

NUMERI PER DECIDERE

LE DISEGUAGLIANZE SOCIALI IN EMILIA-ROMAGNA: LE SOGLIE DI POVERTA'

Valeria Ardito

*Servizio Controllo Strategico e Statistica
Regione Emilia-Romagna*

vardito@regione.emilia-romagna.it

Bologna, 16 novembre 2009



Sommario

- Indagine EU-SILC
- Obiettivi del lavoro della Regione e di Ferrante, Pacei, Fabrizi del Dipartimento di Statistica
- Criteri di povertà
- Soglie di povertà
- Misure di povertà e disuguaglianza nella distribuzione dei redditi oggetto di stima nel lavoro
- Strategia di stima
- Analisi dei risultati

Indagine su reddito e condizioni di vita (EU-SILC)

Obiettivi conoscitivi

L'Indagine EU-SILC è una *indagine sulle famiglie* concordata in sede europea – in sostituzione della precedente indagine ECHP (*European Community Household Panel*), svolta dal 1994 al 2001 – e finalizzata a fornire informazioni attendibili, comparabili e tempestive su reddito, povertà e condizioni di vita delle famiglie in tutti i paesi membri dell'Unione.

Popolazione obiettivo

E' costituita dalle famiglie residenti in Italia e dagli individui che le compongono, anche se temporaneamente all'estero.

Sono escluse le persone che vivono abitualmente all'estero e i membri permanenti di convivenze istituzionali (ospizi, brefotrofi, istituti religiosi, caserme, ecc.). La *lista* è data dai fogli di famiglia delle anagrafi comunali.

Unità di rilevazione

E' la *famiglia di fatto* intesa come insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi, aventi dimora abituale nello stesso comune e coabitanti.

Sono oggetto di intervista *tutti gli individui* appartenenti alle famiglie campione *che abbiano compiuto 15 anni* nell'anno di riferimento del reddito, ovvero l'anno solare precedente a quello dell'intervista.



Regione Emilia-Romagna



Servizio Controllo Strategico e Statistica

Periodicità e riferimento temporale

L'indagine campionaria IT-SILC è condotta *annualmente da ISTAT*, in un periodo successivo alle dichiarazioni dei redditi.

Le informazioni acquisite mediante questionario a livello familiare e individuale fanno riferimento a periodi distinti:

- **al periodo dell'intervista (anno t)**
vanno ricondotte le informazioni sulle condizioni di vita attuali, ad es., caratteristiche dell'abitazione, il possesso di beni durevoli, condizioni di salute degli individui, livello di istruzione, condizione lavorativa, ecc.
- **agli ultimi dodici mesi**
vanno ricondotte, ad es., le principali spese per l'abitazione
- **all'anno precedente la data dell'indagine (anno $t - 1$)**
vanno ricondotte tutte le informazioni sulla situazione economica della famiglia e degli individui, ad es. redditi, mutui e prestiti, ecc.



Regione Emilia-Romagna



Servizio Controllo Strategico e Statistica

Parametri di interesse

L'indagine IT-SILC deve fornire sia stime di tipo *trasversale*, riferite cioè a un prefissato istante temporale, sia stime di tipo *longitudinale*, riferite a un intervallo di tempo.

I principali parametri di interesse sono:

- **ad ogni occasione di indagine t :**
 - valori sintetici del reddito familiare netto;
 - misure di incidenza di povertà;
 - misure di intensità di povertà;
 - misure di disuguaglianza nella distribuzione dei redditi;
- **tra due occasioni di indagine consecutive t e $t + 1$ o più:**
 - la variazione netta del reddito familiare medio;
 - la variazione netta della percentuale di famiglie povere;
 - la percentuale di famiglie che permangono nello stato di povertà per due o più occasioni di indagine.

Domini di studio

Gli **ambiti territoriali** rispetto ai quali sono riferiti i parametri di popolazione oggetto di stima sono:

- l'intero territorio nazionale
- le 5 ripartizioni geografiche (Italia nord-occidentale, nord-orientale, centrale, meridionale, insulare)
- le regioni italiane (a eccezione del Trentino-Alto Adige, le cui stime sono prodotte separatamente per le province autonome di Bolzano e Trento)

Disegno di rilevazione

La scelta del disegno dell'indagine è determinata principalmente dalla necessità di ottenere *stime corrette ed efficienti* di parametri, sia longitudinali sia trasversali, nei domini di studio.

Si è scelto di adottare una *indagine integrata*, per ottenere nel contempo entrambi i tipi di stime.

L'indagine integrata è basata su un *campione di tipo panel ruotato* (con una rotazione quadriennale).

Ad ogni occasione d'indagine t , il campione è costituito da *quattro gruppi di rotazione*. Nella successiva occasione $t+1$, un quarto del campione iniziale viene eliminato e sostituito da un nuovo gruppo e così via, in modo che – a regime (anno $t+3$) – ciascuno gruppo rimane nel campione per quattro anni consecutivi, secondo il seguente schema:

Schema di rotazione dei *panel*

Anno								
t	A	B	C	D				
$t + 1$		B	C	D	E			
$t + 2$			C	D	E	F		
$t + 3$				D	E	F	G	
$t + 4$					E	F	G	H
$t + 5$						F	G	H

Piano di campionamento

Il piano di campionamento adottato segue uno schema ormai consolidato, utilizzato da Istat per le principali indagini sulle famiglie condotte mediante intervista diretta.

Prevede **un disegno a due stadi, con stratificazione delle unità di primo stadio**, i comuni, mentre unità di secondo stadio sono le famiglie estratte dalle anagrafi dei comuni campione.

Primo stadio

Nell'ambito di ciascuna regione italiana, la procedura di stratificazione dei comuni si articola come segue.

1. Suddivisione dei comuni in due sottoinsiemi:
 - l'insieme dei comuni Auto Rappresentativi (AR) costituito dai comuni di maggiore dimensione;
 - l'insieme dei comuni Non Auto Rappresentativi (NAR) costituito dai rimanenti comuni;



Piano di campionamento – Primo stadio

La suddivisione avviene mediante la determinazione di una soglia λ di popolazione, diversa da regione a regione, così definita:

$$\lambda = \frac{\bar{n} \delta}{f}$$

\bar{n} è il numero minimo di famiglie da intervistare in ciascun comune campione della regione

δ è il numero medio regionale di componenti per famiglia;

$f = \frac{n}{N}$ è la frazione di campionamento

I comuni di dimensione maggiore o uguale a λ vengono definiti AR, quelli di dimensione inferiore vengono definiti NAR.

Piano di campionamento – Primo stadio

2. i **comuni AR** formano uno strato a se stante e sono inclusi con certezza nel campione;

i **comuni NAR**, invece, sono suddivisi in strati di ampiezza approssimativamente costante in termini di popolazione residente e all'incirca pari a λm_h , essendo m_h il numero di comuni campione da estrarre da ciascuno strato, posto pari a 4.

3. da ciascuno strato vengono selezionati i comuni campione (4 per ogni strato) con probabilità proporzionale alla loro dimensione demografica; per il generico comune c dello strato h tale probabilità è espressa dalla formula:

$$P_{hc} / P_h$$

Piano di campionamento – Secondo stadio

La **selezione delle famiglie** in ogni comune campione è effettuata da lista anagrafica, mediante una tecnica di selezione **di tipo sistematico**, attuato attraverso le seguenti fasi:

1. vengono messi in sequenza i fogli delle famiglie dell'anagrafe del comune
2. si calcola il passo di campionamento p , come rapporto tra il numero delle famiglie residenti nel comune e il corrispondente numero di famiglie da campionare

$$p = N / n$$

3. si seleziona la prima famiglia e poi di seguito una famiglia ogni p : entrano così a far parte del campione le n famiglie che nella sequenza al punto 1. occupano le seguenti posizioni:

$$1, 1 + p, 1 + 2p, \dots, 1 + (n - 1)p$$

Obiettivo del lavoro

Fornire stime di misure relative alla povertà sperimentata sia dalla popolazione complessiva in ciascuna regione italiana, sia da specifiche tipologie familiari nell'ambito di ciascuna regione, al fine di individuare i segmenti della società maggiormente a rischio di povertà ed esclusione sociale.

I parametri di interesse, scelti tra quelli previsti in ambito UE dal Consiglio di Laeken (2001), sono:

1. **il reddito disponibile equivalente medio**
2. **il reddito disponibile equivalente mediano**
3. **i tassi di povertà basati sulle soglie di povertà nazionali**
4. **i tassi di povertà basati sulle soglie di povertà regionali**
5. **il *poverty gap* mediano basato su SPN**
6. **il *poverty gap* mediano basato su SPR**
7. **l'indice di concentrazione di Gini**
8. **il rapporto interquintilico**



Criteri di povertà

Nelle analisi del fenomeno della povertà occorre precisare un criterio di povertà, sulla cui base sia possibile *identificare* i soggetti come poveri.

La fase di identificazione dei soggetti poveri prevede:

1. la specificazione di un **concetto di povertà** (assoluta o relativa)
2. la scelta di un **indicatore** su cui basare le misure di povertà (reddito o consumo)
3. l'individuazione di una **unità di analisi** (famiglia o individuo)
4. la definizione di una **soglia di povertà**

Le scelte effettuate da Eurostat (con l'avvio alla nuova indagine EU-SILC) e da Istat (che dal 1984 diffonde misure ufficiali di povertà basate sulla Indagine sui consumi delle famiglie), per tali sotto-fasi sono difformi, tranne che nella adozione di un concetto di povertà relativa.

La **povertà relativa** è intesa come una condizione di deficit di risorse monetarie necessarie per mantenere lo standard di vita corrente, definito in funzione del livello medio della distribuzione delle risorse nella popolazione di riferimento.

La **povertà assoluta**, invece, evoca una condizione di deprivazione monetaria tale da impedire di soddisfare i propri bisogni essenziali ed è rilevata dalla incapacità di acquistare un paniere minimo di beni e servizi necessario a garantire la sussistenza.

Le scelte di Istat ed Eurostat

1. **ISTAT** adotta un **concetto di povertà** relativa
2. l' **indicatore** su cui basa le misure di povertà è la **spesa per consumi**
3. l' **unità di analisi** è la **famiglia**
4. la **soglia di povertà** è pari alla spesa media pro-capite per consumi. Per rendere equivalenti la spesa per consumi di famiglie di ampiezza diversa si usa la scala di Carbonaro, che tiene conto solo del numero di componenti

Si definisce **povera** una famiglia di due componenti che abbia una spesa per consumi inferiore o uguale alla soglia di povertà

1. **EUROSTAT** adotta un **concetto di povertà** relativa
2. l' **indicatore** su cui basa le misure di povertà è il **reddito**
3. l' **unità di analisi** è l'**individuo**
4. la **soglia di povertà** è pari al 60% del valore mediano dei redditi familiari equivalenti. Per rendere equivalenti i redditi di famiglie diverse si usa la scala di OCSE modificata, che tiene conto non solo del numero ma anche dell'età dei componenti

Si definisce **povero** un individuo appartenente a una famiglia che disponga di un reddito familiare equivalente inferiore o uguale alla soglia di povertà

Reddito familiare in EU-SILC

Reddito familiare disponibile

E' dato dalla somma delle diverse tipologie di reddito (redditi da lavoro dipendente e autonomo, di quelli da capitale reale e finanziario, delle pensioni e degli altri trasferimenti pubblici e privati) percepite da ciascun membro della famiglia di 15 anni e oltre, al netto del prelievo tributario e di eventuali imposte patrimoniali (imposte personali, contributi sociali a carico dei lavoratori dipendenti e autonomi, ICI, ecc.), nonché degli eventuali trasferimenti versati ad altre famiglie (come, ad esempio, gli assegni di mantenimento versati all'ex coniuge)

Reddito familiare equivalente

E' ottenuto dividendo il reddito disponibile di ciascuna famiglia per il fattore di scala (OCSE modificata)

$$Eq_size = 1 + 0.5 (CF_{14+} - 1) + 0.3 CF_{13-}$$

CF_{14+} = numero di componenti di età maggiore o uguale a 14 anni

CF_{13-} = numero di componenti di età minore o uguale a 13 anni

E' assegnato a ciascun membro della famiglia

Le soglie di povertà relativa in Eu-Silc

A livello nazionale

- 1. soglia di povertà standard (SPN)** pari al 60% del reddito mediano equivalente nazionale
- 2. soglia di povertà estrema o severa (SPN80)** pari all'80% di SP; consente di individuare coloro che versano in condizioni di estrema povertà, in quanto dispongono di un reddito equivalente molto al di sotto (del 20% inferiore) della soglia standard
- 3. soglia di quasi povertà o di rischio di povertà (SPN120)** pari al 120% di SP; consente di individuare coloro che, pur non risultando poveri, percepiscono un reddito equivalente di poco superiore (non più del 20%) rispetto alla soglia standard e che pertanto sono a rischio di povertà

A livello regionale

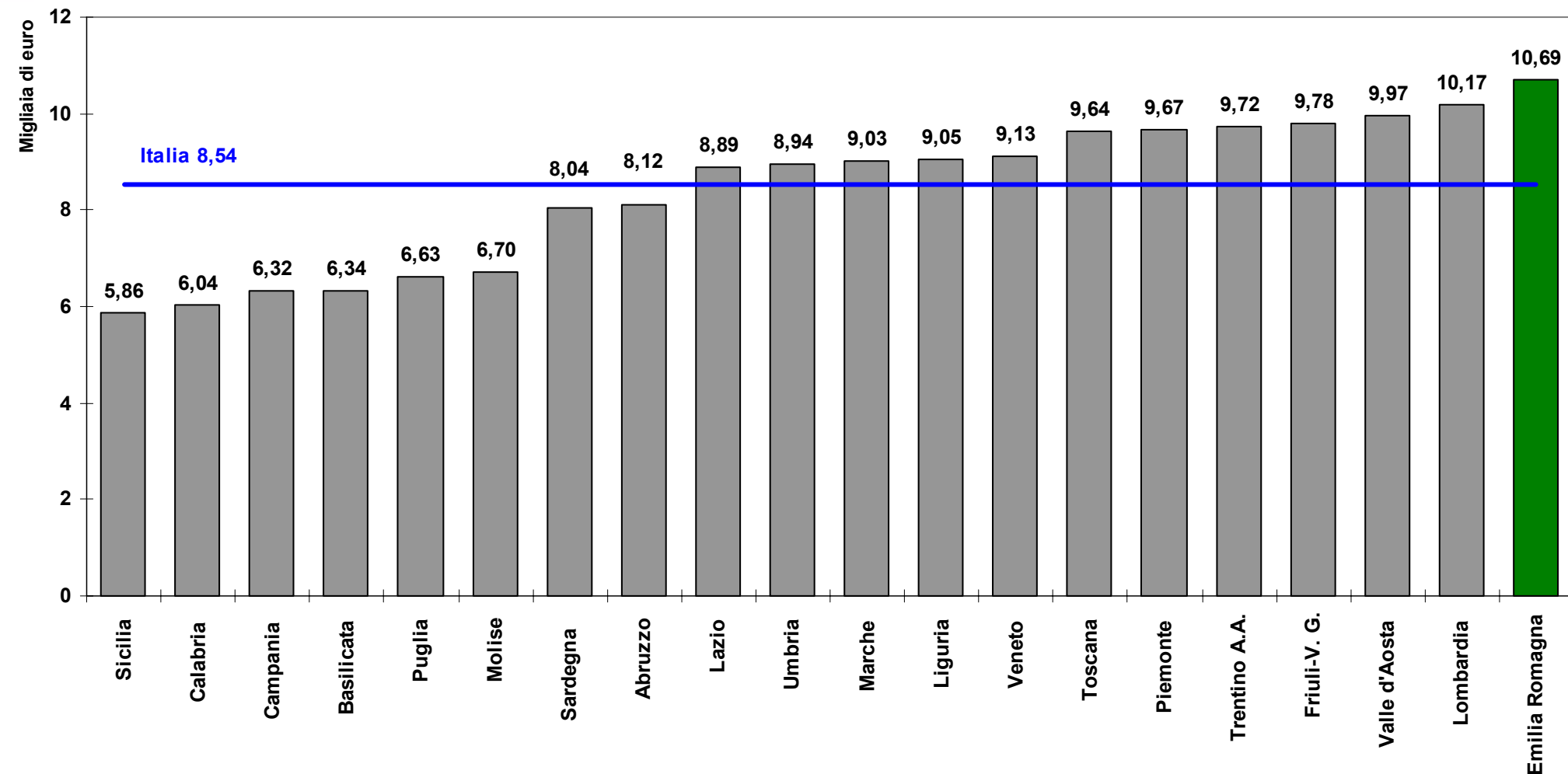
- 1. soglia di povertà standard regionale (SPR)** pari al 60% del reddito mediano equivalente di una data regione
- 2. soglia regionale di povertà estrema (SPR80)** pari all'80% di SPR
- 3. soglia regionale di rischio di povertà (SPR120)** pari al 120% di SPR

Soglie di povertà nazionali e regionali (IT-SILC 2005)

REGIONE	Soglia di povertà estrema	Soglia di povertà standard	Soglia di rischio di povertà
Piemonte	7.733	9.667	11.600
Valle d'Aosta	7.977	9.972	11.966
Lombardia	8.139	10.174	12.209
Trentino A.A.	7.776	9.721	11.665
Veneto	7.304	9.130	10.956
Friuli-V. G.	7.823	9.779	11.735
Liguria	7.242	9.052	10.862
Emilia Romagna	8.551	10.688	12.826
Toscana	7.713	9.642	11.570
Umbria	7.155	8.944	10.733
Marche	7.222	9.028	10.833
Lazio	7.112	8.891	10.669
Abruzzo	6.492	8.115	9.738
Molise	5.364	6.705	8.046
Campania	5.059	6.324	7.589
Puglia	5.301	6.626	7.951
Basilicata	5.069	6.336	7.603
Calabria	4.832	6.040	7.248
Sicilia	4.684	5.855	7.026
Sardegna	6.433	8.042	9.650
Italia	6.835	8.544	10.253



Le regioni italiane ordinate per soglia di povertà regionale



Soglie di povertà in termini di reddito monetario per alcune tipologie familiari

in Italia

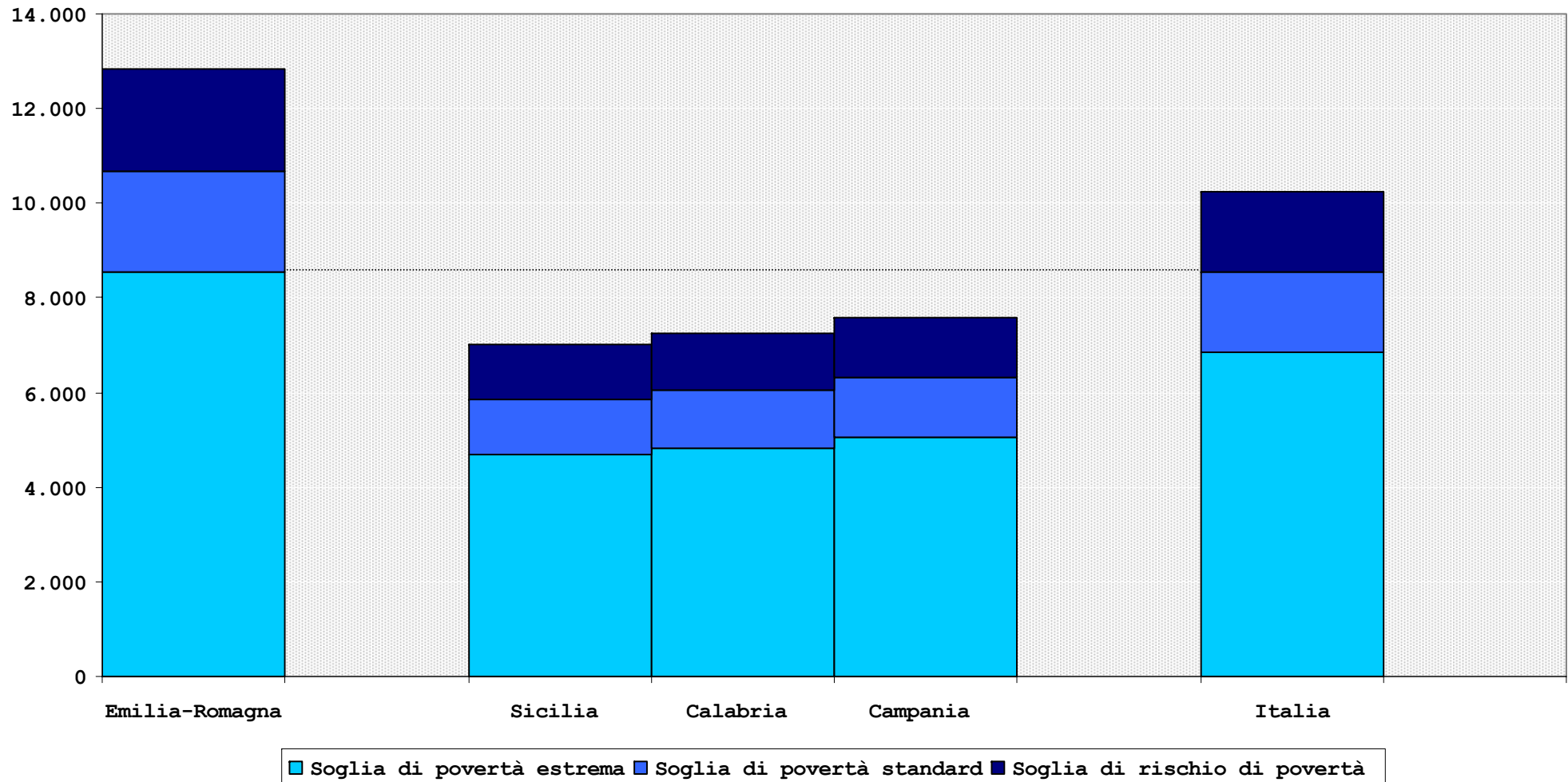
e

in Emilia-Romagna

TIPOLOGIA FAMILIARE	Soglia di povertà nazionale	SPN su base mensile	Fattore di scala
Un adulto	8.544	712	1
Due componenti			
- Due adulti	12.816	1.068	1,5
- Un adulto e un bambino	11.107	926	1,3
Tre componenti			
- Tre adulti	17.088	1.424	2,0
- Due adulti e un bambino	15.379	1.282	1,8
- Un adulto e due bambini	13.670	1.139	1,6
Quattro componenti			
- Quattro adulti	21.360	1.780	2,5
- Due adulti e due bambini	17.942	1.495	2,1
- Un adulto e tre bambini	16.234	1.353	1,9

TIPOLOGIA FAMILIARE	Soglia di povertà regionale	SPR su base mensile	Fattore di scala
Un adulto	10.688	891	1
Due componenti			
- Due adulti	16.032	1.336	1,5
- Un adulto e un bambino	13.894	1.158	1,3
Tre componenti			
- Tre adulti	21.376	1.781	2,0
- Due adulti e un bambino	19.238	1.603	1,8
- Un adulto e due bambini	17.101	1.425	1,6
Quattro componenti			
- Quattro adulti	26.720	2.227	2,5
- Due adulti e due bambini	22.245	1.870	2,1
- Un adulto e tre bambini	20.307	1.692	1,9

Confronto tra le soglie di povertà in Emilia-Romagna, nelle regioni "più povere" del meridione e in Italia



Parametri di interesse

1. Reddito equivalente medio

$$\bar{Y} = \frac{1}{N} \sum_{j=1}^N y_j$$

y_j è il reddito del j -esimo individuo che appartiene al dominio d

N è il numero di individui nel dominio d

Rappresenta il livello di reddito che potrebbe percepire ciascun individuo se il reddito complessivamente detenuto fosse equamente distribuito tra loro

Misura il livello medio di benessere di una certa popolazione

2. Reddito equivalente mediano

$$Me(y) = y_{(c)}$$

$y_{(c)}$ è il reddito dell'individuo che occupa la posizione centrale nella graduatoria dei redditi equivalenti del dominio d ordinati in modo crescente ossia in modo tale che

$$y_{(1)} \leq \dots \leq y_{(c)} \leq \dots \leq y_{(N)}$$

Rappresenta il valore centrale che divide la distribuzione ordinata dei redditi in due parti di pari grandezza, in modo tale che il 50% degli individui abbia un reddito inferiore al valore mediano e il 50% superiore

Non risente della presenza di *outliers*



Regione Emilia-Romagna



Servizio Controllo Strategico e Statistica

Parametri di interesse

3. Tassi di povertà basati sulle soglie di povertà nazionali

$$TPN = \frac{\sum_{j=1}^N \mathbf{1}(y_j \leq SPN)}{N}$$

4. Tassi di povertà basati sulle soglie di povertà regionali

$$TPR = \frac{\sum_{j=1}^N \mathbf{1}(y_j \leq SPR)}{N}$$

Sono dati al rapporto tra il numero di individui del dominio d poveri rispetto alle varie soglie di povertà e il numero totale di individui che appartengono dominio

Misurano l'incidenza del fenomeno (ma nulla dicono sulla sua gravità)

Parametri di interesse

5. *Poverty Gap* mediani basati sulle soglie di povertà nazionali

$$GapN = \frac{Me(SPN - y_{poor})}{SPN}$$

6. *Poverty Gap* mediani basati sulle soglie di povertà regionali

$$GapR = \frac{Me(SPR - y_{poor})}{SPR}$$

Sono dati dalla mediana degli scarti (*gap*) tra le varie soglia di povertà e il reddito equivalente a disposizione di ciascun individuo povero che appartiene al dominio d , rapportato alla corrispondente soglia (in modo tale da ottenere un indice che varia tra 0 e 1)

Misurano l'intensità del fenomeno ossia lo stato di indigenza in cui versano i poveri



Parametri di interesse

7. Indice di concentrazione di Gini

$$G = \frac{2Cov(y, RF(y))}{\bar{Y}}$$

$RF(y)$ è il rango frazionario ossia il posto che l'individuo occupa nella graduatoria dei redditi equivalenti del dominio d ordinati in modo crescente, successivamente diviso per N

E' un indice che varia tra 0 e 1. E' uguale a 0 nel caso di perfetta equità della distribuzione dei redditi e pari a 1 nel caso di massima concentrazione

Misura il grado di concentrazione della distribuzione dei redditi

8. Rapporto interquintilico

$$\frac{S_{80}}{S_{20}}$$

E' dato dal rapporto tra l'ammontare di reddito percepito dal quinto di individui più "ricchi" (S_{80}) del dominio d e quello detenuto dal quinto di individui più "poveri" (S_{20})

Misura il livello di disuguaglianza nella distribuzione dei redditi



Domini di studio

I parametri di interesse sono stimati con riferimento a diversi domini di studio:

1. per l'intero territorio nazionale

2. per ciascuna regione

per confrontare la situazione della povertà in Emilia-Romagna, con quella nazionale e delle altre regioni

3. per ciascuna tipologia familiare

per evidenziare i segmenti della società maggiormente esposti al rischio di povertà ed esclusione sociale

4. per ciascuna tipologia familiare all'interno di ogni regione

per confrontare la struttura "demografica" della povertà nelle diverse aree del paese

1. e 2. sono **domini pianificati** dall'indagine IT-SILC, per i quali si riescono a ottenere stime con un livello di precisione sufficientemente elevato

3. pone problemi in fase di stima utilizzando la procedura adottata da Istat

4. sono **domini non pianificati**, con numerosità campionare molto basse, per i quali è necessario predisporre metodologie di stima molto più complessi



Tipologie familiari (EU-SILC)

- A. Monocomponenti
- B. Due adulti, senza figli a carico, entrambi età inferiore a 65 anni
- C. Due adulti, senza figli a carico, almeno uno età superiore a 65 anni
- D. Altra tipologia, senza figli a carico
- E. Monogenitore, con uno o più figli a carico
- F. Due adulti, con un figlio a carico
- G. Due adulti, con due figli a carico
- H. Due adulti, con tre o più figli a carico
- I. Altra tipologia con figli a carico

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

Considerando le 9 tipologie familiari all'interno delle 20 regioni italiane si ottengono 180 domini di studio (non pianificati) che si configurano, in termini di stima, come “**piccole aree**” ossia come sotto-popolazioni per cui i dati campionari non sono sufficienti a produrre stime di affidabili e che quindi presuppongono l'adozione di apposite metodologie di stima

Famiglie rispondenti per regione e tipologia familiare. Indagine IT-SILC 2005

	Tipologia									
Regione	A	B	C	D	E	F	G	H	I	Totale
Piemonte	441	219	223	179	36	190	145	21	54	1508
Valle D'Aosta	136	67	51	39	16	41	36	4	15	405
Lombardia	600	374	390	359	58	271	239	46	144	2481
Trentino A.A.	233	106	107	123	33	94	117	39	58	910
Veneto	432	229	225	309	50	209	193	32	122	1801
Friuli V.G.	296	129	150	132	31	100	87	9	29	963
Liguria	354	122	158	124	32	105	61	13	37	1006
Emilia-Romagna	465	206	276	262	56	164	155	19	79	1682
Toscana	381	192	224	274	41	195	134	29	90	1560
Umbria	234	90	175	157	22	112	109	15	67	981
Marche	242	115	171	176	35	128	111	26	91	1095
Lazio	428	181	249	207	59	160	188	19	115	1606
Abruzzo	113	32	91	98	16	47	75	15	54	541
Molise	122	37	69	64	5	42	57	7	18	421
Campania	231	100	180	185	46	124	207	82	148	1303
Puglia	223	85	115	155	20	102	155	43	88	986
Basilicata	94	51	88	63	6	38	81	31	34	486
Calabria	132	57	85	73	16	50	96	29	53	591
Sicilia	253	110	160	125	22	116	158	43	91	1078
Sardegna	116	80	67	114	17	63	82	24	65	628
Totale	5526	2582	3254	3218	617	2351	2486	546	1452	22032

Individui rispondenti per regione e tipologia familiare. Indagine IT-SILC 2005

	Tipologia									
Regione	A	B	C	D	E	F	G	H	I	Totale
Piemonte	441	438	446	579	84	570	580	107	236	3481
Valle D'Aosta	136	134	102	124	42	123	144	21	64	890
Lombardia	600	748	780	1179	146	813	956	235	639	6096
Trentino A.A.	233	212	214	420	78	282	468	205	297	2409
Veneto	432	458	450	1047	117	627	772	165	567	4635
Friuli V.G.	296	258	300	432	81	300	348	48	137	2200
Liguria	354	244	316	400	69	315	244	67	165	2174
Emilia-Romagna	465	412	552	858	139	492	620	100	357	3995
Toscana	381	384	448	905	95	585	536	145	419	3898
Umbria	234	180	350	536	52	336	436	76	323	2523
Marche	242	230	342	612	83	384	444	134	432	2903
Lazio	428	362	498	701	153	480	752	99	526	3999
Abruzzo	113	64	182	338	44	141	300	76	255	1513
Molise	122	74	138	217	14	126	228	35	87	1041
Campania	231	200	360	649	128	372	828	419	717	3904
Puglia	223	170	230	533	56	306	620	217	403	2758
Basilicata	94	102	176	219	16	114	324	160	158	1363
Calabria	132	114	170	253	40	150	384	154	237	1634
Sicilia	253	220	320	420	56	348	632	224	435	2908
Sardegna	116	160	134	397	40	189	328	126	291	1781
Totale	5526	5164	6508	10819	1533	7053	9944	2813	6745	56105



Procedura di stima di ISTAT

La costruzione delle stime trasversali si basa sull'utilizzo di una famiglia di stimatori, noti in letteratura come **calibration estimator** (stimatori diretti o di ponderazione vincolata), utilizzati correntemente da Istat nelle varie indagini campionarie, che prevedono la determinazione di un **peso** (detto **coefficiente di riporto all'universo**) w_i da assegnare a ciascuna unità campionaria.

Il principio su cui si basano è che le unità appartenenti al campione devono rappresentare anche le unità della popolazione che non sono state incluse. A tale scopo, ad ogni unità campionaria è attribuito un peso che indica quante unità della popolazione essa deve rappresentare.

(Così, ad es., se all'unità campionaria i -esima è attribuito un peso $w_i = 30$, allora questa unità rappresenta se stessa e altre 29 unità della popolazione non incluse nel campione.)

A tutti i componenti di una stessa famiglia si assegna il medesimo peso.

Procedura di determinazione dei pesi

Il **peso finale** da attribuire a ciascuna famiglia si ottiene mediante una procedura complessa che:

- tiene conto del disegno di campionamento (la **probabilità di inclusione** di ciascuna famiglia nel campione)
- corregge l'effetto distorsivo della **mancata risposta totale** (MRT), dovuta all'impossibilità di intervistare alcune delle famiglie selezionate per irreperibilità o per rifiuto all'intervista;
- tiene conto della **conoscenza di totali noti** di variabili ausiliarie, correlate alle variabili oggetto di interesse, disponibili da fonti esterne all'indagine, nel senso che le stime campionarie dei totali di tali variabili ausiliarie, ottenute mediante l'utilizzo dei pesi finali, devono coincidere con i valori noti di popolazione.

Procedura di determinazione dei pesi

La strategia per la costruzione dei pesi da attribuire a ciascuna famiglia rispondente i si articola nelle seguenti fasi:

1. si calcola un peso iniziale, d_i , detto **peso diretto**, ottenuto come reciproco della probabilità d'inclusione della famiglia i nel campione;
2. si determinano i **fattori correttivi** per MRT, ${}_1c_i$, come l'inverso del tasso di risposta del comune in cui la famiglia risiede;
3. si ottengono i **pesi base** b_i , o pesi corretti per MRT, moltiplicando il peso diretto d_i per il corrispondente fattore correttivo per MRT;
4. si costruiscono i fattori correttivi, ${}_2c_i$, che consentono di soddisfare la condizione di uguaglianza tra i totali noti delle variabili ausiliarie e le corrispondenti stime campionarie;
5. si ottengono i **pesi finali** w_i come prodotto del peso base b_i per i fattori correttivi ${}_2c_i$;



Regione Emilia-Romagna



Servizio Controllo Strategico e Statistica

Procedura di costruzione dei pesi finali: Passo 1

Il peso iniziale o diretto, d_i , per la famiglia i , al passo 1 è calcolato come reciproco della probabilità d'inclusione π_i della famiglia i nel campione:

$$d_i = \frac{1}{\pi_i} = m_h \frac{P_h}{P_{hc}} \frac{N_{hc}}{n_{hc}}$$

m_h è il numero di comuni campione nello strato h

P_h indica il totale di persone residenti nello strato h

P_{hc} è il totale di persone residenti nel comune c dello strato h

N_{hc} è il totale di famiglie residenti nel comune c dello strato h

n_{hc} è il numero di famiglie campione nel comune c dello strato h



Regione Emilia-Romagna



Servizio Controllo Strategico e Statistica

Procedura di costruzione dei pesi finali: Passo 2

Nella fase di raccolta delle informazioni presso le famiglie campione, alcune di queste si trovano nella impossibilità di partecipare alla indagine.

Se il meccanismo che genera la MRT è **trascurabile**, si assume che il comportamento dei rispondenti sia del tutto simile a quello dei non rispondenti.

In tal caso, il fattore di correzione per MRT al passo 2 è calcolato come reciproco della probabilità di risposta della famiglia i

$${}_1C_i = \frac{1}{\rho_i} = \frac{n_{hc}}{n_{hc}^R}$$

n_{hc}^R è il numero di famiglie rispondenti nel comune c dello strato h

Procedura di costruzione dei pesi finali: Passo 3

Al passo 3, i pesi base per la famiglia i , b_i , corretti per MRT, si ottengono moltiplicando il peso diretto d_i per il corrispondente fattore correttivo ${}_1c_i$ per MRT

$$b_i = d_i {}_1c_i = m_h \frac{P_h}{P_{hc}} \frac{N_{hc}}{n_{hc}} \frac{n_{hc}}{n_{hc}^R} = m_h \frac{P_h}{P_{hc}} \frac{N_{hc}}{n_{hc}^R}$$



Regione Emilia-Romagna



Servizio Controllo Strategico e Statistica

Procedura di costruzione dei pesi finali: Passi 4 e 5

I fattori correttivi ${}_2c_j$ del passo 4 sono ottenuti dalla risoluzione di un problema di minimo vincolato, in cui la funzione obiettivo è

$$\min_{{}_2c_i} \left\{ \sum_i dist(w_i; b_i) \right\}$$

ove $dist(;)$ è una opportuna funzione di distanza tra i pesi base, b_i , e i pesi finali, $w_i = b_i {}_2c_i$.

I vincoli sono definiti dalla condizione di uguaglianza tra stime campionarie dei totali noti di popolazione X e i valori noti degli stessi.

$$\hat{X}_w = \sum_{i=1}^n x_i w_i$$

$$X = \sum_{j=1}^N x_j$$

Totali noti adottati da ISTAT

Gli stimatori di ponderazione vincolata si basano sull'utilizzo di informazioni ausiliarie, sintetizzate nei totali noti di variabili ausiliarie che, correlate con le variabili di interesse, hanno la funzione di aumentare l'accuratezza delle stime.

I totali noti utilizzati nella *indagine IT-SILC 2005*, definiti da EUROSTAT, sono:

1. la popolazione residente per ripartizione territoriale, sesso e classi di età (0-15, 16-19, ... , 70-74, 75 e più) al 31.12.2004;
2. il totale delle famiglie residenti in ciascuna regione al 31.12.2005 .

Quelli aggiunti da ISTAT per esigenze informative interne:

3. la popolazione residente per regione, sesso e classi di età (0-15, 16-19, ... , 70-74, 75 e più) al 31.12.2004;
4. la popolazione residente per ripartizione, sesso e classi di età (0-15, 16-19, ... , 70-74, 75 e più) al 31.12.2005;
5. la popolazione straniera maggiorenne residente per ripartizione, sesso e provenienza (UE, non UE) al 31.12.2004.



Regione Emilia-Romagna



Servizio Controllo Strategico e Statistica

Lo stimatore diretto

Data una variabile obiettivo y , osservato un campione di valori $(y_1, \dots, y_i, \dots, y_n)$ nel dominio d cui è associato il vettore di pesi $(w_1, \dots, w_i, \dots, w_n)$, lo **stimatore diretto** del totale Y della variabile obiettivo nel dominio d è dato da:

$$\hat{Y} = \sum_{i=1}^n y_i w_i$$

che per la media diventa

$$\hat{\bar{Y}} = \frac{1}{N} \sum_{i=1}^n y_i w_i$$

dove

$$N = \sum_{i=1}^n w_i$$

Strategia di stima dei parametri di povertà adottata nel lavoro

1. Produrre **stime dirette** dei parametri di interesse per tutti i domini di studio
2. Valutare **l'affidabilità delle stime dirette** in ciascun dominio di studio
3. Sostituire le stime dirette con stime ottenute con metodologie più complesse (**stimatori per piccole aree**) per i domini in cui le prime non risultino sufficientemente affidabili.

Al punto 1. non possono essere applicati i pesi utilizzati da ISTAT, in quanto non permettono di ottenere stime affidabili a livello dei domini di nostro interesse (domini non pianificati in IT-SILC).

La procedura di calibrazione adottata da ISTAT impone che le stime delle variabili ausiliarie rispetto a cui si è calibrato (popolazione per classi d'età, sesso, ecc.) riproducano esattamente i totali noti (da fonte anagrafica), ma non garantiscono che questi totali siano rispettati anche a livello di sottopopolazioni.



Regione Emilia-Romagna



Servizio Controllo Strategico e Statistica

La procedura di calibrazione impiegata nel lavoro

I **pesi finali** utilizzati nel lavoro per produrre le stime dirette sono stati ottenuti, partendo dai pesi base b_i identificati da ISTAT e modificando il passo di calibrazione (passi 4 e 5).

Per ottenere stime affidabili in tutti i domini di interesse, i fattori di correzione al passo 4. sono stati ottenuti vincolando ai seguenti **totali noti**:

- popolazione residente per regione, sesso e classe d'età al 31/12/2005;
- totale di famiglie residenti per regione e tipologia familiare come da stime ottenute sulla base della Rilevazione Continua sulle Forze di Lavoro (media 2005)

in modo tale che

$$P_d = \sum_{i=1}^{n_d} w_i, \quad d = 1, \dots, 180$$

Trattamento di *outliers* e valori negativi

Nel data set di IT-SILC sono presenti valori molto elevati o negativi del reddito disponibile, non necessariamente dovuti ad errore. Tuttavia

- valori molto elevati influenzano la stima del valor medio e dell'indice di Gini
- valori negativi possono determinare valori superiori a 1 per i gap mediani e l'indice di Gini

Pertanto, si è proceduto a un trattamento degli *outliers* e dei valori negativi (in tutto 272 unità su 56.105), mediante winsorizzazione, riportando gli stessi entro soglie prefissate delle due code della distribuzione dei redditi equivalenti.

Per i **valori estremi** (124 unità) è stata fissata la soglia di 80.000 euro, che corrisponde al percentile 0,995

Per i **valori negativi** (148 unità) è stata fissata la soglia di 0 euro

Valutazione dell'affidabilità delle stime dirette

Dato che il disegno campionario è complesso e i parametri oggetto di stima sono funzioni non lineari dei dati (fatta eccezione per il valor medio del reddito disponibile equivalente), non è facile ottenere stime della varianza degli stimatori diretti dei parametri di povertà

Pertanto, per valutare l'affidabilità delle stime dirette, si è fatto ricorso a metodi di stima della varianza basati sul ricampionamento, mediante la costruzione di un **algoritmo di tipo *bootstrap***

Come misura dell'affidabilità delle stime è stato adottato il coefficiente di variazione delle stime

Valori sintetici dei coefficienti di variazione delle stime dirette per regione e tipologia familiare

Per tutte le regioni italiane

	Reddito equivalente medio	Reddito equivalente mediano	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	Rapporto interq,
<i>min cv</i>	2,71	2,10	9,69	7,21	6,01	9,26	7,06	5,44	0,43	2,57	3,79	8,27
<i>avg cv</i>	8,65	10,07	46,60	33,66	24,79	46,41	35,63	24,77	57,64	49,58	11,89	34,18
<i>max cv</i>	33,77	83,88	178,56	140,47	140,47	178,56	238,45	140,47	484,02	331,29	47,67	574,32

Per l'Emilia-Romagna

	Reddito equivalente medio	Reddito equivalente mediano	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	Rapporto interq,
<i>min cv</i>	2,98	2,67	18,49	10,14	8,32	10,18	8,47	5,44	7,67	8,05	4,65	8,75
<i>avg cv</i>	6,25	6,68	47,02	38,98	33,11	38,32	42,73	21,03	55,54	34,42	9,35	19,37
<i>max cv</i>	13,95	20,72	107,77	75,40	75,45	75,45	164,13	46,23	192,46	51,94	18,37	43,93



Metodologia di stima per piccole aree

Rispetto agli stimatori diretti (da disegno), gli SME (da modello) assumono un modello probabilistico per spiegare come i dati di popolazione relativi ad una certa variabile siano stati generati. In altre parole, si assume che tra i parametri di interesse esista una “relazione strutturale” che tipicamente coinvolge informazioni ausiliarie, utili per spiegare la variabilità sottostante ai parametri stessi e che permette di sfruttare, per la stima in un dominio, le informazioni raccolte in tutti gli altri domini.

A tal fine occorre scegliere

1. le informazioni ausiliarie
2. il tipo di modello per piccole aree, composto dal “sampling model” che definisce le assunzioni sullo stimatore diretto e la loro relazione con i relativi parametri di dominio, ed il “linking model” che pone in relazione i parametri alle informazioni ausiliarie specifiche di dominio.

Al punto 1. è stata considerata informazione ausiliaria contenuta in Contabilità Regionale, Rilevazione Continua sulle FL e altri fonti amministrative, selezionata secondo criteri di efficacia statistica, disponibilità e interpretabilità, anche economica, dei modelli specificati.

Si è scelta la stima del PIL pro-capite regionale ottenuta dalla Contabilità Regionale.

Al punto 2. è stato scelto un modello univariato di tipo Beta-Logistico, specificato “a livello di area” secondo un approccio bayesiano gerarchico, che permette di tener conto di tutte le fonti di incertezza riguardanti i dati e quindi di valutare in modo esaustivo l’affidabilità delle stime prodotte.

Modello univariato Beta-logistico

Sia Y_{ij} un parametro di povertà da stimare nel dominio $d = ij$, per $i = 1, \dots, 20$ e $j = 1, \dots, 9$ e \hat{Y}_{ij} la corrispondente stima diretta. Allora si assume

“Sampling model”

$$\hat{Y}_{ij} | Y_{ij} \sim \text{Beta}(a_{ij}, b_{ij}), \quad \text{tale che } a_{ij} > 0 \text{ e } b_{ij} > 0$$

dipendono dal parametro da stimare e dagli elementi diagonali della matrice di covarianze delle stime dirette

“Linking model”

$$\eta_{ij} \sim \text{Ntrunc}(\mu_{ij}, \sigma^2),$$

dove

$$\eta_{ij} = \ln[Y_{ij} / (1 - Y_{ij})]$$

è la trasformata *logit* del parametro da stimare e la normale è troncata in modo da garantire la positività dei parametri della distribuzione Beta. Ancora si assume che

$$\mu_{ij} = \alpha_j + x_i \beta_j$$

con α_j intercetta e β_j coefficiente della regressione lineare del valore atteso μ_{ij} dalla informazione ausiliaria x_i data dal PIL pro-capite della regione i .

Infine si assegnano le seguenti distribuzioni a priori (diffuse) sui parametri di secondo livello

$$\alpha_j \sim N(0, 100), \beta_j \sim N(0, 100) \text{ e } \sigma^2 \sim \text{Unif}(0, 100)$$



Regione Emilia-Romagna



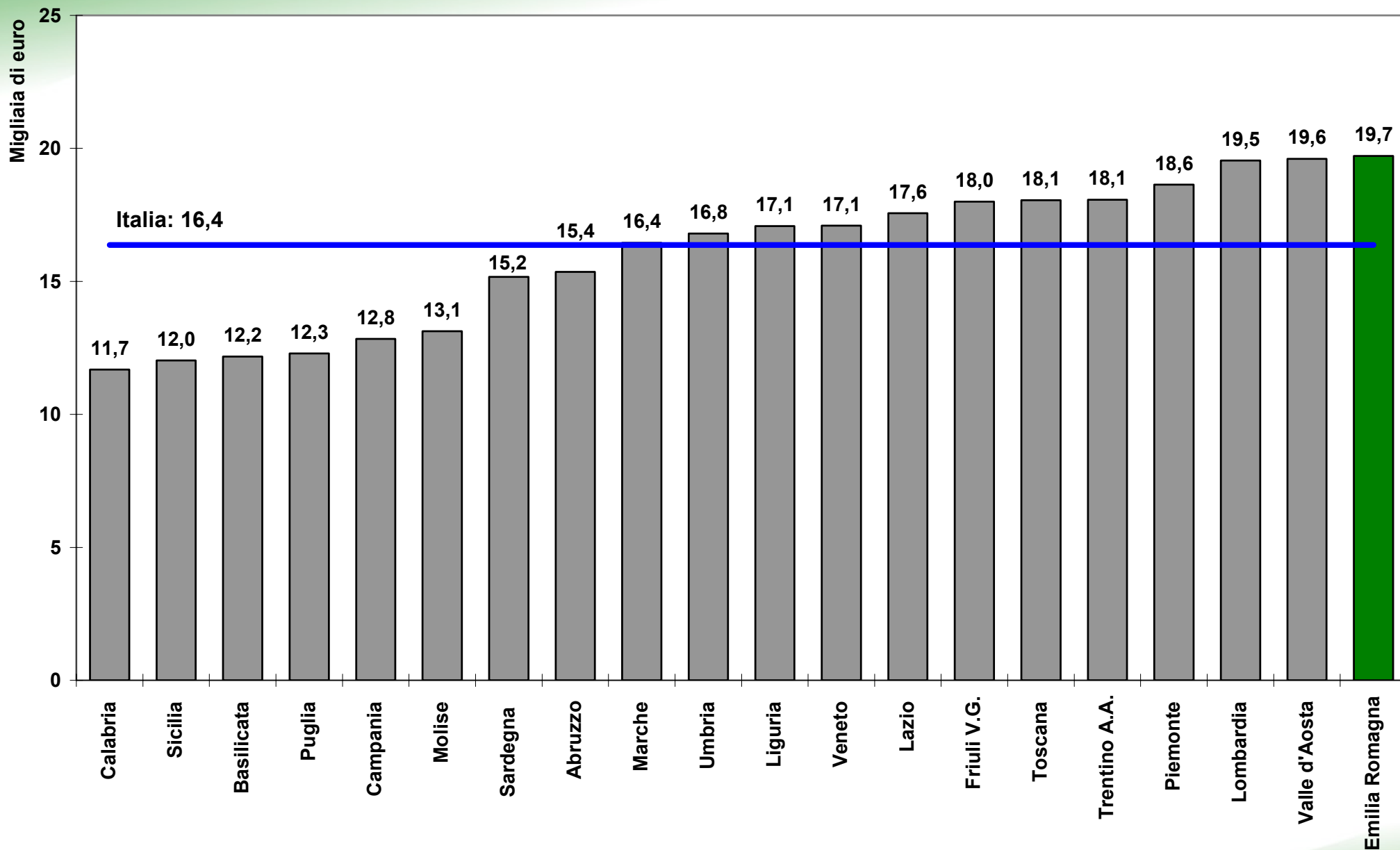
Servizio Controllo Strategico e Statistica

Analisi dei risultati di IT-SILC 2005 in Italia e per regione

REGIONE	Reddito equivalente medio	Reddito equivalente mediano	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	Rapporto interquartil.
Piemonte	18.634	16.111	5,3	11,3	17,1	8,2	15,5	23,2	0,19	0,22	30,6	4,6
Valle d'Aosta	19.611	16.620	3,8	5,7	11,9	5,0	8,9	19,5	0,22	0,23	30,4	4,0
Lombardia	19.546	16.957	4,7	9,6	16,9	7,9	16,4	25,0	0,19	0,19	32,2	4,7
Trentino A.A.	18.070	16.201	3,6	8,9	16,9	6,3	13,9	25,6	0,17	0,17	29,1	4,1
Veneto	17.100	15.216	5,0	10,8	20,1	6,5	13,5	23,9	0,18	0,20	28,4	4,0
Friuli-V. G.	17.999	16.298	4,1	10,2	17,8	6,9	16,5	25,3	0,17	0,18	27,6	4,1
Liguria	17.080	15.087	7,9	15,1	23,4	8,9	17,5	25,5	0,22	0,22	31,6	5,0
Emilia Romagna	19.719	17.814	4,6	9,1	14,5	9,1	16,3	25,7	0,20	0,22	30,2	4,4
Toscana	18.053	16.069	4,2	9,8	15,5	7,1	12,9	23,0	0,17	0,21	27,1	4,0
Umbria	16.798	14.907	6,8	14,7	23,5	7,9	17,0	26,1	0,18	0,17	29,0	4,4
Marche	16.448	15.046	6,3	13,2	21,5	7,7	15,7	24,6	0,19	0,19	27,7	4,2
Lazio	17.564	14.818	9,2	16,0	24,2	10,7	18,1	25,8	0,24	0,23	34,1	5,5
Abruzzo	15.357	13.525	9,3	18,4	29,2	8,5	15,8	25,5	0,20	0,22	29,8	4,3
Molise	13.125	11.175	13,5	28,5	44,6	8,6	13,1	24,0	0,17	0,33	30,3	4,5
Campania	12.841	10.540	21,0	31,8	47,8	11,4	17,8	24,7	0,29	0,31	35,4	6,3
Puglia	12.288	11.043	21,5	33,9	45,5	11,9	20,0	29,0	0,27	0,26	30,9	5,2
Basilicata	12.171	10.560	17,6	33,6	48,9	10,4	14,8	26,2	0,22	0,29	30,1	4,9
Calabria	11.681	10.067	23,3	39,4	51,3	12,9	17,3	27,9	0,24	0,38	33,1	6,0
Sicilia	12.027	9.759	24,7	40,0	53,3	12,5	18,9	27,5	0,28	0,30	36,0	6,6
Sardegna	15.173	13.403	12,7	20,4	30,2	10,7	18,0	26,7	0,25	0,27	31,8	5,2
ITALIA	16.368	14.240	10,9	19,1	24,3	--	--	--	0,24	--	32,0	4,6



Le regioni italiane ordinate per reddito equivalente medio

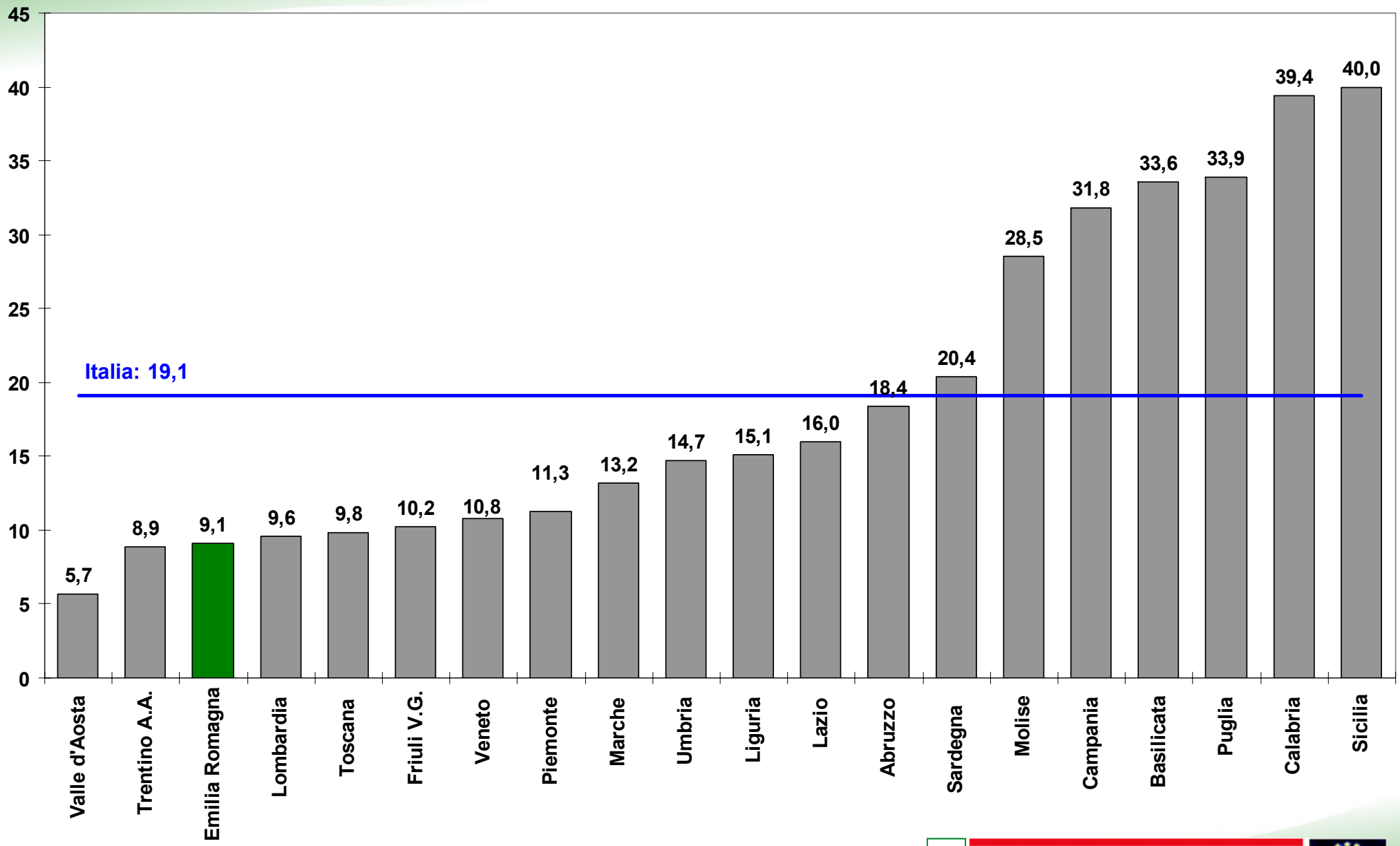


Regione Emilia-Romagna

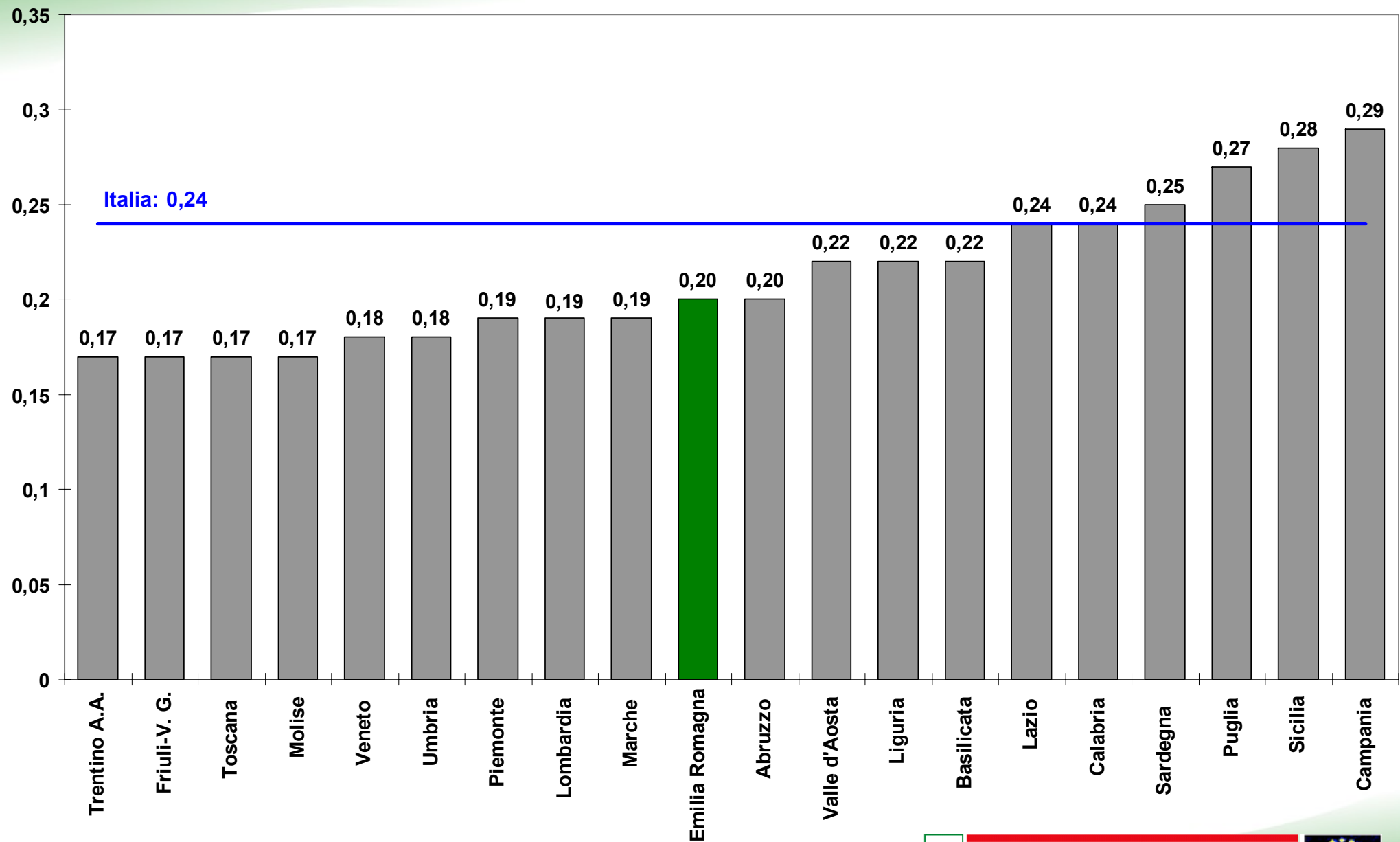


Servizio Controllo Strategico e Statistica

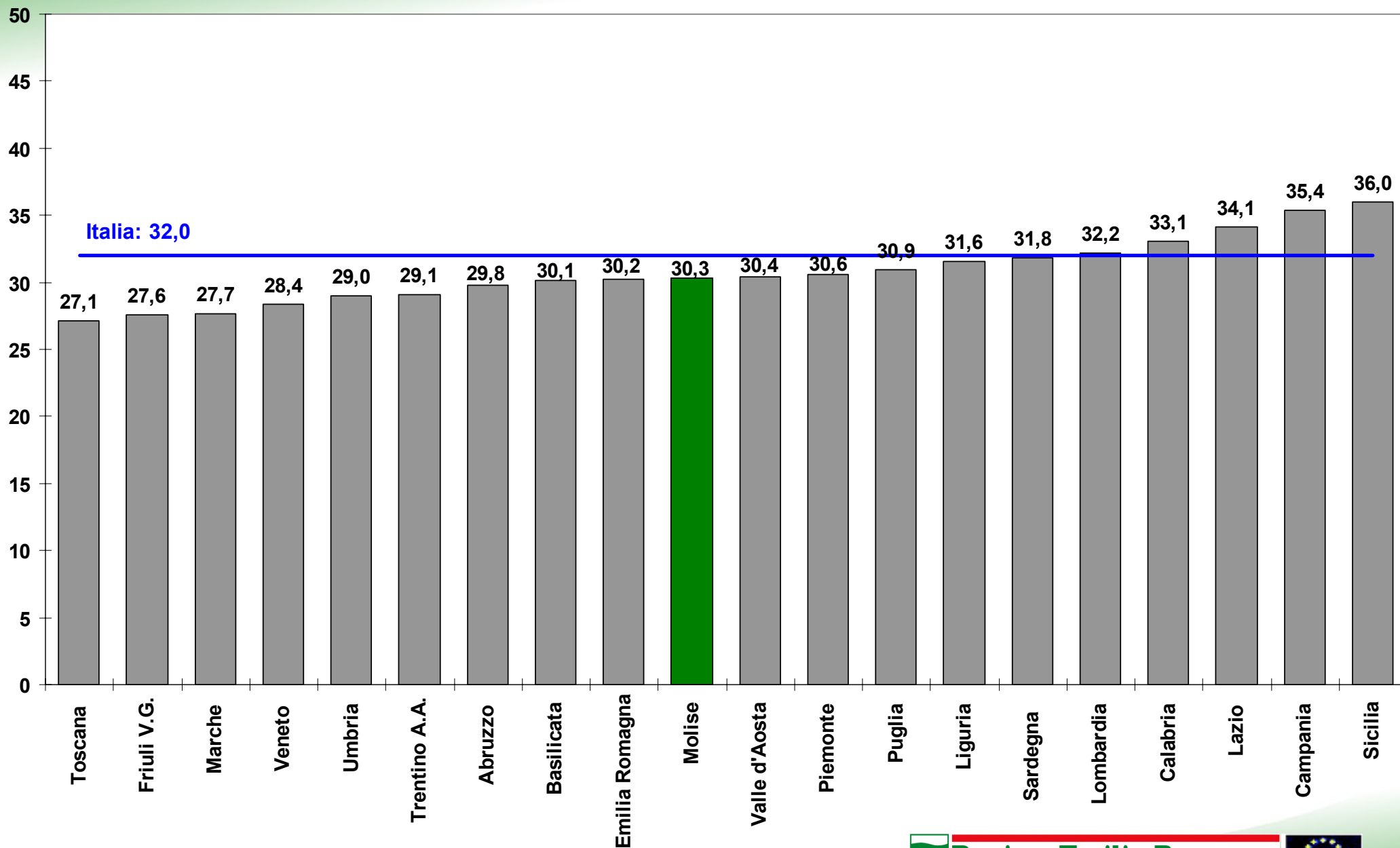
Le regioni italiane ordinate per tasso di povertà rispetto a SPN



Le regioni italiane ordinate per *poverty gap* rispetto a SPN



Le regioni italiane ordinate per indice di Gini



Regione Emilia-Romagna



Servizio Controllo Strategico e Statistica

Analisi dei risultati per tipologia familiare a livello nazionale

Tipologia familiare	Reddito equivalente medio	Reddito equivalente mediano	Tasso povertà severa, SN	Tasso povertà, SN	Tasso di rischio di povertà, SN	Gap mediano, SN	Indice di Gini	Rapporto interq.
A	14.436	12.246	14,2	28,1	37,1	0,33	34,1	6,1
B	20.300	18.185	5,5	9,7	16,8	0,31	30,5	4,9
C	15.158	12.763	7,0	19,5	32,0	0,20	29,6	4,3
D	19.355	17.369	5,5	9,4	15,1	0,29	27,8	4,3
E	11.475	10.490	27,0	37,0	47,2	0,46	36,0	9,1
F	17.269	15.396	9,6	15,0	23,7	0,36	30,8	5,2
G	15.033	13.179	12,9	21,7	33,6	0,31	31,4	5,2
H	12.685	10.722	24,9	36,0	47,5	0,35	35,7	6,9
I	15.155	13.461	11,2	20,4	31,4	0,29	30,5	5,0

Legenda tipologie famigliari:

A: Monocomponente;

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.

Analisi dei risultati per tipologia familiare in Emilia-Romagna

Tipologia familiare	Reddito pro capite	Reddito mediano pro capite	Tasso povertà severa, SN	Tasso povertà, SN	Tasso di rischio di povertà, SN	Gap mediano, SN	Indice di Gini	Rapporto interq.	Tasso povertà severa, SR	Tasso povertà SR	Tasso di rischio di povertà, SR	Gap mediano, SR
A	15.649	13.711	10,2	21,2	28,2	0,19	30,4	4,9	20,6	30,4	43,8	0,27
B	23.580	20.758	2,8	5,1	9,1	0,20	30,3	4,8	3,5	9,7	14,6	0,15
C	16.782	14.021	3,4	10,8	20,1	0,11	28,0	3,9	9,6	21,8	38,8	0,18
D	22.288	19.858	1,3	2,7	5,7	0,09	23,6	3,3	2,3	5,4	9,0	0,18
E	12.857	11.108	15,7	24,5	36,1	0,32	28,4	4,3	22,2	39,5	52,2	0,26
F	20.114	18.256	3,9	5,6	12,8	0,20	29,8	5,1	4,7	12,6	20,9	0,19
G	18.728	16.856	3,5	6,9	14,1	0,23	25,3	3,6	6,6	14,8	26,4	0,15
H	21.007	17.998	7,1	14,9	19,9	0,10	27,1	3,9	12,3	17,5	41,3	0,16
I	21.350	20.236	1,4	3,2	3,8	0,09	22,7	3,0	3,0	6,7	19,7	0,11

Legenda tipologie famigliari:

A: Monocomponente;

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;

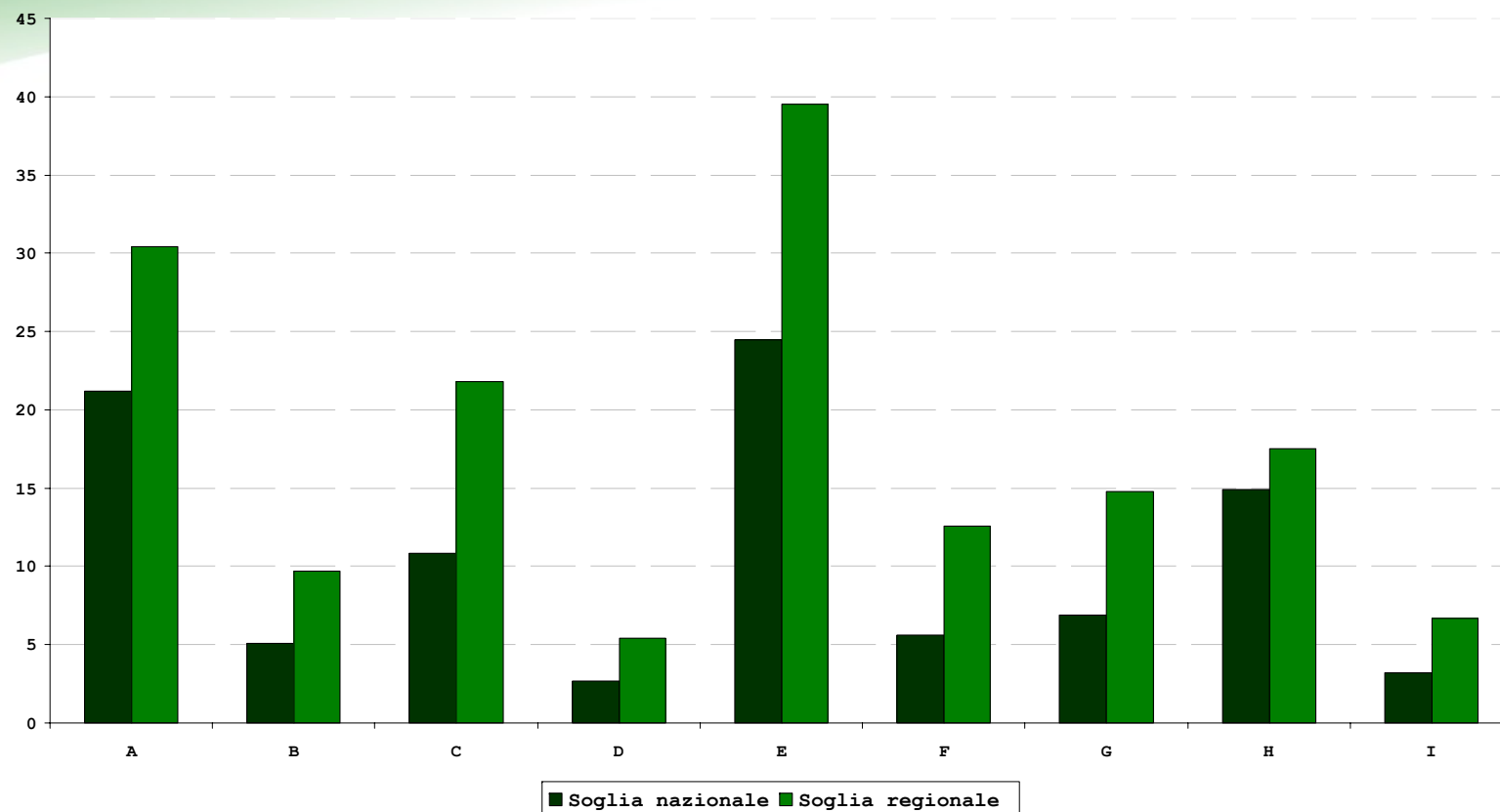
G: 2 adulti, 2 figli dipendenti;

H: 2 adulti, 3 o più figli dipendenti;

I: Altra tipologia con figli dipendenti.



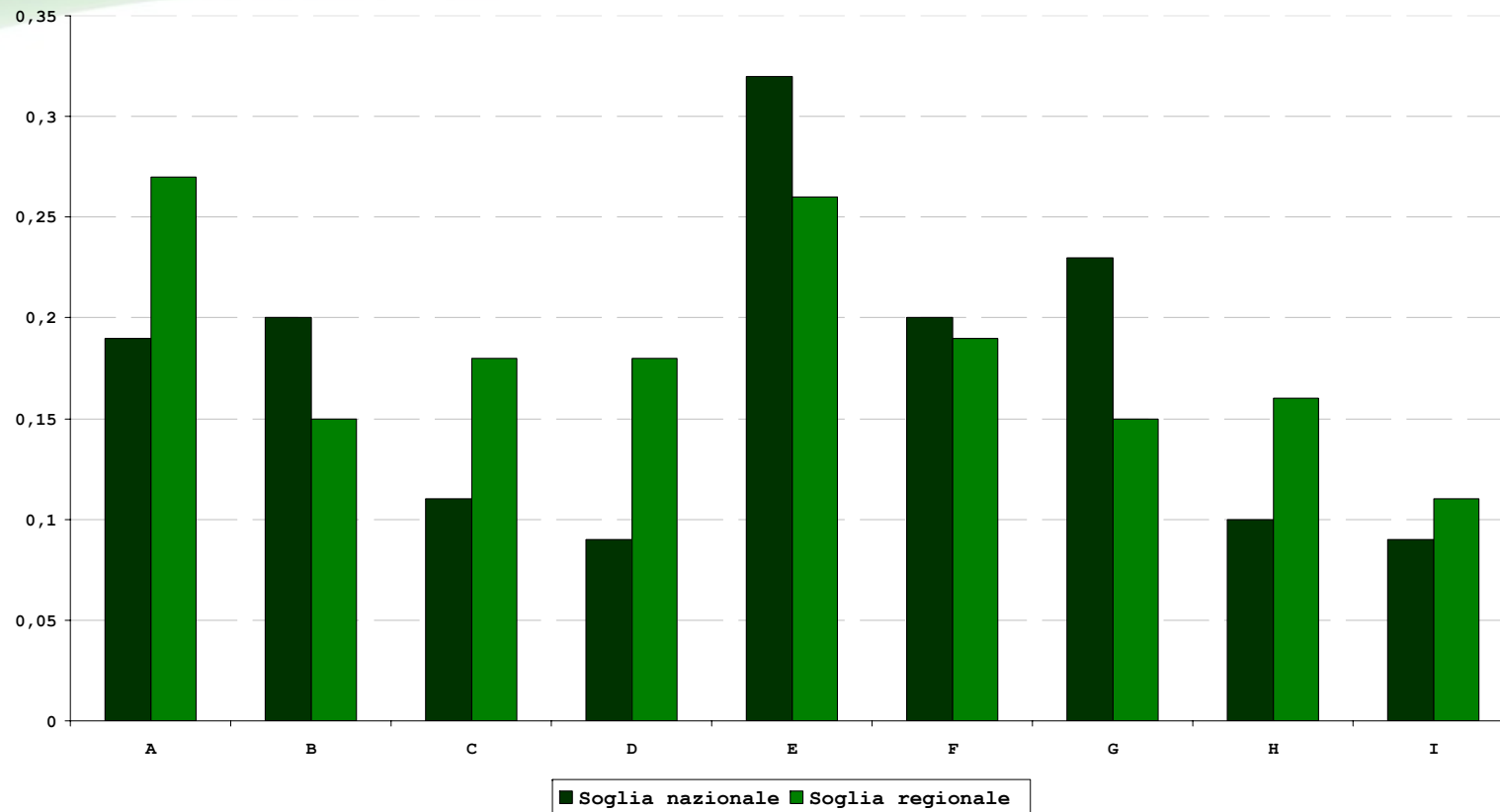
Tassi di povertà per tipologia familiare in ER (rispetto a SPN e SPR)



Legenda tipologie familiari:

- A: Monocomponente;
- 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.

Gap mediani per tipologia familiare in ER (rispetto a SPN e SPR)



Legenda tipologie familiari:

- A: Monocomponente;
- 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.

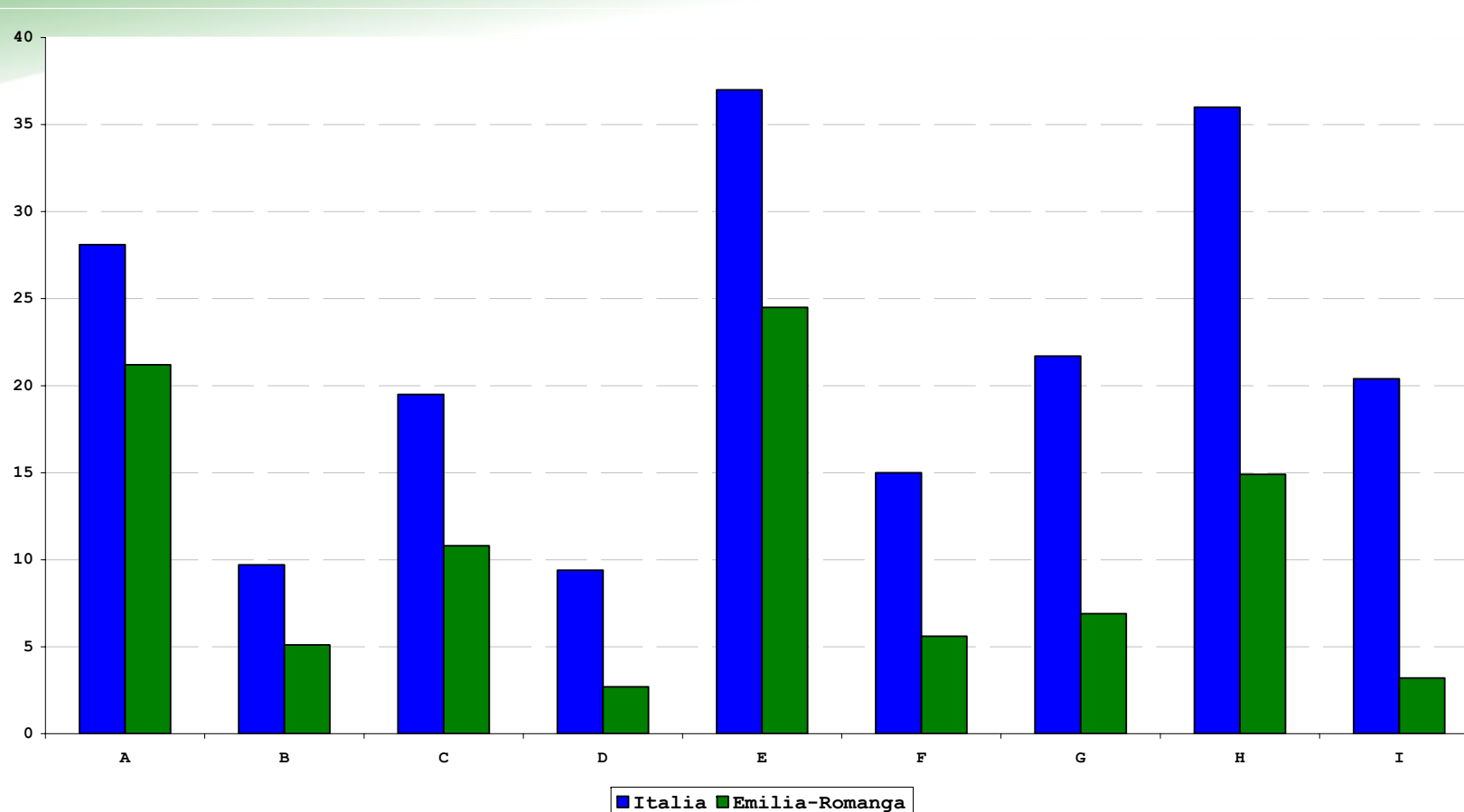


Regione Emilia-Romagna



Servizio Controllo Strategico e Statistica

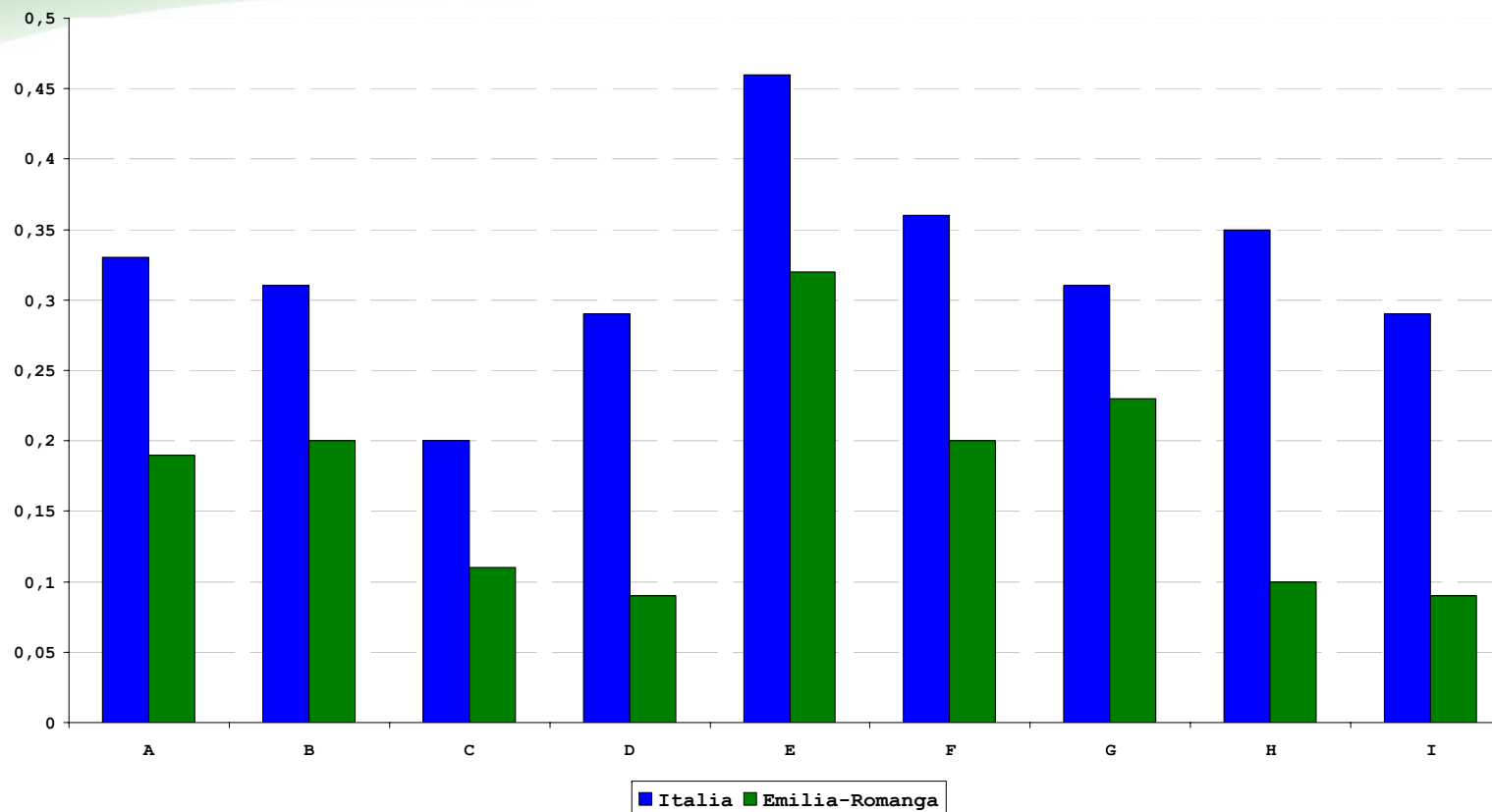
Tassi di povertà per tipologia familiare in Italia e in ER



Legenda tipologie familiari:

- A: Monocomponente;
- 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.

Gap mediani per tipologia familiare in Italia e ER



Legenda tipologie familiari:

- A: Monocomponente;
- 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.

GRAZIE

<http://www.regione.emilia-romagna.it/wcm/statistica/publicazioni/publicazioni/reddito.htm>