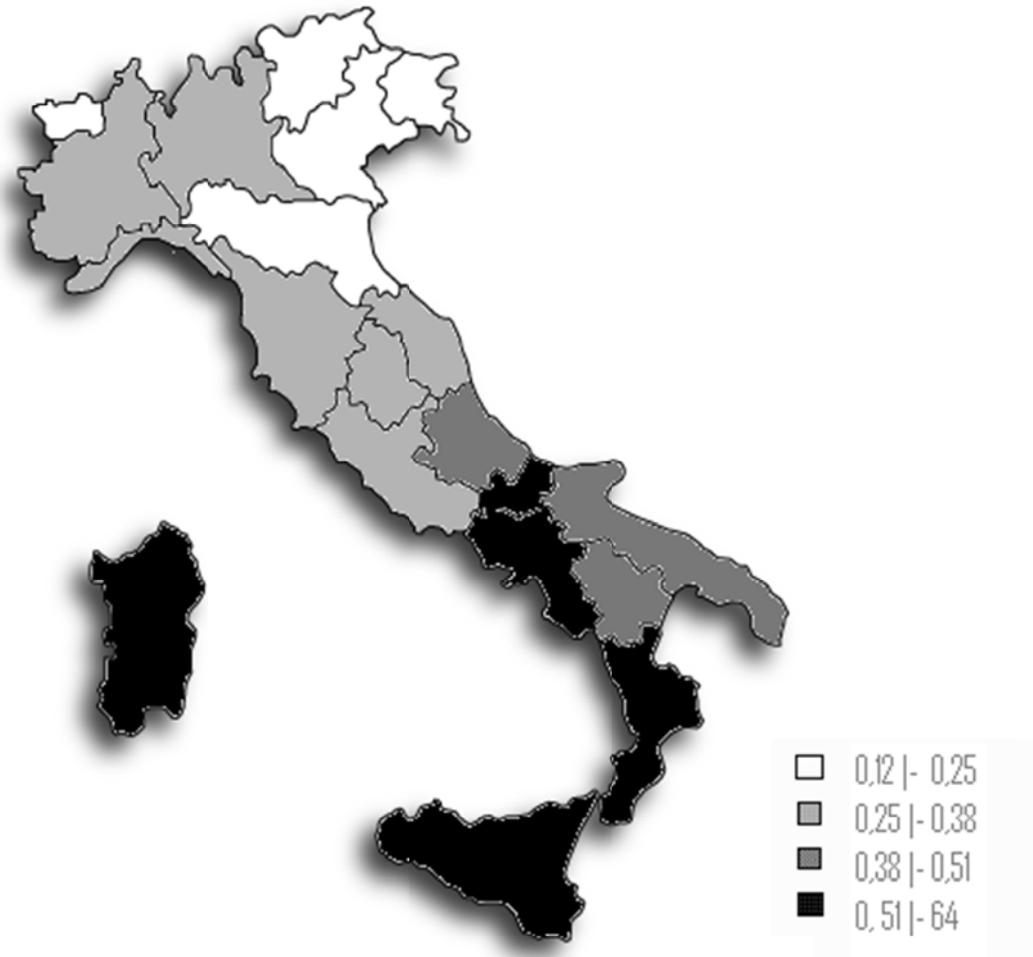
## Le stime dei parametri di povertà e disuguaglianza per regione e tipologia familiare

180 domini: non è possibile proporre una descrizione dettagliata di tutti i risultati

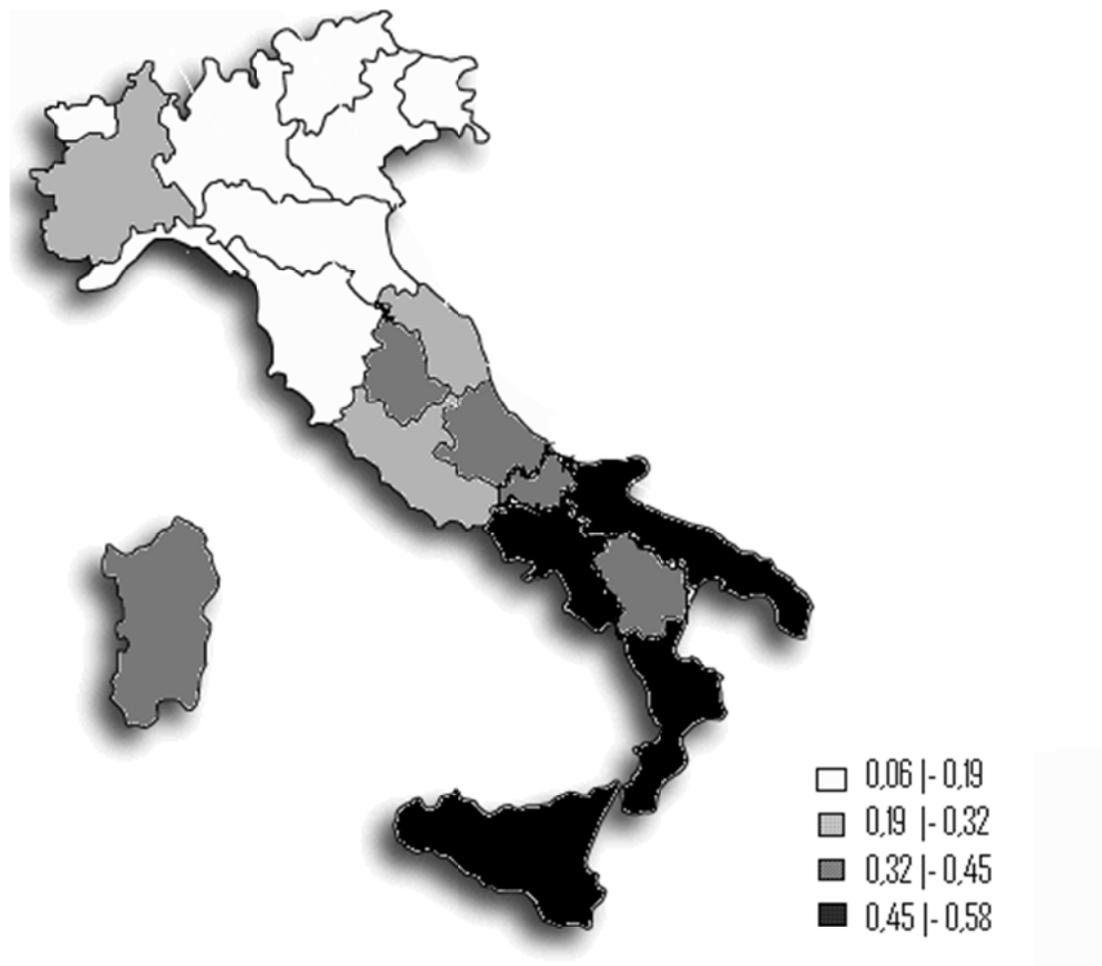
Alcune caratteristiche di fondo evidenziate dall'analisi dalle stime “marginali” per regione e per tipologia familiare permangono anche disaggregando le stime:

- l'ER è per tutte le tipologie fra le regioni con i tassi (SN) più bassi
- per tutti i tassi considerati, le tipologie a cui corrispondono i valori più elevati sono: “un adulto con bambini”, “monocomponenti” e “due adulti con tre o più bambini”.
- l'ordinamento delle regioni è simile per i tre tassi quando si guarda alle singole tipologie (il divario tra le regioni del Nord e le regioni del Sud, infatti, si riproduce nelle stime relative a tutte le tipologie)

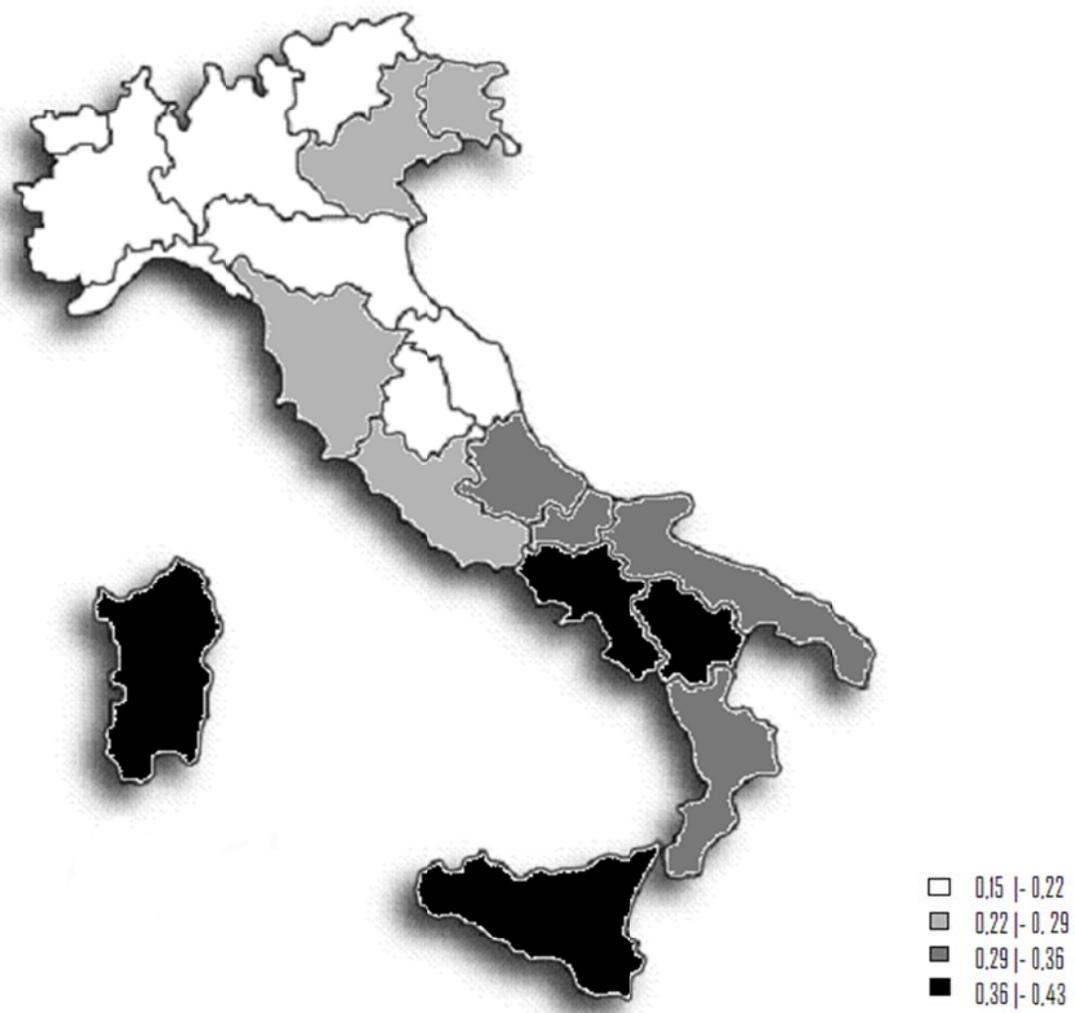
*Tasso di povertà (soglia 60% del reddito mediano nazionale). Famiglie monogenitore con uno o più figli dipendenti*



*Tasso di povertà (soglia 60% del reddito mediano nazionale). Famiglie composte da due adulti, tre o più figli dipendenti*



*Tasso di povertà (soglia 60% del reddito mediano nazionale). Famiglie monocomponenti*

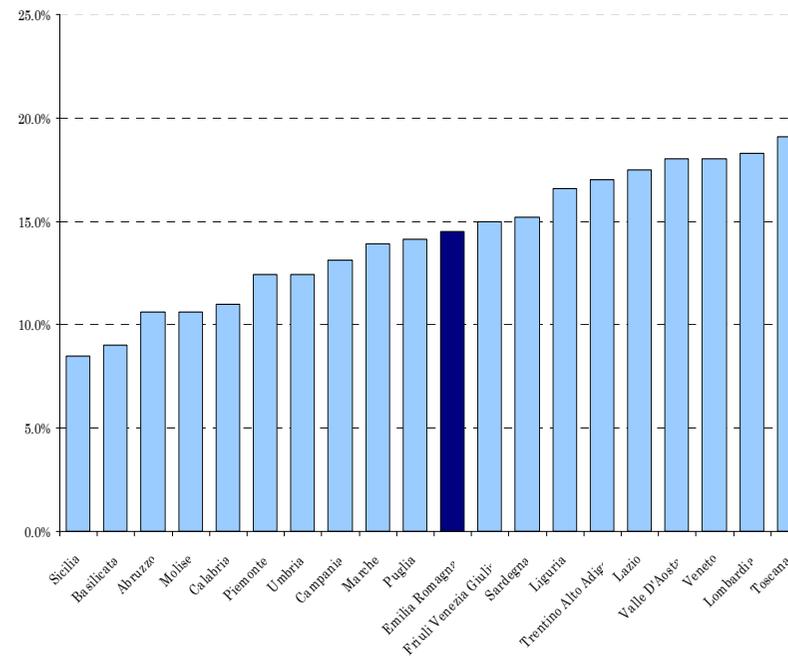
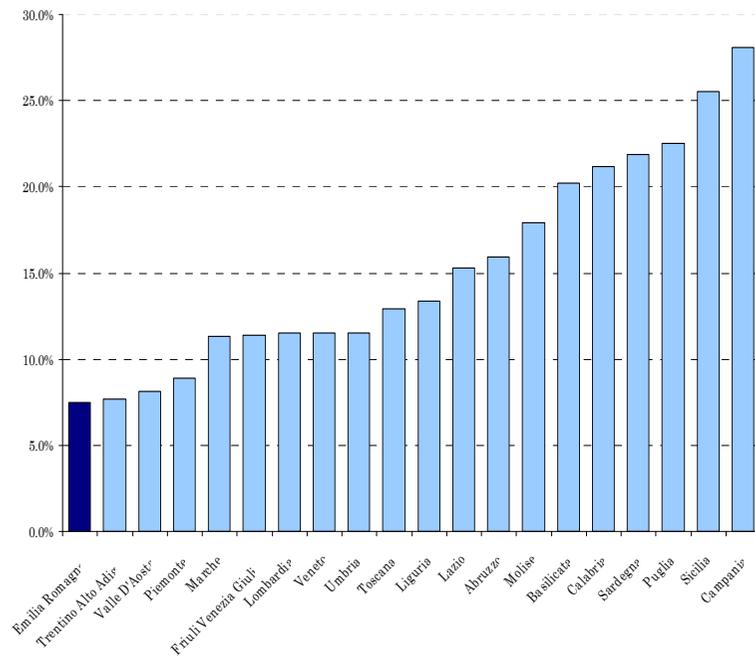


**Tassi di povertà severa basati su soglia regionale a confronto con quelli su soglia nazionale**

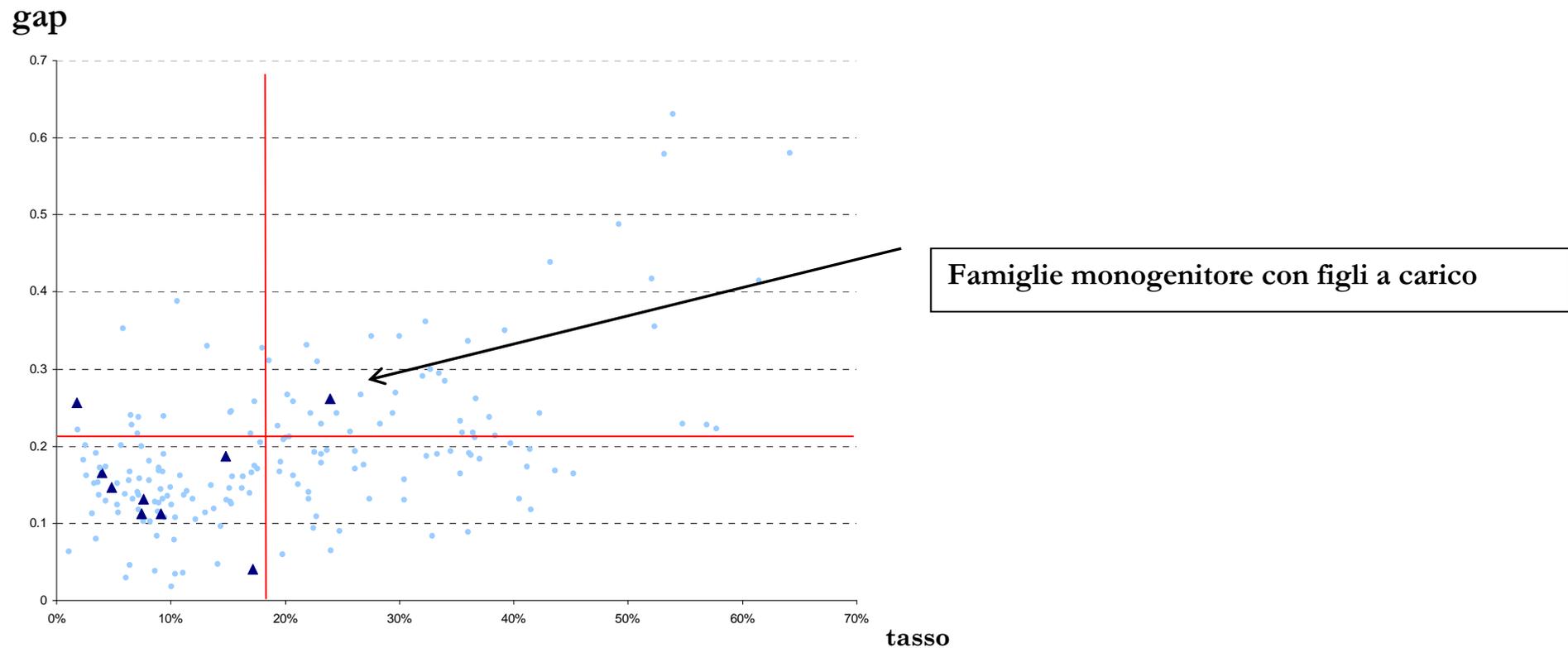
*Regioni italiane ordinate per tasso di povertà severa - Famiglie monocomponente (tipologia A).*

*Soglia nazionale*

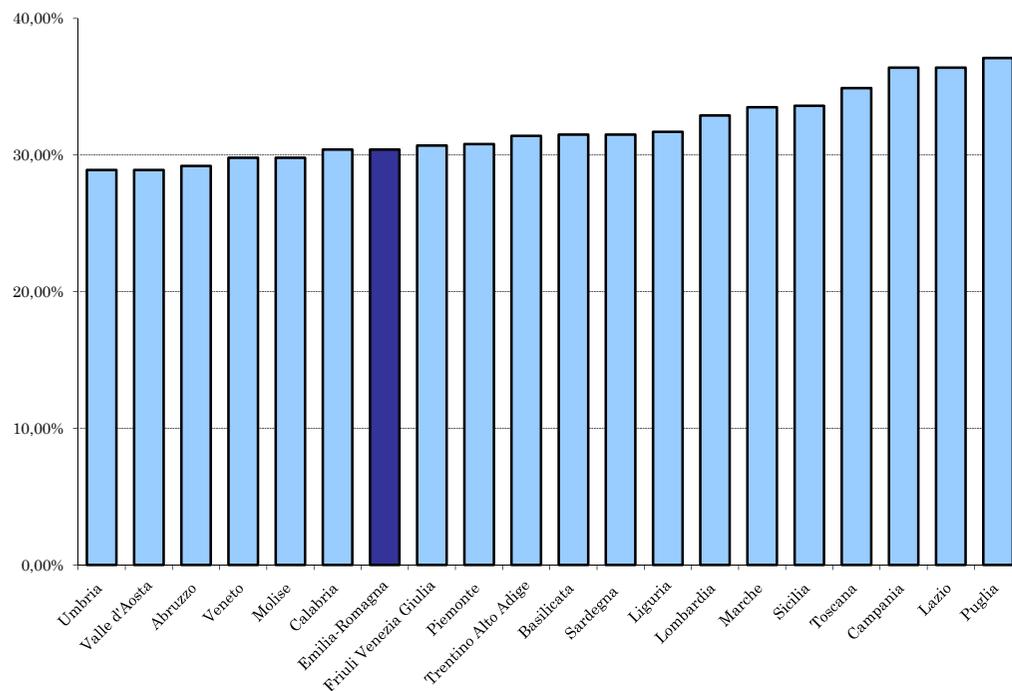
*Soglia regionale*



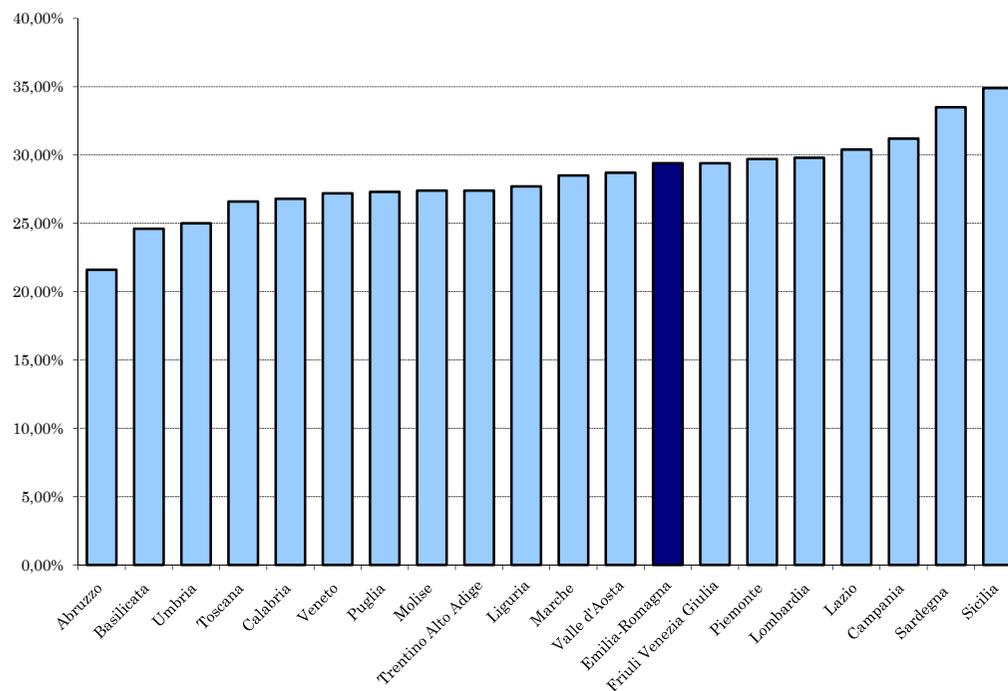
*Tasso di povertà e poverty gap mediano (soglia di povertà 60% reddito mediano nazionale). In colore più scuro i punti relativi alle tipologie familiari in Emilia-Romagna*



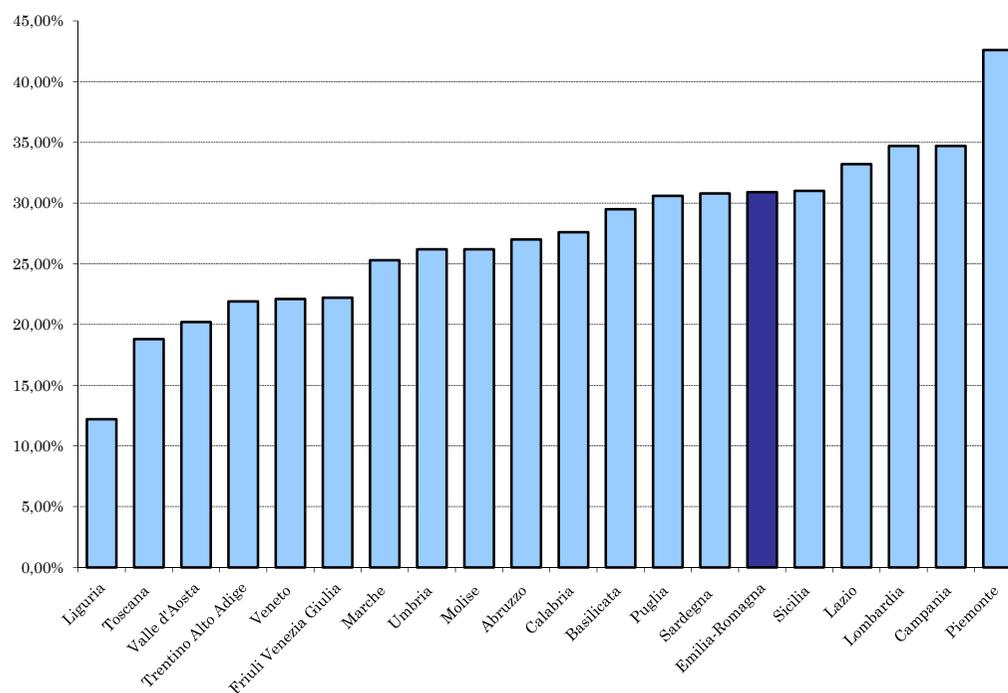
*Regioni ordinate per indice di concentrazione di Gini.  
Famiglie monocomponente (tipologia A)*



*Regioni ordinate per indice di concentrazione di Gini.  
Famiglie composte da 2 adulti, di cui uno di età superiore a 65 anni, senza figli (tip. C).*



*Regioni ordinate per indice di concentrazione di Gini.  
Famiglie composte da 2 adulti e 3 o più figli dipendenti (tipologia H)*

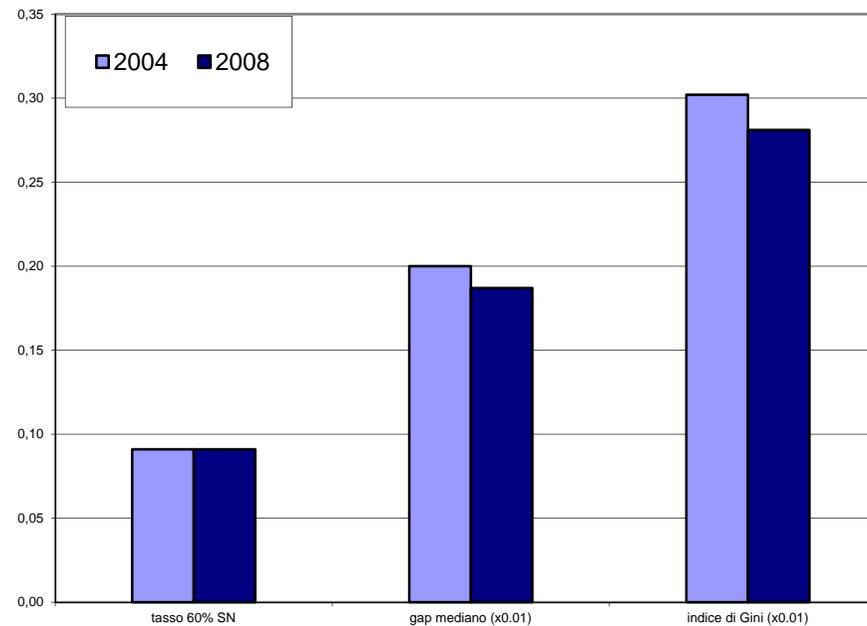


In sintesi, in Emilia-Romagna si nota che:

- le tipologie con **tassi di povertà** maggiori (soglia nazionale, 60% della mediana) sono: a) “un adulto con uno o più bambini”, 24% circa; b) “monocomponenti”, 15% circa; c) “due adulti con 3 o più figli”, 17% circa;
- le tipologie caratterizzate dal **gap mediano** maggiore sono: a) “un adulto con uno o più bambini”, 26% circa; b) la categoria residuale di “altro, senza figli”;
- le tipologie con i valori più elevati dell'**indice di Gini** sono: a) “un adulto con uno o più bambini”, 33% circa; b) “monocomponenti”, 30% circa; c) “due adulti con 3 o più figli”, 31% circa;
- le tipologia che presentano congiuntamente i valore più elevati **sia del tasso di povertà calcolato su soglia nazionale, sia di quello calcolato su soglia regionale** (60% della mediana), come emerge dal grafico 4.15 sono: a) “un adulto con uno o più bambini”, 37% circa; b) “monocomponenti”, 23% circa; c) “due adulti con 3 o più figli”, 24% circa;

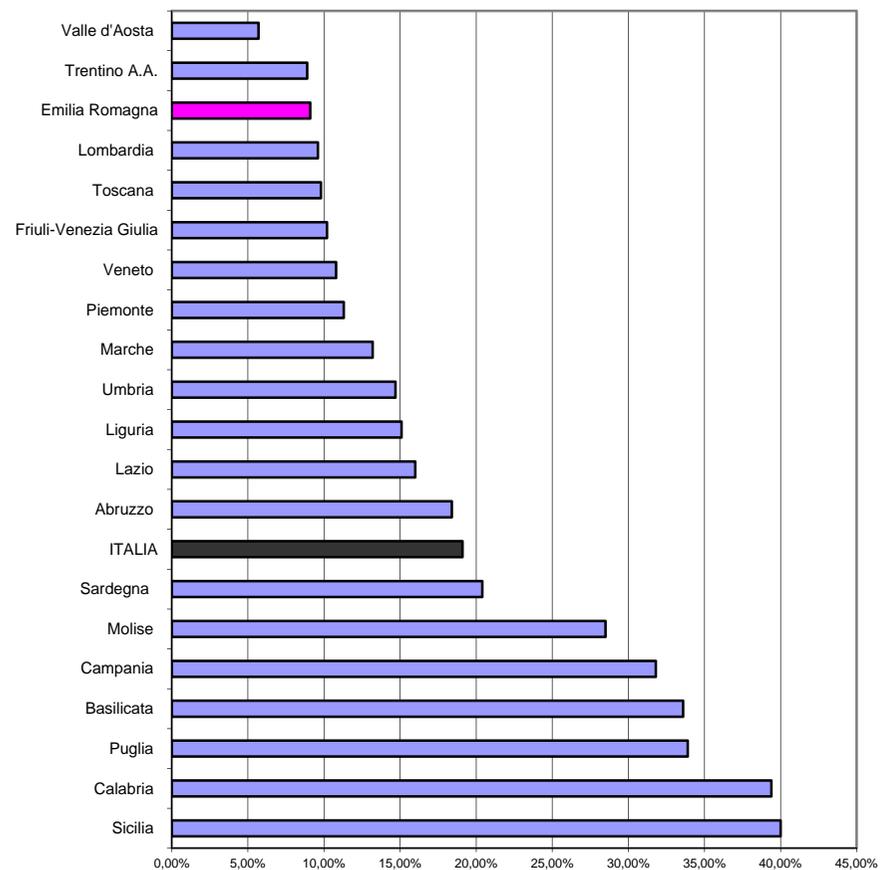
## Le variazioni intervenute dal 2004 al 2008 in ER

*Tasso di povertà (soglia 60% reddito disponibile equivalente mediano nazionale), gap mediano ed indice di Gini dell'Emilia Romagna, anni 2004 e 2008*

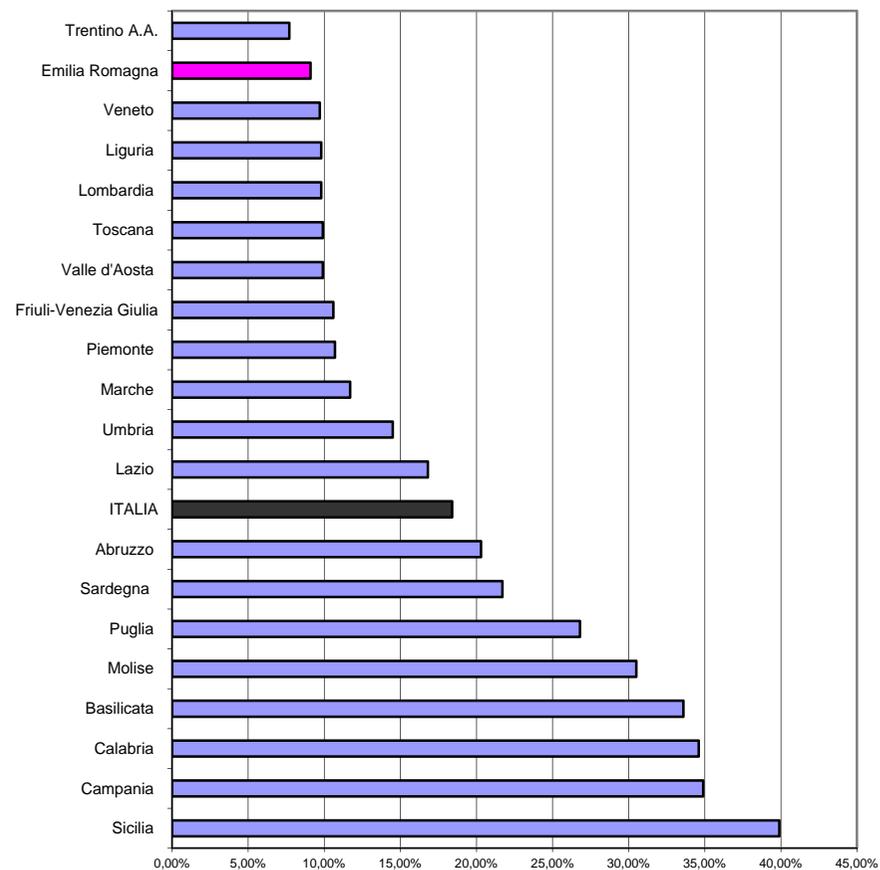


*Regioni italiane ordinate per tasso di povertà (soglia 60% reddito mediano nazionale), 2004 e 2008*

Anno 2004

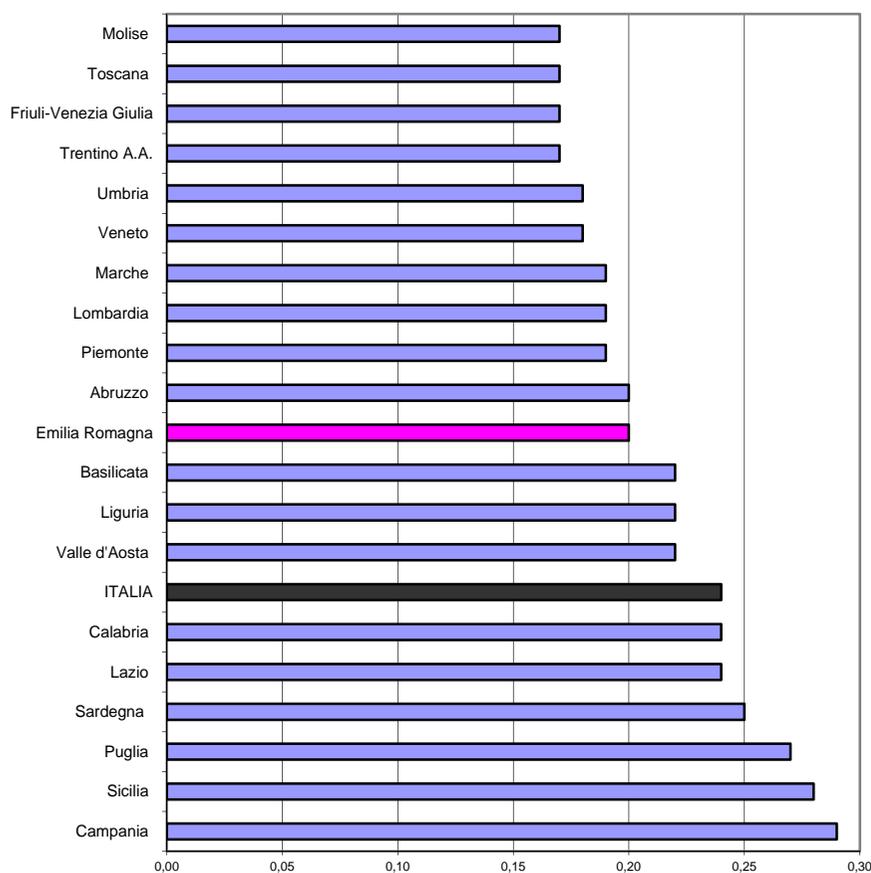


Anno 2008



*Le regioni italiane ordinate per gap mediano, anni 2004 e 2008*

*Anno 2004*

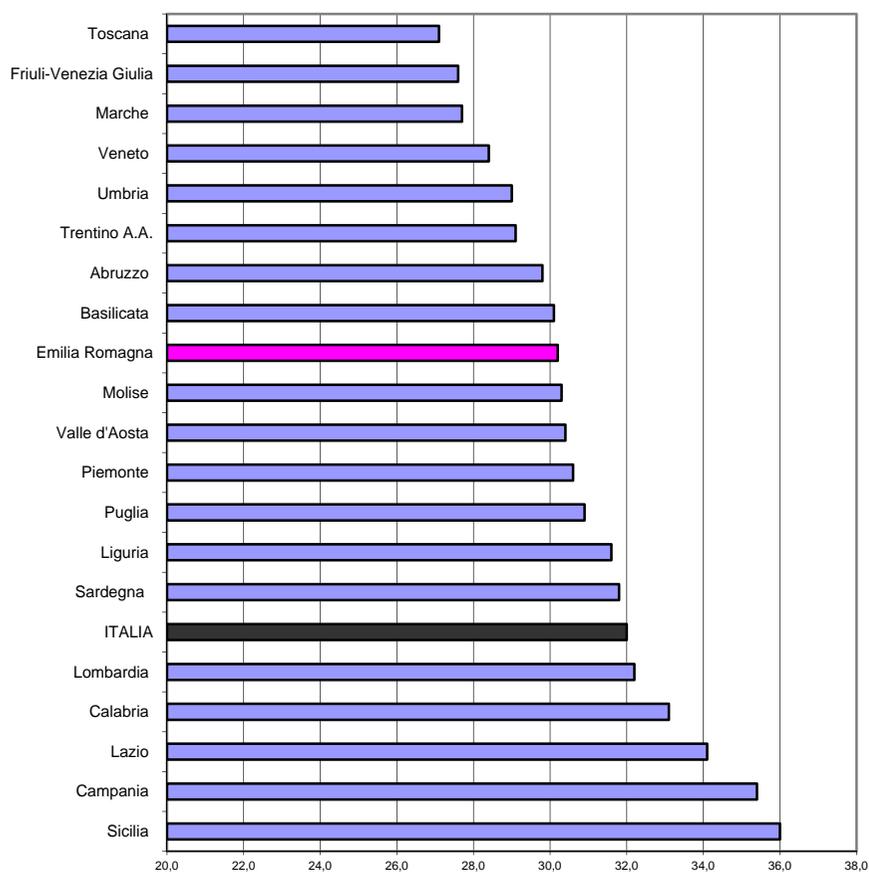


*Anno 2008*

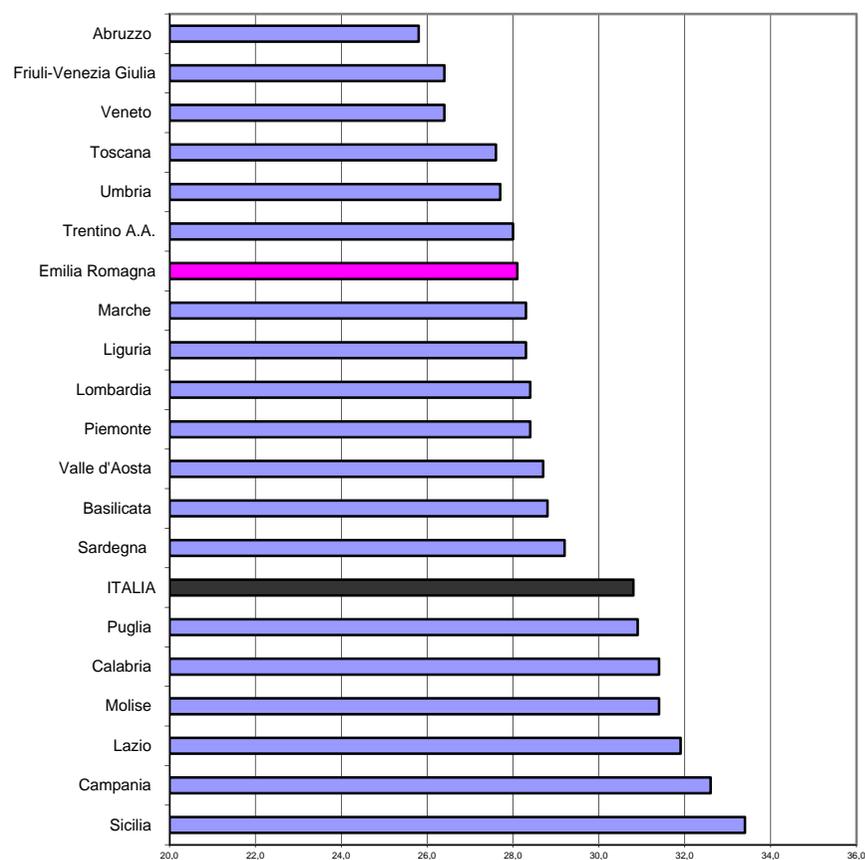


*Le regioni italiane ordinate per indice di Gini, anni 2004 e 2008*

*Anno 2004*

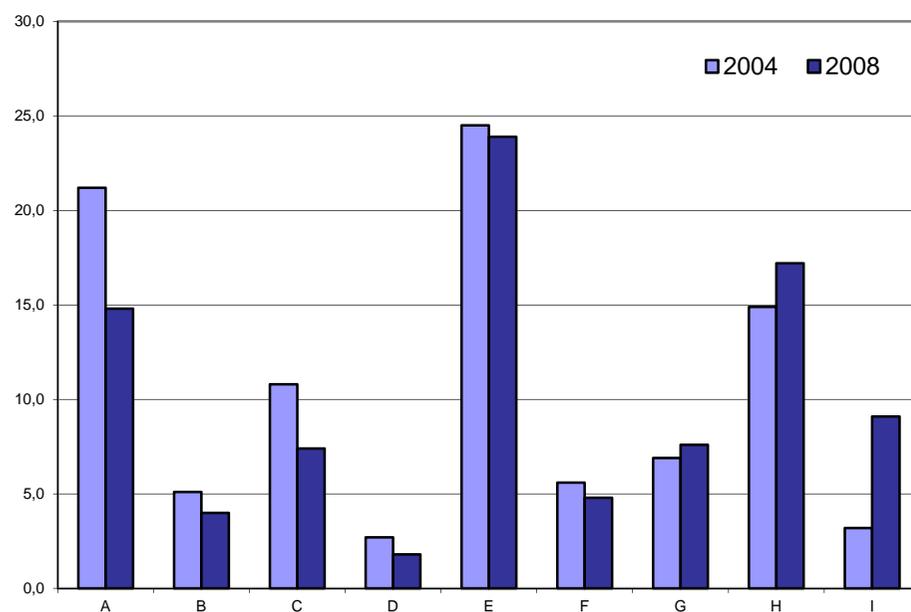


*Anno 2008*

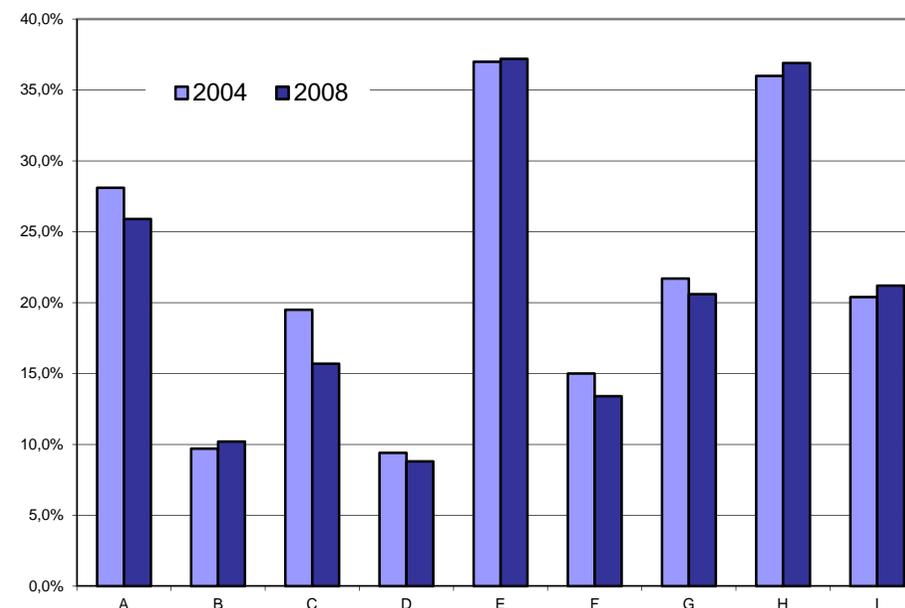


*Tasso di povertà per le tipologie familiari (2004-2008)*

*Emilia Romagna*

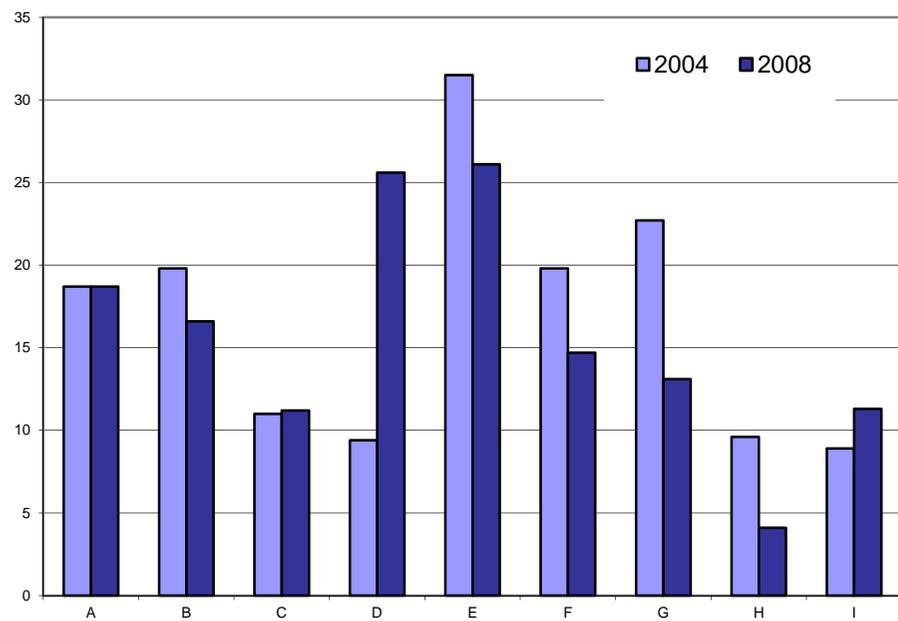


*Italia*

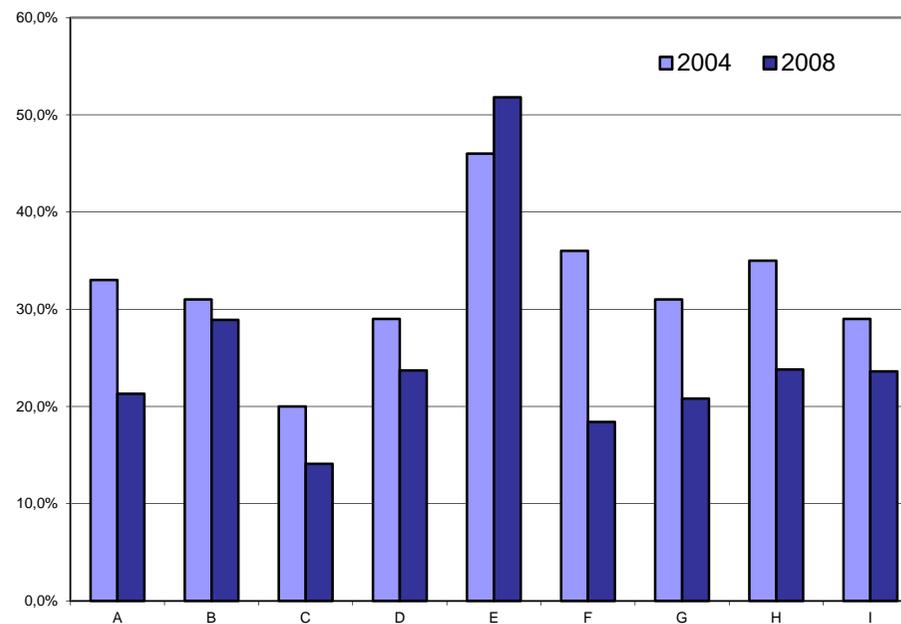


## Gap mediano per le tipologie familiari, (2004- 2008)

### in Emilia Romagna

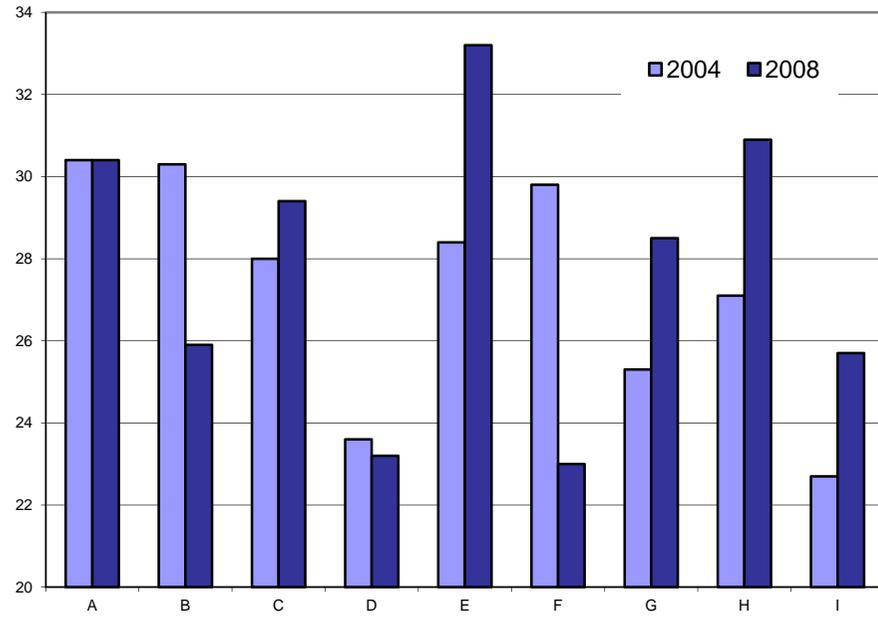


### Italia

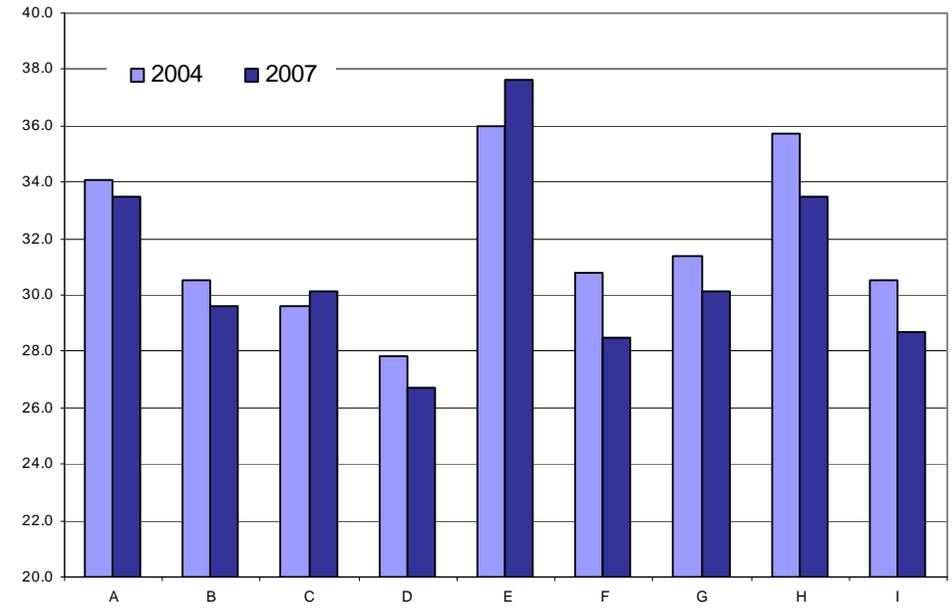


*Indice di Gini per le tipologie familiari (2004-2008)*

*Emilia Romagna*



*Italia*



**STIMA DI PARAMETRI DI POVERTÀ E DEPRIVAZIONE PER I DISTRETTI SOCIO-SANITARI DELLA REGIONE EMILIA-ROMAGNA**

**Parametri da stimare:**

- tasso di povertà
- gap mediano
- indice di Gini
- tasso di deprivazione materiale (o prevalenza delle famiglie in “stato di disagio economico”)

EUROSTAT utilizza come indicatore sintetico di **disagio economico** la quota di famiglie che presentano almeno tre deprivazioni su:

1) non riuscire a sostenere spese impreviste, 2) non potersi permettere una settimana di ferie lontano da casa in un anno, 3) avere arretrati (mutuo, affitto, bollette, debiti diversi dal mutuo), 4) non potersi permettere un pasto adeguato almeno ogni due giorni, 5) non potersi permettere di riscaldare adeguatamente l'abitazione, non potersi permettere: 6) lavatrice, 7) televisione a colori, 8) telefono, 9) automobile.

## Domini di studio

38 distretti socio sanitari della regione Emilia Romagna.

popolazioni di dimensione variabile, in molti casi piccole (media 115 mila, mediana 102 mila)

<b>Distretto</b>	<b>Popolazione</b>
CASTELNUOVO NE' MONTI	34,508
PAVULLO NEL FRIGNANO	41,461
VALLI TARO E CENO	46,666
...	.....
REGGIO EMILIA	222,278
CITTA' DI BOLOGNA	377,220

al fine di disporre di un numero di unità statistiche maggiore abbiamo considerato anche i distretti della Toscana

## Campioni EU-SILC nei domini

- per 4 domini non ci sono famiglie campionate (Montecchio, Guastalla, Piacenza ponente, Ferrara ovest)
- per gli altri la dimensione campionaria varia da 2 (Rubicone) a 122 (Città di Bologna). Media = 42, mediana = 33
- per alcuni distretti non sono presenti famiglie povere tra quelle campionate

## Problemi:

- Gli stimatori diretti sono “inaffidabili”, in molti casi neppure calcolabili;
- E' difficile calcolare la varianza associata agli stimatori diretti.

### Stime dirette - pesi

- pesi pubblicati da ISTAT riportano a totali regionali;
- ricalibrazione per ottenere coefficienti di riporto a livello di distretto fattibile, ma sconsigliabile: eccessiva variabilità dei pesi, conseguente instabilità dei pesi;
- soluzione adottata: ricalibrazione ad un livello intermedio: la provincia. Pesi calibrati sui totali provinciali della consistenza della popolazione per sesso e cinque classi di età

### Stima della varianza - soluzioni

- Bootstrap “semplificato” (ri-campionamento diretto delle famiglie all’interno dei domini. Ricalibrazione dei campioni replicati a livello provinciale);
- Calcolo delle varianze usando le formule elementari associate al campionamento casuale semplice e loro inflazione sulla base del fattore di Kish (ipotesi: bassa correlazione tra pesi campionari e variabili oggetto di interesse).

I due metodi portano a stime in sostanziale accordo tra loro

## I modelli per piccole aree

**Problema:** il ristretto ambito territoriale (la ridotta variabilità tra le stime dirette) determina la impossibilità di individuare covariate efficaci

### Soluzione proposta:

sotto ipotesi distributiva del reddito (Log-Normalità) sono state derivate relazioni funzionali tra i 3 parametri di povertà/disuguaglianza economica che danno luogo a modelli multivariati

Ad esempio:

- tasso di povertà =  $f(\text{indice di Gini}, \text{media del log}(\text{reddito}))$
- poiché l'indice di Gini è più affidabile del tasso, può costituire una sorta di covariata per il tasso

I primi risultati sono incoraggianti