

2023

Individuazione di gruppi omogenei di assistiti Caritas con l'analisi multivariata dei dati





Individuazione di gruppi omogenei di assistiti Caritas con l'analisi multivariata dei dati

Sommario

1. Obiettivi del lavoro	3
2. Metodologia	3
2.1. Panoramica.....	4
2.2. Attività di pre-elaborazione preparatorie	6
2.3. Realizzazione	9
3. Analisi dei risultati	18
3.1. I profili dei gruppi	18
3.2. Incroci dei gruppi con le variabili principali.....	24
3.3. Commento ai risultati.....	34



1. Obiettivi del lavoro

La classificazione dei beneficiari in categorie omogenee può aiutare i decisori a misurare e a mappare i benefici apportati da un'organizzazione e a comprendere meglio la relazione tra i beneficiari stessi e il *contesto*, intendendo per contesto sia quello esterno (economia, demografia, fenomeni sociali) che quello che attiene all'organizzazione stessa, ai servizi resi, ai punti di erogazione di tali servizi.

Ciò è vero anche nel caso di Caritas, i cui beneficiari sono superiori alle 250mila persone, con una rete di assistenza ramificata e diffusa sul territorio che non uguali sul territorio nazionale. La capillarità della rete di punti di erogazione della Caritas fa sì che Caritas intercetti tutte le particolarità e le sfumature degli ambiti socio-economici nei quali è presente, e di conseguenza anche la platea delle persone assistite è multiforme e necessita di essere anzitutto compresa.

Dall'esigenza di approfondire tale conoscenza è nato questo lavoro di analisi multivariata dei dati, con lo scopo di far emergere un certo numero di tipologie omogenee principali di assistiti.

Questo report descrive le metodologie utilizzate per realizzare la classificazione e i risultati ottenuti.

2. Metodologia

Il database di partenza utilizzato per realizzare la classificazione degli assistiti Caritas 2022 è ottenuto attraverso l'imputazione dei dati delle Schede Utente, compilate dagli operatori dei Centri di Ascolto all'atto della presa in carico della persona. Attraverso la Scheda Utente vengono rilevati i dati anagrafici della persona, quelli del nucleo familiare se presente, ma anche dati relativi alla condizione professionale, allo status occupazionale, al permesso di soggiorno e alla situazione alloggiativa. Vengono poi rilevati i bisogni in base a undici macro-voci, a loro volta suddivise in micro-voci di bisogno. Nella parte finale della Scheda vengono infine annotate le Richieste di



intervento, effettuate dal soggetto sotto la guida degli operatori, nonché gli Interventi realizzati.

Nel complesso il Db iniziale includeva:

- n° 26 variabili nominali (o categoriali), delle quali 9 *booleane* (sì/no o presenza/assenza)
- n° 1 variabile ordinale
- n° 7 variabili cardinali, o di scala, autonome
- n° 1 variabile multi-item non collegata a bisogni-richieste-interventi
- n° 116 variabili relative a micro-voci di bisogno, raggruppate in 11 macro-voci di bisogno, di norma cardinali discrete (in quanto conteggi di occorrenze di bisogno)¹
- n° 93 variabili relative a micro-voci di intervento, raggruppate in 12 macro-voci di intervento, di norma cardinali discrete (in quanto conteggi degli interventi erogati)²
- n° 93 variabili relative a micro-voci di richieste, raggruppate in 12 macro-voci di richiesta, di norma cardinali discrete (in quanto conteggi delle richieste effettuate)³.

Si tratta di una ingente mole di dati, potenzialmente ricchissima e unica nel campo dell'assistenza alle persone bisognose in Italia.

2.1. Panoramica

Per perseguire l'obiettivo di classificare le persone assistite in categorie omogenee è stato scelto un approccio di analisi multivariata dei dati.

¹ Di norma, in quanto in alcune zone del Paese, laddove i CdA Caritas utilizzano altri applicativi per l'imputazione dei dati, tali variabili possono essere di tipo booleano.

² Anche per quanto concerne gli interventi i dati relativi ad alcune zone del Paese riportano la fruizione o meno degli interventi (presenza/assenza)

³ I dati relativi alle richieste effettuate sono speculari a quelli sugli interventi, per cui anche in questo caso si registra una commistione tra conteggi e indicatori booleani.



La classificazione in gruppi omogenei viene realizzata normalmente attraverso procedure di aggregazione e classificazione che ricadono nella categoria delle tecniche di *clustering*. Nell'ambito delle tecniche di *clustering* possono essere annoverate tecniche applicabili a soli dati quantitativi, tecniche applicabili a soli dati categoriali o tecniche miste. Nel caso in oggetto le tecniche utilizzabili sul formato originario del database sono quelle per sole variabili categoriali (tra le quali la principale è la tecnica *Classification Tree*, -in italiano *Alberi di Risposta*-) o quelle che trattano situazioni miste (p.e. la metodologia *Two Steps*). Vista la preponderanza delle variabili categoriali nel database Caritas, ci si dovrebbe orientare decisamente verso la metodologia *Classification Tree*. Si evidenzia però che per le caratteristiche peculiari di tale metodologia essa produce infine una classificazione basata su poche variabili del dataset, solo quelle che spiegano la maggior quantità della devianza totale dei dati. Per esempio una classificazione in otto gruppi, che di per sé è già ampia, può coinvolgere al massimo sette variabili, mentre tutte le altre rimangono alla fine inutilizzate. Si tratta di un esito decisamente parziale e che non valorizza un database così ricco.

In questi casi è d'uso utilizzare piuttosto una procedura in due fasi, così articolata:

- A. Una prima fase di ***individuazione di dimensioni latenti***, soggiacenti a tutte le variabili in gioco, attraverso una delle tecniche di analisi multivariata della scuola di Benzecrì
- B. Una seconda fase di ***effettivo clustering delle unità statistiche***, realizzata attraverso i punteggi che ogni unità statistica (la singola persona assistita) assume su un numero significativo delle dimensioni principali individuate nella fase A.

Tale procedura in due fasi valorizza tutto il database, consentendo di ottenere una classificazione multidimensionale. Inoltre, dato che i punteggi delle dimensioni principali sono numerici, è possibile infine applicare le tecniche di *clustering* per variabili quantitative.

Ciò premesso, prima dell'applicazione della procedura descritta, sono state svolte una serie di attività preparatorie, descritte nel prossimo paragrafo, per creare nuove variabili



o trasformare quelle già disponibili nel modo più appropriato per il loro uso nella procedura in due fasi.

2.2. Attività di pre-elaborazione preparatorie

La prima attività di pre-elaborazione svolta è stata un'analisi esplorativa dei dati. Nel corso dell'analisi esplorativa sono stati anzitutto calcolati gli indicatori sintetici di ogni variabile (*media, mediana o moda*) e le rispettive distribuzioni di frequenza.

Le distribuzioni di frequenza hanno rivelato un'ampia incidenza di valori non rilevati per diverse variabili del database, in molti casi per intere sezioni del questionario. Alcune variabili sono risultate di fatto non utilizzabili per analisi dati di tipo multivariato, proprio a causa della forte incidenza dei valori mancanti.

Stabilita nel 40% la soglia indicativa massima accettabile di valori mancanti, sono state pre-selezionate un gruppo di variabili del database sulle quali costruire la procedura di raggruppamento e classificazione delle tipologie di utenza. In virtù dell'importanza di due variabili inizialmente escluse, quali l'istruzione e la condizione professionale, è stato deciso di verificare se tali variabili potessero essere recuperate, stimando una parte dei loro valori mancanti sulla base delle correlazioni con altre variabili del database.

Effettivamente sono state individuate alcune variabili correlate con le due in oggetto, cosicché si è proceduto:

- i. alla stima dei dati mancanti sul livello di istruzione utilizzando una combinazione delle classi di età con i gruppi di nazionalità più ricorrenti.
- ii. per quanto concerne la situazione occupazionale, invece, si è proceduto alla stima dei dati mancanti sulla base dell'incrocio tra classi di età e bisogno di occupazione espresso o NON-espresso.

Varie modalità di diverse variabili selezionate per l'analisi multivariata sono state aggregate tra loro, in modo che ciascuna modalità aggregata avesse un'incidenza sul



totale non inferiore al 15%⁴. Sono state tuttavia ammesse alcune eccezioni riguardanti modalità di particolare interesse, come p.e. gli under 18.

Alcune variabili consecutive della Scheda Utente sono state combinate tra loro per ottenere un maggior dettaglio:

- La cittadinanza è stata combinata con la nazione ottenendo, con successive aggregazioni, cinque classi (*Italiana, Nordafricana, Africa Occidentale, Ucraina, Resto del mondo*)
- La dimora con la condizione alloggiativa, (*Casa di proprietà, Affitto/subaffitto, Ospitato o alloggio gratuito, Senza dimora, Alloggio non specificato*).

Inoltre sono state create nuove variabili, tra le quali:

- La Durata dell'assistenza (a partire dall'anno di ingresso nel sistema di assistenza Caritas)
- L'Intensità di Bisogno (a partire dal conteggio delle micro-voci di bisogno rilevate)
- L'Intensità di Richiesta e l'Intensità degli Interventi fruiti
- Tre variabili di multidimensionalità di bisogni – richieste – interventi, calcolate conteggiando gli ambiti (macro-voci) attivati, a prescindere dal numero di micro-voci interessate per ogni ambito
- Una variabile di profilo dei bisogni (ottenuta attraverso una *cluster analysis* applicata alle macro-voci di bisogno)
- Una variabile di profilo delle richieste (ottenuta attraverso una *cluster analysis* applicata alle macro-voci di richiesta)
- Una variabile di profilo degli interventi (ottenuta attraverso una *cluster analysis* applicata alle macro-voci di intervento).

E' stata infine svolta un'analisi di correlazione tra tutte le variabili ottenute, calcolando l'indicatore di connessione V di Cramer per ogni coppia di variabili.

⁴ La presenza di categorie "rare" può destabilizzare le analisi multivariate, per cui le aggregazioni preventive rendono più stabili i risultati ottenuti.

Nella matrice sottostante sono stati valorizzati solo gli incroci con una V di Cramer superiore a 20 (in giallo-rosso quelli superiori a 30). Per quanto concerne Istruzione e Condizione Professionale i valori riportati sono quelli precedenti alla stima dei valori mancanti.

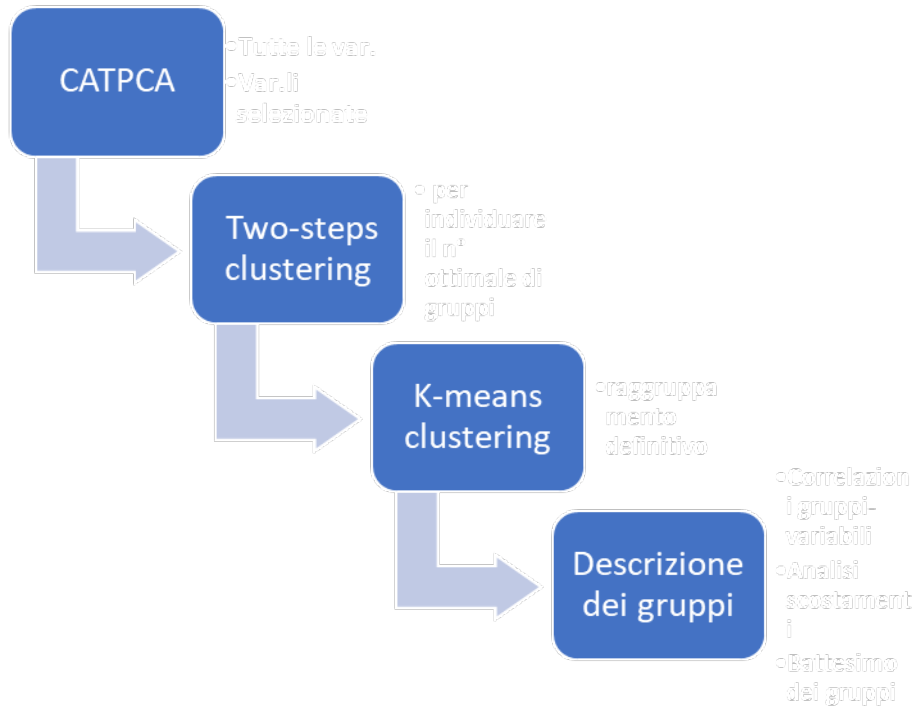
Variabile trasformata - descrizione	Variabile trasformata - database	livelloterritoriale_cod	Diocesi_class	MACRO_5_reg_civile	Macro_3_reg_civile	Macroaree_5_reg_ecc	Macroaree_3_reg_ecc	Sesso_cod	StatoCivile_aggr	Istruz_aggr	Nazionalit3	Abitativa	Sit_lavor	Con_chi_aggr	Ha_figli_cod	classi_eta	Num_comp_aggr	CI_BIS_stim	CI_INT	CI_RIC	Intens_BIS	Intens_INT	Intens_RIC	Multidim_BISOGNI	Multidim_INTERVENTI	Multidim_RICHIESTE	BIS_OCC_si_no	BIS_POV_si_no	Anno_apertura_scheda_classi	
livelloterritoriale	livelloterritoriale_cod	39	23																		25									
diocesi	Diocesi_class		35	29	35	29	24										22				24					20			22	
regione geografica	MACRO_5_reg_civile																													
	Macro_3_reg_civile																													
regione ecclesiastica	Macroaree_5_reg_ecc																													
	Macroaree_3_reg_ecc																													
genere	Sesso_cod								22		25	24		25	24		26													
stato civile	StatoCivile_aggr													32	47		22	28												
Istruzione	Istruz_aggr																													
cittadinanza & nazione	Nazionalit3											20					20													
dimora & cond. Alloggiativa	Abitativa														27	23		20												
Condizione professionale	Sit_lavor																26											26		
Con chi vive	Con_chi_aggr																													
Ha figli?	Ha_figli_cod																													
Classi di età	classi_eta																													
Numero componenti	Num_comp_aggr																													
Cluster dei Bisogni	CI_BIS_stim																													
Cluster Interventi fruiti	CI_INT																					72	65	23		82	31		20	25
Cluster Servizi Richiesti	CI_RIC																						27	35		35	48		24	
Intensità dei bisogni	Intens_BIS																													
Intensità di assistenza fruita	Intens_INT																													
Intensità di assistenza richiesta	Intens_RIC																													
Multidimensionalità dei bisogni	Multidim_BISOGNI																													
Multidimensionalità dell'assistenza fruita	Multidim_INTERVENTI																													
Multidimensionalità delle richieste	Multidim_RICHIESTE																													
Bisogno Occupativo	BIS_OCC_si_no																											25	35	
Bisogno Povertà	BIS_POV_si_no																													39
Anzianità di assistenza	Anno_apertura_scheda_classi																													24

Tabella delle connessioni tra le variabili selezionate per l'analisi multivariata.

Al netto delle correlazioni attese tra variabili che descrivono lo stesso aspetto si nota un'elevata correlazione tra *numero di componenti del nucleo* con *chi vive* e *presenza di figli*, tra *stato civile* e *con chi vive*, tra *stato di civile* e *presenza di figli*, tra *tipologia di diocesi – livello dimensionale e ripartizione territoriale Istat*, tra *richieste e interventi*, tra *bisogno di povertà e bisogno di occupazione*.

2.3. Realizzazione

Nel dettaglio la classificazione è stata realizzata secondo il seguente schema:



Schema delle procedure di analisi multivariata utilizzate.

dove:

- CATPCA è la procedura di **analisi in componenti principali categoriale**, attraverso la quale sono state individuate le dimensioni principali che “spiegano” la maggior parte della variabilità insita nei dati, passando da una rappresentazione con 26 variabili a una rappresentazione basata su solo 4 dimensioni
- La procedura di **clustering** denominata *Two-steps* è uno strumento esplorativo progettato per rivelare raggruppamenti naturali (i *cluster*) di unità statistiche all'interno di un set di dati. Una delle caratteristiche più importanti di questa procedura è l'individuazione “automatica” del numero di *cluster* ottimale. E in effetti la *Two-steps cluster* è stata utilizzata proprio per individuare preliminarmente un numero ottimale di *cluster* da utilizzare nella procedura *K-means*



- La procedura di clustering *K-means* mira invece a partizionare le unità statistiche in k cluster (k scelto dall'utente a priori), in cui ciascuna osservazione appartiene al cluster con il centroide più vicino. I centroidi utilizzati a tale scopo vengono selezionati dapprima casualmente e poi modificati via via in modo coerente con le nuvole di punti che si vanno formando, attraverso una procedura iterativa.

L'analisi in componenti principali categoriale

La procedura di riduzione delle dimensioni dei dati è stata applicata a sua volta in due step: a) nel primo sono state utilizzate tutte le variabili in gioco, mentre b) nel secondo il dataset è stato “ripulito” da alcune variabili-copia e dalle variabili irrilevanti ai fini dell'analisi.

Il primo lancio della procedura ha rivelato una elevata incidenza della variabilità spiegata dalle prime dimensioni principali:

- 51,2% della varianza spiegata dalle prime tre dimensioni e
- 62,2% della varianza spiegata dalle prime quattro dimensioni.

L'indice *Alpha di Cronbach* ha assunto un valore superiore a 0,7 (considerato in letteratura come valore soglia significativo) per le prime tre dimensioni, avvicinandosi molto alla soglia per la quarta. Ciò significa che il numero ottimale di dimensioni da mantenere per la descrizione dei dati Caritas 2022 è, appunto, compreso tra tre e quattro. In base ai risultati sulla varianza spiegata da ogni singola variabile, sono state eliminate:

- Le ripartizioni territoriali ecclesiastiche, sia a tre che a cinque ripartizioni, nonché la ripartizione geografica Istat a tre zone (Nord, Centro, Sud), in quanto variabili-proxi della ripartizione geografica Istat a cinque zone (Nord-Ovest, Nord-Est, Centro, Sud, Isole), più performante nella procedura CATPCA
- Le variabili di intensità dei bisogni, delle richieste e degli interventi, in quanto variabili-proxi delle rispettive multidimensionalità, più performanti nella procedura CATPCA.



- Livello di istruzione, Situazione occupazionale e Genere, in quanto hanno apportato un contributo alla spiegazione della varianza complessiva⁵ inferiore all'1%.

Al secondo lancio sono state pertanto utilizzate solo 17 variabili, con un aumento della capacità discriminante della procedura.

Infatti le prime quattro dimensioni spiegano ora quasi il 71% della varianza iniziale, che rappresenta un risultato molto positivo in analisi di questo tipo.

La tabella seguente riporta i contributi di ogni variabile alla spiegazione della varianza, per ogni dimensione e sul totale delle quattro dimensioni principali. I contributi sono espressi in valore assoluto all'interno delle dimensioni e in valore percentuale sul totale nella colonna di destra della tabella.

Le celle con colorazione verde indicano un contributo alto, quelle in rosso un contributo basso, le varie gradazioni intermedie i valori intermedi.

⁵ L'eliminazione di tali variabili ha riguardato solo la procedura CATPCA. In seguito tali variabili sono state recuperate nella fase di descrizione dei cluster.

Variabile	Dimensioni				Somma	%
	1	2	3	4		
Con chi vive aggregato*	0,118	0,659	0,122	0,019	0,918	8,0%
Ha figli?*	0,071	0,401	0,124	0,009	0,605	5,3%
Condizione Abitativa*	0,018	0,184	0,036	0,138	0,376	3,3%
Stato Civile aggregato*	0,032	0,279	0,053	0,008	0,372	3,3%
Livello Territoriale	0,015	0,054	0,272	0,326	0,667	5,8%
Tipo località diocesi	0,022	0,025	0,19	0,234	0,471	4,1%
MAcro_5_reg_civile	0,011	0,057	0,056	0,593	0,717	6,3%
Nazionalità aggregata 5 voci	0,015	0,011	0,077	0,04	0,143	1,3%
classi_eta	0,005	0,117	0,322	0,088	0,532	4,7%
Numero di componenti aggregato	0,129	0,667	0,131	0,022	0,949	8,3%
Cluster delle tipologie di Bisogno	0,037	0,177	0,202	0,183	0,599	5,2%
Cluster delle tipologie di Intervento	1,175	0,118	0,021	0,009	1,323	11,6%
Cluster delle tipologie di Richiesta	1,44	0,09	0,114	0,019	1,663	14,5%
Multidimensionalità dei bisogni	0,045	0,14	0,139	0,226	0,55	4,8%
Multidimensionalità degli interventi	0,339	0,142	0,145	0,031	0,657	5,7%
Multidimensionalità delle richieste	0,19	0,016	0,283	0,011	0,5	4,4%
Anno_apertura_scheda_classi	0,038	0,142	0,204	0,004	0,388	3,4%
Totale	3,685	3,27	2,415	1,92		
<i>varianza spiegata</i>	23,0%	20,4%	15,1%	12,0%		

Richieste/interventi
Profilo famiglia
[interrelazioni nascoste]
Territorio (e bisogni)

Quadro dei contributi di ogni variabile alla varianza spiegata, per dimensione e totale.

Si nota che:

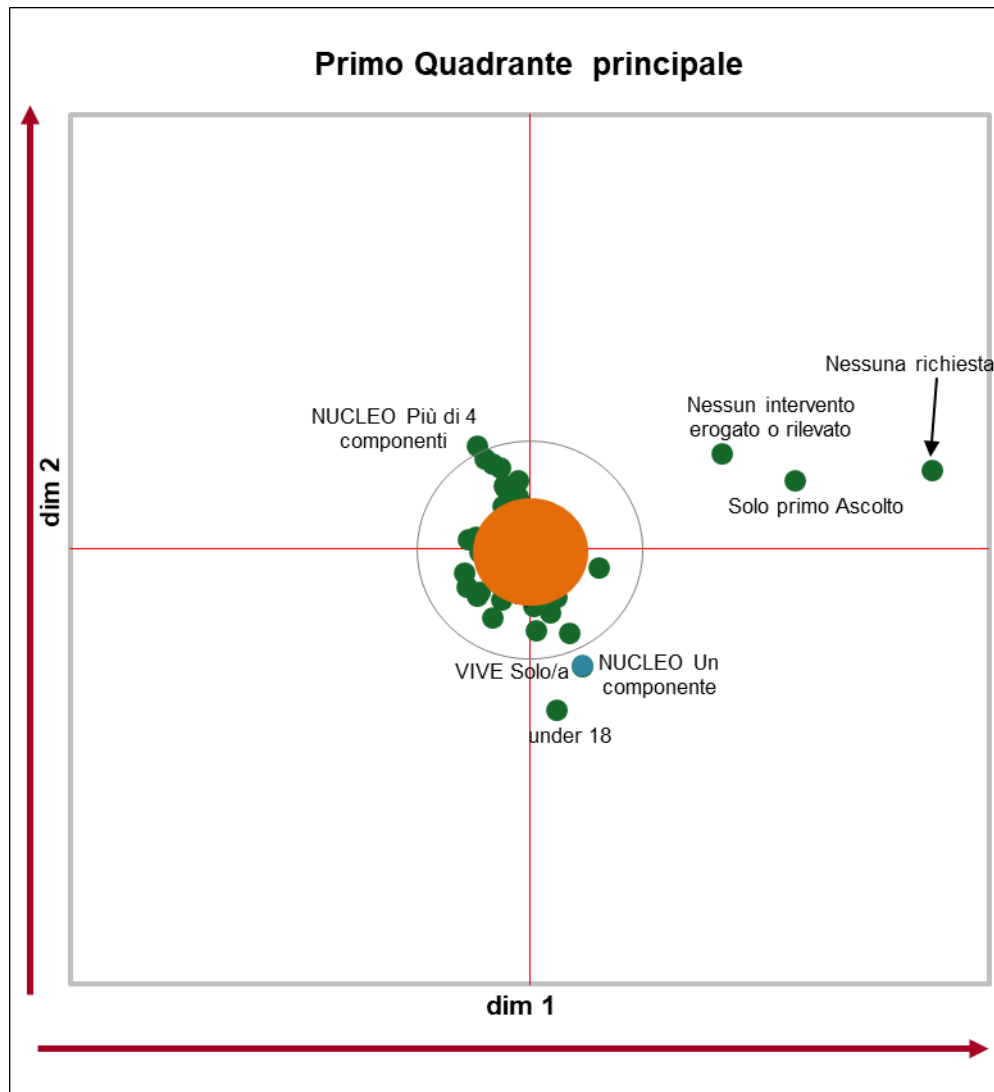
- Nel complesso delle quattro dimensioni, le variabili che sono più rilevanti per la spiegazione della varianza sono il *cluster di tipologia di richiesta*, il *cluster di tipologia di intervento*, il *numero di componenti del nucleo*, *con chi vive* la persona assistita.
- La prima dimensione è caratterizzata fortemente dalle tipologie di intervento e richiesta (che, ricordiamo, sono molto correlate tra loro)
- La seconda dimensione è spiegata soprattutto da variabili attinenti al profilo del nucleo associato alla persona (intendendo qui per nucleo anche l'eventuale condizione di *Solo*)



- La terza dimensione presenta contributi piuttosto schiacciati intorno alla media. Emergono sugli altri la classe di età, la multidimensionalità delle richieste, il livello territoriale... L'interpretazione è in questo caso meno agevole, la dimensione ha intercettato probabilmente una serie di interrelazioni articolate tra più variabili
- La quarta dimensione è invece molto caratterizzata da variabili di tipo territoriale e dimensionale, oltre che da multidimensionalità e tipologie di bisogno.

Emergono in definitiva le componenti che, come vedremo, caratterizzano i gruppi, ossia l'essere solo piuttosto che inserito in un nucleo familiare, i bisogni rilevati, le tipologie di intervento fruite, la distribuzione sul territorio nazionale.

Il grafico seguente illustra la posizione delle modalità più eccentriche del *primo piano principale* (ossia quello che incrocia le dimensioni 1 e 2).



Primo Piano Principale (con sovrapposizione del *cerchio di indifferenza* e del cerchio unitario)

Il cerchio arancione sovrapposto al grafico è il *cerchio di indifferenza* e lo scopo della sua presenza è di oscurare le modalità più vicine all'origine degli assi, che sono anche quelle meno o per niente rilevanti nelle prime due dimensioni individuate dalla CATPCA. Rimangono invece visibili i punti che individuano le modalità con distanza superiore a 0,5 dall'origine degli assi, anche se le etichette sono state apposte solo alle modalità esterne al cerchio di raggio 1.

Notiamo, nella parte destra, le modalità *Nessuna richiesta*, *Solo primo ascolto* e *Nessun intervento (...)*, a testimonianza del fatto che la prima dimensione principale ha isolato soprattutto le persone per le quali non sono stati registrati bisogni o interventi. Le altre modalità etichettate, che si muovono verso il basso o verso l'alto lungo la seconda



dimensione, sono tutte collegate alle caratteristiche del nucleo della persona. Nella parte inferiore si nota una correlazione svelata tra la categoria degli *under 18* e la condizione *Solo*.

Quadrante 4 (dim1 bassa - dim2 alta)	Quadrante 1 (dim1 alta - dim2 alta)
NUCLEO Più di 4 componenti	Nessuna richiesta
NUCLEO Tre o quattro componenti	Solo primo Ascolto
VIVE Con famiglia/partner	Nessun intervento erogato o rilevato
Coniugato/a	
BIS Solo Povertà	
FIGLI Sì	
Affitto/subaffitto	
INT Sussidi con Ascolto/ Beni	
NUCLEO Due componenti	
Quadrante 3 (dim1 bassa - dim2 bassa)	Quadrante 2 (dim1 alta - dim2 bassa)
Tre o + ambiti di bisogno	under 18
Tre o + ambiti di INT	VIVE Solo/a
Due ambiti di INT	NUCLEO Un componente
Tre o + ambiti di RIC	FIGLI No
INT Tipologie residuali	Senza dimora
Bisogni rari e combinazioni multiple	RIC non espresse o rilevate
	Celibe o nubile
	18-34 anni

Primo Piano Principale – riepilogo delle modalità poste al di fuori del cerchio unitario

Lo schema testuale delle modalità esterne al cerchio unitario completa quanto emerso dall'analisi del grafico. Notiamo che nel quadrante Uno non si sono aggiunte altre modalità, mentre gli altri quadranti presentano dalle sei alle nove modalità. Scorrendo l'elenco delle modalità che più caratterizzano i quattro quadranti si possono intuire già dei profili di *personas*, che però, va sottolineato, sono limitati alle prime due dimensioni. A volte ci sono più modalità della stessa variabile accumulate nello stesso quadrante, a rafforzare vicendevolmente la loro presenza. Per esempio, nel quadrante 4 hanno trovato collocazione sia le persone con nuclei molto numerosi, che quelle con 3-4 o 2 componenti. In definitiva: tutte le categorie della variabile Numero di componenti eccetto che l'opzione *Un componente*. In breve, si tratta delle persone NON-sole. E non a caso i Soli trovano posto nel quadrante opposto (il n° 2).



Two-step cluster

Come anticipato, la procedura di *clustering* denominata *two-steps* è stata utilizzata solo per avere un'idea solida sul numero di cluster da richiedere alla procedura *K-means*, considerata generalmente più performante.

La procedura *two-steps* applicata ai punteggi di ogni persona rispetto alle prime quattro dimensioni emerse dalla CATPCA ha individuato 11 cluster distinti, dei quali uno con meno di 3mila persone e complessivamente sette con meno di 25mila persone. Ricordando che la soglia di 25mila persone rappresenta circa il 10% degli assistiti totali, se si volessero mettere in campo gruppi con almeno il 10% di peso sul totale, solo quattro degli undici cluster individuati supererebbero la soglia. Gli altri dovrebbero essere aggregati tra loro per superarla. Oppure dovrebbero essere aggregati a uno dei quattro cluster più numerosi. Da queste considerazioni emerge quindi un numero minimo di **quattro** cluster (i quattro più popolati ai quali sarebbero aggregati uno o più dei cluster minori), fino a un massimo di **sette** (ottenuto sommando ai quattro più popolati ipotetiche aggregazioni a due a due dei cluster rimanenti).

Clustering K-means

A seguito delle considerazioni svolte è stata lanciata la procedura K-means, a partire da un obiettivo di 7 cluster, per poi passare a 6 e infine a 5. L'ipotesi a 5 cluster è stata infine accettata in quanto solida rispetto a tutti i parametri di diagnosi.

Per verificare la qualità della clusterizzazione sono state utilizzate sia le statistiche dell'Analisi della Varianza (ANOVA) che la numerosità dei cluster costruiti. Anticipato che il test ANOVA è stato sempre superato positivamente, assicurando così la buona qualità dei clusters in termini di capacità discriminante, le scelte sono state in realtà impiegate sulle numerosità dei cluster.

Infatti:

- Nella soluzione a sette cluster, ben tre di essi assommavano meno di 20mila persone. Uno in particolare appariva assolutamente di nicchia includendo solo 259 persone, ossia circa lo 0,1% del totale.



- Nella soluzione a sei cluster, due di essi includevano meno di 20mila persone
- Nella soluzione a cinque cluster solo uno dei cinque era popolato da meno di 15mila persone. Ritenendo questa soluzione accettabile si è proceduto a validare la soluzione a cinque cluster, elevando poi il numero di iterazioni richieste alla procedura da 13 a 25, onde ottenere una classificazione ancora più precisa⁶.

Le numerosità dei cinque cluster ottenuti sono, rispettivamente, di 52.711, 58.101, 17.282, 63.159 e 36.194 persone, per un totale di 227.447 persone classificate. Si evidenzia che circa 30mila persone sono state escluse dalla procedura a causa della mancanza di molti dati nel loro profilo.

⁶ Come già detto, la procedura *K-means* prevede l'aggiornamento iterativo dei centroidi, per cui stoppare le iterazioni troppo presto significa accettare una soluzione meno precisa.



3. Analisi dei risultati

3.1. I profili dei gruppi

Per individuare le caratteristiche dei cluster sono state calcolate tutte le distribuzioni di frequenza doppie tra ogni variabile del database iniziale di Ospoweb e l'appartenenza ai cinque cluster. Per le variabili ordinali sono state altresì calcolate alcune statistiche di posizione, come la mediana e i quartili, mentre per le variabili cardinali (numeriche) sono stati calcolati anche la media aritmetica, la mediana, i quartili, i percentili 5% e 95% e il minimo-massimo.

Sia per le variabili nominali che per quelle ordinali è stata calcolata anche la connessione bidirezionale tra la variabile in oggetto e l'appartenenza al cluster, utilizzando ancora una volta la V di Cramer. Per quanto concerne le variabili cardinali è stato invece calcolato l'indice di dipendenza in media Eta di Pearson.

Si evidenzia che per il calcolo delle distribuzioni doppie non sono state utilizzate tutte le persone classificate nei cinque cluster, bensì solo quelle *ben rappresentate* dal rispettivo cluster di appartenenza. È stata pertanto calcolata preliminarmente la distanza di ogni persona dal centroide del *cluster* (nell'iperspazio delle quattro dimensioni principali), provvedendo ad accantonare in questa fase tutte le persone con una distanza superiore a 1,5. L'incidenza delle 31.970 persone accantonate per le analisi descrittive dei profili è del 14,1% sul totale delle persone classificate.

Nella tabella seguente è riportata la graduatoria delle variabili più connesse al cluster di appartenenza, limitatamente a quelle nominali o ordinali.

Variabile	V di Cramer (su scala 0-100)
Con chi vive	91
Ha figli minori conviventi?	75
Cluster delle tipologie di Bisogno	66
Multidimensionalità dei Bisogni	64
Intensità di Bisogno	64
Ha figli?	64
Livello Territoriale	63
Ripartizione territoriale Istat	62
Classe dimensionale località diocesi	56
Convive con partner?	55
Condizione Abitativa	52
Stato Civile	52
---	...

Tabella delle connessioni superiori a 50 tra variabili del database Ospoweb e il cluster di appartenenza.

Tra le variabili cardinali, il cui grado di associazione con il cluster di appartenenza è stato misurato con l'*Eta di Pearson*, emerge il **Numero di Componenti (del nucleo)**, con un *Eta* pari a 58%.

Questi dati anticipano che i cluster sono caratterizzati soprattutto per la diversità degli appartenenti relativa alla variabile *Con chi vive*, che trascina con sé variabili ad essa molto correlate come il *Numero di figli*, la *Convivenza con un partner*, lo *Stato Civile*. Molto rilevanti anche tipologia e numerosità dei bisogni, così come alcune variabili demografiche (ripartizioni e classe dimensionale) o organizzative interne alla Caritas (il livello territoriale).



Profilo statistico dei cluster

Ciò premesso, la descrizione statistica dei gruppi è la seguente:

- A. **Cluster 1.** Adulti per lo più tra i 35 e i 60 anni, in prevalenza di genere maschile. Oltre la metà di loro sono celibi. Presentano una maggiore incidenza di divorziati. Molto spesso vivono soli, uno su tre al nord-est. Oltre uno su tre è senza dimora. Molto spesso presentano bisogni multipli (il 65% almeno tre tipologie, il 60% in almeno tre ambiti diversi), comprese diverse tipologie a bassa incidenza (casa 50%, problemi di immigrazione 22%, di salute 22%, problemi familiari 20%). E' significativamente più elevata la presenza di fragilità quali solitudine, abusi, maltrattamenti, prostituzione, bullismo. Si registra in questo gruppo un'incidenza significativamente superiore di interventi in ambito alloggio (8%) e socio-assistenziali (7% - soprattutto in termini di sostegno diurno socio-educativo). Cluster di assistiti in larga prevalenza da CdA diocesani.
- B. **Cluster 2.** Adulti per lo più tra i 35 e i 60 anni, due su tre sono di genere femminile, I due terzi sono coniugati e molto spesso hanno figli, spesso minori conviventi. Vivono quasi sempre con i propri familiari o in convivenze di fatto, in nuclei di 2-4 persone. Quasi il 60% di essi presenta solo bisogno di povertà. Oltre il 40% di essi è in carico a Caritas da almeno 5 anni. Quasi la metà di essi sono assistiti da CdA parrocchiali.
- C. **Cluster 3.** Giovani stranieri, di età media 25 anni, in grande maggioranza celibi e per la quasi totalità maschi. Oltre la metà di loro è di nazionalità africana. Gran parte delle persone di questo gruppo si sono accumulate al confine italo-francese nel tentativo di raggiungere altri paesi europei, trovando assistenza soprattutto nella diocesi di Ventimiglia, anche se diverse decine sono assistite da altre diocesi liguri. Tre persone su quattro, in questo gruppo, sono senza dimora. Nonostante ciò, molti di loro sono accompagnati da propri familiari o loro conoscenti. Quasi la metà di essi si dichiara studente. Presentano SEMPRE bisogni multipli (il 65% almeno tre tipologie, oltre il 60% in almeno tre ambiti diversi), comprese diverse tipologie a bassa incidenza. In particolare, il 62% di essi presenta problemi di

immigrazione, mentre il 60% lamenta la necessità di un alloggio e il 36% di istruzione. Tuttavia la quasi totalità di essi ha richiesto e fruito solo di interventi di erogazione di beni o servizi, anche se spesso di diversa tipologia. Sono tutti nuovi ascolti.

D. **Cluster 4.** Adulti per lo più tra i 35 e i 60 anni, due su tre sono di genere femminile. Il 60% di loro sono coniugati. Quasi sempre hanno figli, per lo più minori conviventi. Vivono quasi sempre con i propri familiari o in convivenze di fatto, in nuclei mediamente più numerosi rispetto agli altri gruppi. Quasi il 60% di loro vive in affitto, in generale le loro condizioni abitative sono mediamente migliori rispetto agli altri gruppi. E' un gruppo che caratterizza tutto il Sud e la fascia appennino-adriatica (Umbria, Marche, Emilia-Romagna, Veneto, Friuli V.G.). La nazionalità ucraina ha un'incidenza del 13% in questo gruppo, quasi il doppio della media nazionale, (il 18% tra le donne). Molto spesso presentano bisogni multipli (in oltre la metà dei casi in tre o più ambiti diversi), comprese diverse tipologie a bassa incidenza (problemi familiari 26%, di immigrazione 20%, di istruzione 14%). Due su tre esprimono un bisogno di occupazione. Alcune tipologie di fragilità, quali abusi, maltrattamenti, prostituzione, solitudine, bullismo, presentano un'incidenza superiore alla media generale. Tra gli interventi fruiti è significativamente più elevata l'incidenza di coinvolgimenti (5%) e di interventi in ambito istruzione (4%).

E. **Cluster 5.** Adulti per lo più tra i 35 e i 65 anni, di età media più alta rispetto agli altri cluster, in lieve prevalenza di genere maschile. Oltre la metà di loro sono celibi e quasi sempre senza figli. Presentano una maggiore incidenza di vedovi e di pensionati. Molto spesso vivono soli, due su tre al nord-ovest, i rimanenti al Centro (e soprattutto in Toscana e Lazio). Il 40% di essi vive in grandi città (>500mila abitanti). Quasi la metà di essi presenta solo bisogno di povertà. Si evidenzia un'incidenza doppia rispetto alla media di richiesta di assistenza socio-assistenziale (2%). Quasi la metà di essi sono assistiti da CdA parrocchiali.



Battesimo dei cluster

Sulla base dei profili è stata effettuato uno *brainstorming* per scegliere i nomi dei gruppi. Il nome del gruppo deve descrivere sinteticamente il gruppo utilizzando un basso numero di parole (non più di 3, al massimo 4).

Di seguito i nomi scelti per ogni gruppo:

Cluster 1. I soli vulnerabili

Cluster 2. Le famiglie povere

Cluster 3. I giovani stranieri in transito

Cluster 4. I genitori fragili

Cluster 5. I poveri soli.

Profilo accentuato dei cluster

In attività come quella descritta in questo report, laddove l'obiettivo delle elaborazioni è quello di ricavare dai dati una classificazione ben definita di gruppi omogenei di unità statistiche (in questo caso di persone assistite, in altri casi di clienti/utenti di beni e servizi), è necessario accentuare i contorni dei gruppi, così da renderli più netti e definiti. Questa è anche un'esigenza comunicativa, nel senso che il lettore deve essere accompagnato verso una comprensione nitida delle similarità e delle differenze che esistono, rispettivamente, tra unità appartenenti allo stesso cluster o, viceversa tra unità appartenenti a cluster diversi.

Ma non è solo un'esigenza comunicativa. In effetti la fotografia che ci restituiscono le procedure di analisi statistica multivariata è affetta da tutte quelle particolarità che i dati esaminati contengono intrinsecamente: i dati mancanti, gli errori di registrazione, le risposte imprecise... Ne deriva un certo grado di variabilità, a causa della quale una parte delle unità statistiche attribuite a un *cluster* presenta delle dissomiglianze, delle eccezioni, delle sfumature, che possono indurre confusione nel lettore.

Anche la scelta della soglia numerica dei cluster, in questo caso cinque piuttosto che 6-7-8..., implica alcune approssimazioni, dato che gli individui che in una soluzione a 8



cluster appartengono a gruppi diversi, si ritrovano invece ad appartenere insieme allo stesso gruppo in una soluzione a cinque *cluster*.

Per tutti questi motivi è d'uso procedere a una accentuazione dei contorni descrittivi dei cluster, ripulendo i gruppi dalle eccezioni, dalle mezze misure, dai “*ma...*” e dai “*tuttavia...*”. Questa ridefinizione dei profili conduce a quelli che vengono in genere definiti “*profili commerciali*” dei cluster. Nel nostro caso, non essendo “commerciale” lo scopo posto alla base dell'analisi svolta, è stato usato il termine di “*profilo accentuato*”.

- A. ***I SOLI VULNERABILI***. Adulti di genere maschile, per lo più celibi, vivono soli in Italia, e presentano una elevata incidenza di divorziati. Assistiti da CdA diocesani, uno su tre al Nord-Est, spesso sono senza dimora o comunque in situazione abitativa precaria. Presentano una molteplicità di bisogni, comprese diverse tipologie di bisogno a bassa incidenza quali casa, salute, problemi di immigrazione, problemi familiari. Evidenziano spesso fragilità quali solitudine, abusi, maltrattamenti, prostituzione, bullismo. In stretta correlazione con il loro profilo, queste persone hanno fruito più spesso degli altri di interventi in ambito alloggiativo e socio-assistenziale (soprattutto in termini di sostegno diurno socio-educativo).
- B. ***LE FAMIGLIE POVERE***. Donne adulte per lo più coniugate e molto spesso con figli minori conviventi. Vivono con i propri familiari o in convivenze di fatto, in nuclei di 2-4 persone. Presentano spesso il solo bisogno di povertà. Due su cinque sono in carico a Caritas da almeno 5 anni.
- C. ***I GIOVANI STRANIERI IN TRANSITO***. Giovani stranieri maschi, di età media 25 anni, per lo più celibi, uno su due è di nazionalità africana. Sono stati presi in carico tutti nel corso del 2022 (nuovi ascolti). Gran parte delle persone di questo gruppo si sono accumulate al confine italo-francese nel tentativo di raggiungere altri paesi europei, trovando assistenza soprattutto nella diocesi di Ventimiglia, anche se diverse decine sono assistite da altre diocesi liguri. Molto spesso sono senza dimora, anche se spesso in compagnia di propri familiari o

conoscenti. Presentano SEMPRE bisogni multipli, comprese diverse tipologie a bassa incidenza, con una elevata incidenza di problemi di immigrazione, necessità di alloggio e di istruzione. Tuttavia hanno fruito solo di interventi di erogazione di beni o servizi, anche se spesso di diversa tipologia.

D. I GENITORI FRAGILI. Genitori di figli minori conviventi, in prevalenza di genere femminile, per lo più coniugati. Vivono con i propri familiari o in convivenze di fatto, ma in nuclei mediamente più numerosi rispetto agli altri gruppi. Le loro condizioni abitative sono mediamente migliori rispetto agli altri gruppi. Tra le donne, una su cinque è di nazionalità ucraina. Sono particolarmente presenti in tutto il Sud e nelle regioni della fascia appennino-adriatica (Umbria, Marche, Emilia-Romagna, Veneto, Friuli V.G.). Presentano molto spesso bisogni multipli, tra i quali molto elevato il bisogno di occupazione, ma anche diverse tipologie di bisogno a bassa incidenza quali problemi familiari, di immigrazione, di istruzione, comprese diverse tipologie di fragilità, come abusi, maltrattamenti, prostituzione, solitudine, bullismo. Tra gli interventi fruiti sono più elevati rispetto alla media l'incidenza dei coinvolgimenti di altre organizzazioni assistenziali e gli interventi in ambito istruzione.

E. I POVERI SOLI. Adulti di genere maschile, ma di età media più alta rispetto agli altri cluster, vivono soli in Italia, e presentano una elevata incidenza di celibi, vedovi e pensionati. Sono presenti in prevalenza al nord-ovest o nelle regioni tirreniche del Centro, quasi la metà di essi in grandi città. Uno su due presenta solo bisogno di povertà. Richiedono più spesso degli altri un'assistenza di tipo socio-assistenziale.

3.2. Incroci dei gruppi con le variabili principali

In questa sezione si riportano alcuni incroci tra l'appartenenza al cluster e alcune variabili del database Ospoweb selezionate. Tra le variabili selezionate ci sono quelle più significative ai fini della spiegazione della varianza, oltre che le principali variabili

di struttura (genere ed età). In continuità con le scelte effettuate per la descrizione dei cluster sono state escluse le persone con distanza superiore a 1,5 dal centroide del proprio *cluster*.

Genere ed età

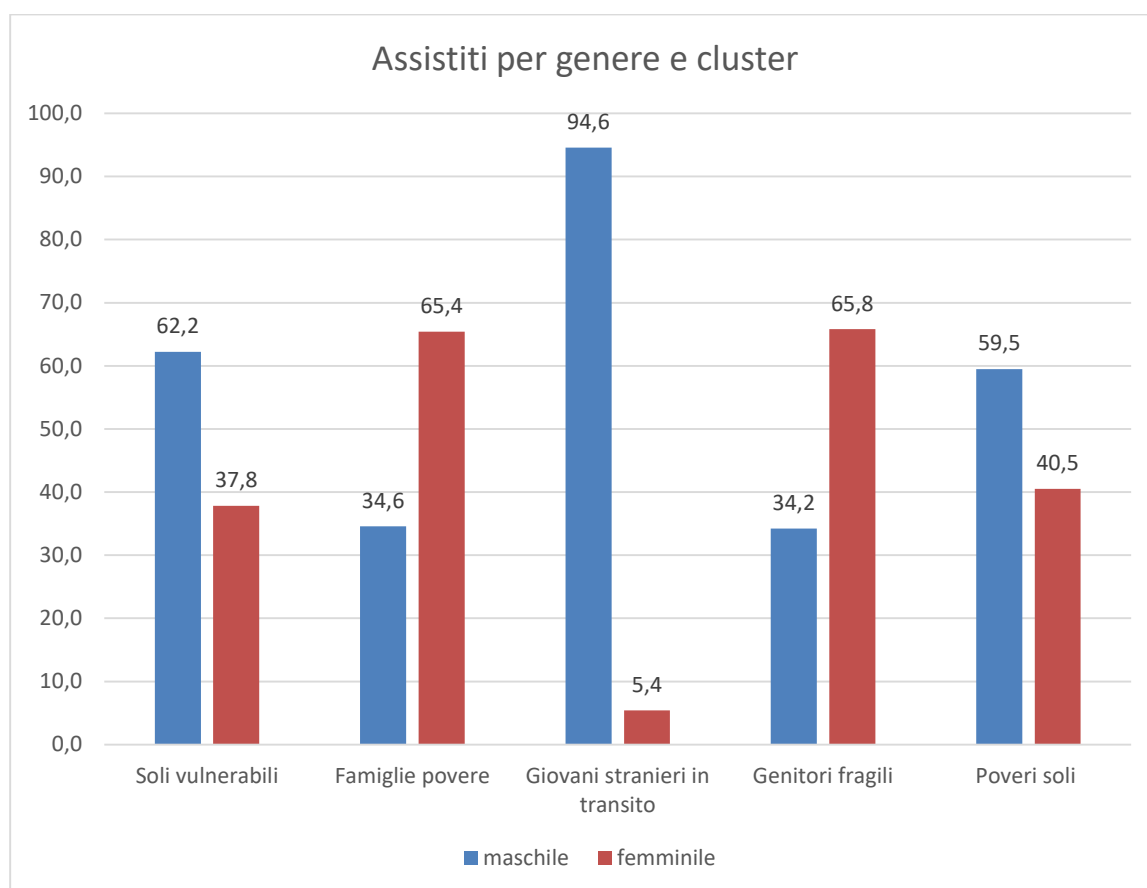
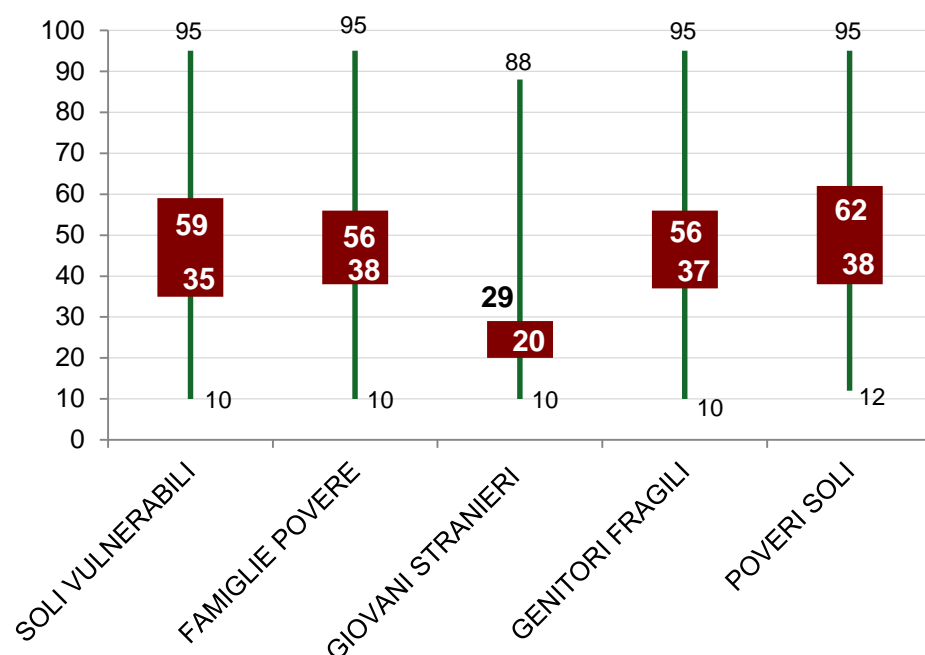


Grafico della distribuzione degli assistiti Caritas per genere all'interno di ogni cluster.

Il grafico evidenzia la netta prevalenza di maschi nel *cluster* GIOVANI STRANIERI, ma anche la larga prevalenza di femmine nei *cluster* FAMIGLIE POVERE e GENITORI FRAGILI.

Grafico della distribuzione degli assistiti Caritas per età all'interno di ogni cluster.

Distribuzione degli assistiti per età



La distribuzione per età è stata rappresentata attraverso grafici box-plot. Le scatole *bordeaux* racchiudono i valori compresi tra i due quartili⁷.

Con chi vive, figli minori, numero componenti

Assistiti per Con chi vive e Cluster

Con chi vive/Cluster	SOLI VULNERABILI	FAMIGLIE POVERE	GIOVANI STRANIERI	GENITORI FRAGILI	POVERI SOLI	Tutti
Solo	83,7%	0,2%	1,6%	0,0%	82,1%	26,3%
In nucleo con conoscenti o soggetti esterni alla propria famiglia	8,5%	5,1%	10,8%	7,6%	8,8%	7,2%
Presso istituto, comunità, ecc.	2,5%	0,5%	2,1%	1,0%	2,2%	1,3%

⁷ Il primo quartile individua la persona che lascia sotto di sé il 25% delle età più basse, mentre il terzo quartile individua la persona che lascia sopra di sé il 25% delle età più alte



Coabitazione di più famiglie	0,3%	0,3%	1,3%	1,0%	0,1%	0,5%
In famiglia di fatto (in nucleo con partner, con o senza figli)	0,6%	10,0%	6,0%	10,4%	0,5%	7,2%
In nucleo con familiari/parenti	2,8%	83,3%	77,2%	78,4%	5,3%	56,3%
Altro	1,6%	0,7%	1,0%	1,5%	1,1%	1,2%
TOTALE	100%	100%	100%	100%	100%	100%
VAL. PREVALENTE	Solo	In nucleo con familiari/parenti	In nucleo con familiari/parenti	In nucleo con familiari/parenti	Solo	In nucleo con familiari/parenti
VARIABILITA'	34,1%	32,0%	44,2%	40,1%	35,7%	61,2%

Tabella della distribuzione degli assistiti Caritas per *Con-chi-vive* all'interno di ogni cluster.

Come già emerso nella descrizione dei *cluster*, i SOLI VULNERABILI e i POVERI SOLI sono ad ampia prevalenza di persone che vivono sole. Negli altri *cluster*, compreso il GIOVANI STRANIERI, prevale invece l'opzione *In nucleo con familiari/parenti*. Si nota anche una certa incidenza delle famiglie di fatto (10% nei *cluster* FAMIGLIE POVERE e GENITORI FRAGILI) e dei conoscenti (11% nel *cluster* GIOVANI STRANIERI).

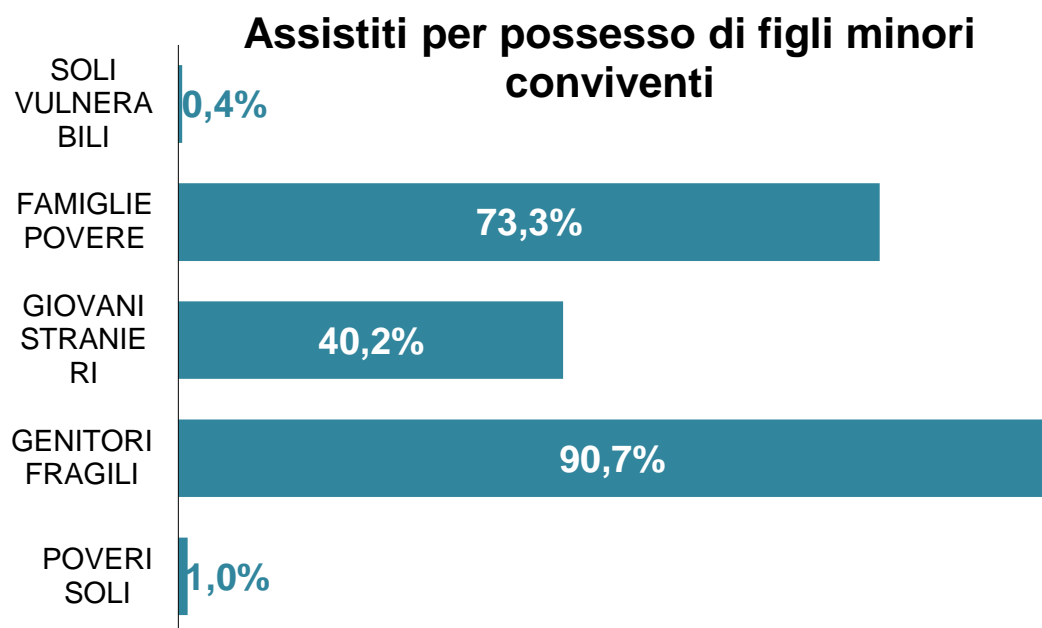


Grafico della distribuzione degli assistiti Caritas per possesso di figli minori conviventi, all'interno di ogni cluster.

Coerentemente alle caratteristiche dei profili, si registrano pochissimi casi di figli minori conviventi nei *cluster* SOLI VULNERABILI e POVERI SOLI. Viceversa l'incidenza è molto alta nel *cluster* GENITORI FRAGILI, e rilevante nel *cluster* FAMIGLIE POVERE. Da notare che anche nel *cluster* GIOVANI STRANIERI si rileva un'incidenza di figli minori al seguito superiore al 40%.

Distribuzione degli assistiti per numerosità del nucleo di convivenza

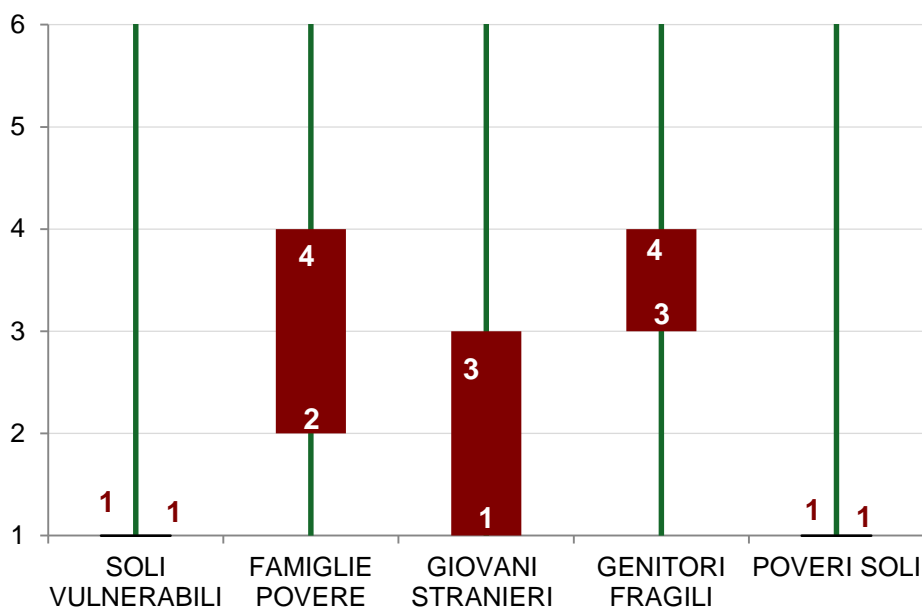


Grafico della distribuzione degli assistiti Caritas per numero di componenti del nucleo, all'interno di ogni cluster.

Anche il grafico per numero di componenti, nel quale è stata troncata la scala per motivi grafici, è coerente con le caratteristiche dei *cluster*. I nuclei mediamente più numerosi sono quelli dei *cluster* FAMIGLIE POVERE e GENITORI FRAGILI, mentre nei SOLI VULNERABILI e POVERI SOLI gran parte delle persone vivono, come già descritto,

da sole. I valori massimi, rispettivamente pari a 55, 96, 22, 111 e 96 sono associati a convivenze presso istituti e comunità.

Ripartizione territoriale

Assistiti per Ripartizione territoriale e Cluster

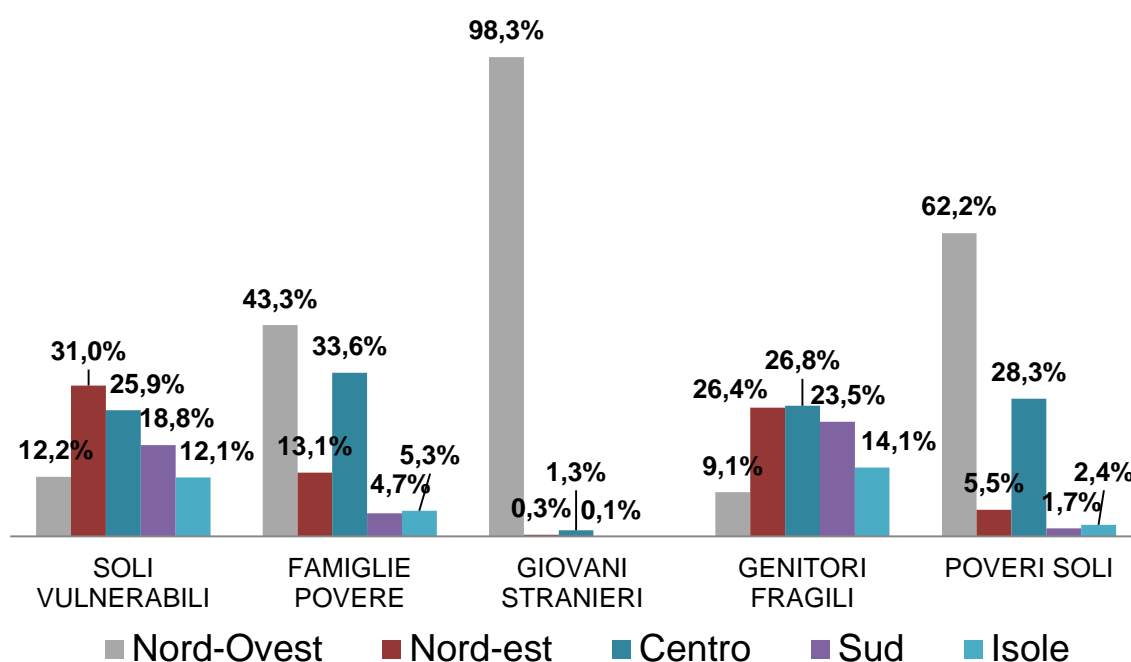


Grafico della distribuzione degli assistiti Caritas per ripartizione territoriale, all'interno di ogni cluster.

Gli assistiti del *cluster* GIOVANI STRANIERI vivono tutti all'interno della ripartizione Nord-Ovest, per la quasi totalità in Liguria. Anche il *cluster* POVERI SOLI (62%) e in misura minore il FAMIGLIE POVERE (43%), sono molto presenti al Nord-Ovest. In entrambi i casi è rilevante anche l'incidenza di coloro che vivono al Centro (risp.te: 28% e 34%). I SOLI VULNERABILI e i GENITORI FRAGILI sono invece dispersi nel Paese, con prevalenza al Nord-Est nel caso dei SOLI VULNERABILI, e al Centro nel caso del *cluster* GENITORI FRAGILI.

I bisogni

Assistiti per Tipologia di Bisogno e Cluster

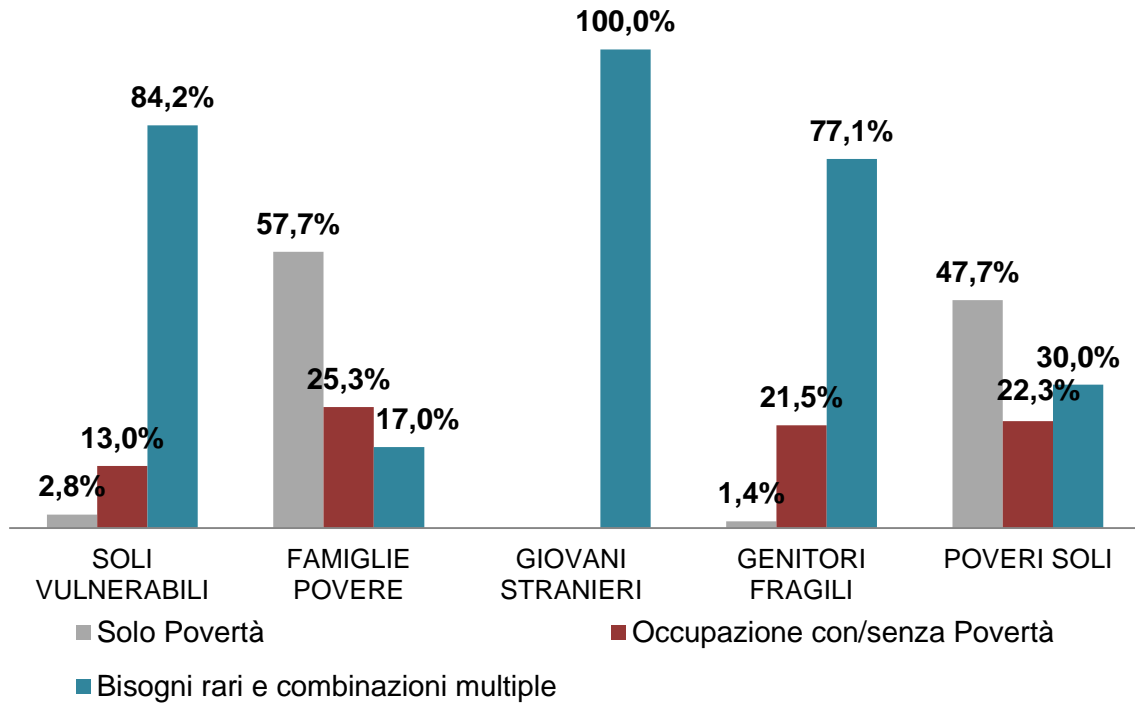


Grafico della distribuzione degli assistiti Caritas per tipologia di bisogni, all'interno di ogni cluster.

Assistiti per Multidimensionalità dei bisogni

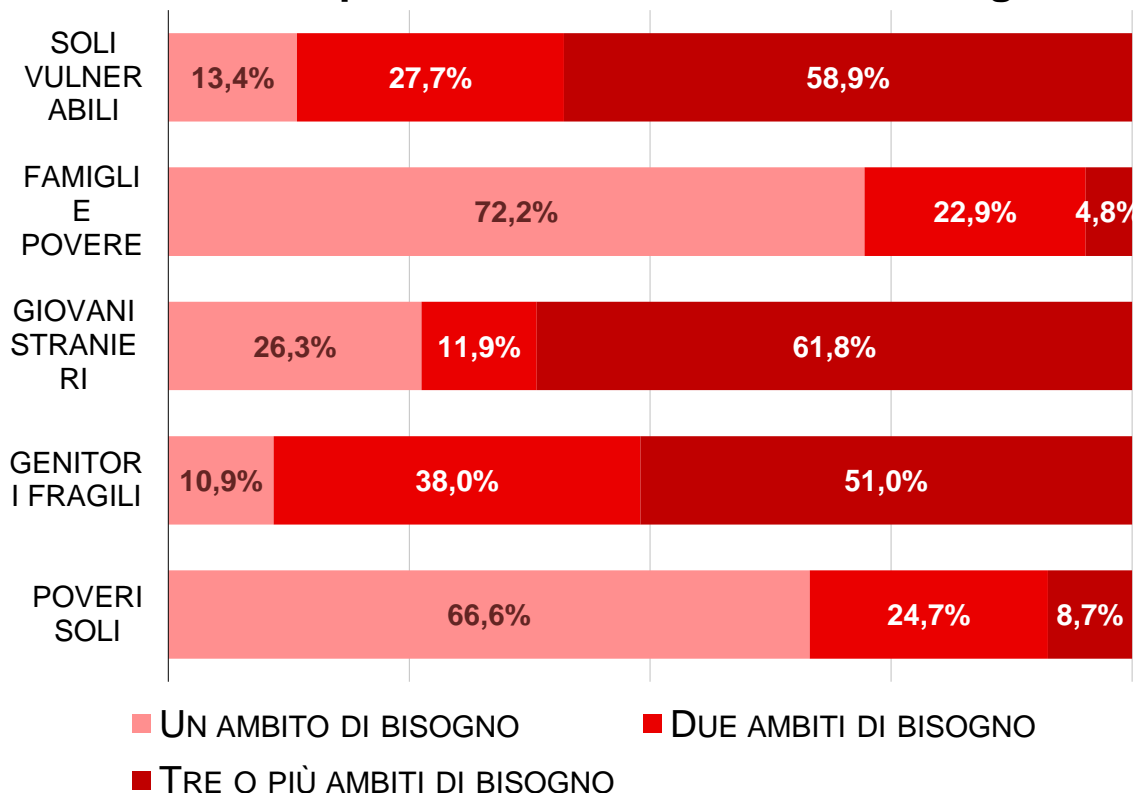


Grafico della distribuzione degli assistiti Caritas per multidimensionalità dei bisogni, all'interno di ogni cluster.

I due grafici relativi ai bisogni evidenziano notevoli differenze tra i *cluster*. Il *cluster* GIOVANI STRANIERI è caratterizzato dalla tipologia “*Bisogni rari e combinazioni multiple*” che comprende sia persone con più tipologie di bisogno, che persone con una sola tipologia di bisogno diversa da povertà od occupazione, che sono le più rilevanti nel database Ospoweb. Si nota dal secondo grafico, che circa il 62% delle persone di questo cluster presentano tre o più ambiti di bisogno, che è il valore più alto rispetto a tutti i cluster.

Anche i SOLI VULNERABILI e i GENITORI FRAGILI sono in maggioranza portatori di svariate tipologie di bisogni, con un'incidenza molto elevata di combinazioni multiple.

I *cluster* FAMIGLIE POVERE e POVERI SOLI sono invece sbilanciati verso la presenza del solo bisogno di povertà, (risp.te: 58% e 48%). In entrambi i casi l'incidenza delle persone che presentano un solo ambito di bisogno è però più alta (risp.te: 72% e

67%), per cui combinando i due dati si può affermare che nel 14% e nel 19% dei casi, rispettivamente, gli assistiti di questo gruppo presentano un solo ambito di bisogno che è diverso da quello di povertà.

Livello territoriale

Assistiti per Livello territoriale e Cluster

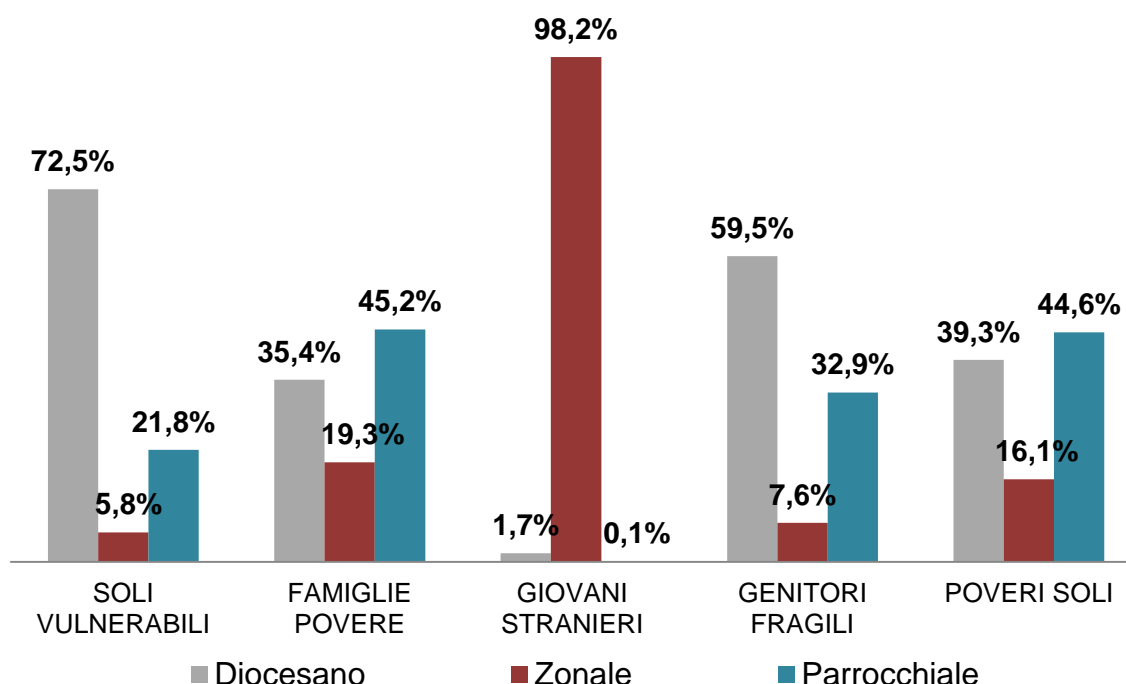


Grafico della distribuzione degli assistiti Caritas per livello territoriale, all'interno di ogni cluster.

La distribuzione per livello territoriale conferma l'unicità del *cluster* GIOVANI STRANIERI, caratterizzato dalla quasi totalità di assistiti da CdA zonali. Tra gli altri *cluster*: i SOLI VULNERABILI presentano un'incidenza molto elevata di CdA diocesani (72%), mentre sia nel *cluster* FAMIGLIE POVERE che nel *cluster* POVERI SOLI si presenta una moderata prevalenza di assistiti da CdA parrocchiali.

Dati per regione

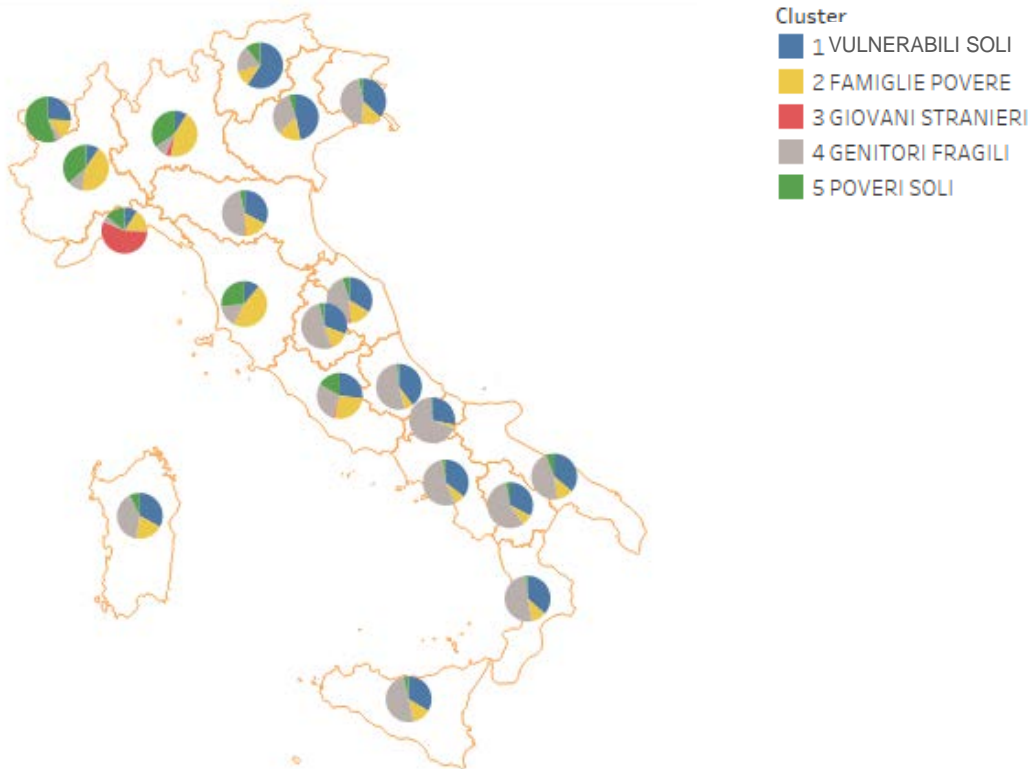
I dati per regione geografica sono stati calcolati per restituire alle direzioni regionali una fotografia della situazione sul proprio territorio. Dal punto di vista statistico i dati di alcune regioni sono associati a una variabilità elevata delle stime per cui devono essere valutati con prudenza. E' il caso, soprattutto, di Molise e Valle d'Aosta, ma anche di Basilicata, Trentino A.A. e Friuli V.G.

NB: diversamente dalle altre elaborazioni presentate in questa sezione, i dati per regione sono stati calcolati su tutte le persone assegnate a *cluster*, indipendentemente dalla distanza dal centroide del *cluster* stesso. Le differenze tra le due distribuzioni sono comunque contenute: i massimi scostamenti riguardano l'incidenza del cluster POVERI SOLI in Toscana e in Lombardia, e oscillano intorno al 3%.

Regione geografica	VULNERABILI FAMIGLIE		GIOVANI STRANIERI	GENITORI FRAGILI	POVERI SOLI	Totale
	SOLI	POVERE				
Piemonte	9,4%	43,1%	0,7%	10,3%	36,5%	100%
Valle d'Aosta	25,7%	14,2%	0,4%	4,1%	55,6%	100%
Lombardia	8,8%	44,3%	3,5%	8,8%	34,6%	100%
Trentino-Alto Adige	59,3%	11,2%	0,6%	18,3%	10,6%	100%
Veneto	46,8%	16,2%	0,3%	31,7%	5,0%	100%
Friuli-Venezia Giulia	36,9%	13,7%	0,0%	46,4%	3,0%	100%
Liguria	9,2%	16,5%	55,6%	4,3%	14,3%	100%
Emilia-Romagna	32,7%	16,8%	0,8%	45,9%	3,7%	100%
Toscana	11,0%	46,5%	0,5%	15,4%	26,6%	100%
Umbria	30,7%	13,5%	0,6%	51,8%	3,4%	100%
Marche	33,6%	15,7%	1,0%	44,8%	4,8%	100%
Lazio	26,7%	27,1%	1,3%	28,4%	16,5%	100%
Abruzzo	39,7%	6,1%	0,1%	52,4%	1,6%	100%
Molise	28,4%	2,4%	0,0%	68,5%	0,7%	100%
Campania	35,6%	7,3%	0,2%	54,6%	2,2%	100%
Puglia	36,1%	11,5%	0,2%	46,5%	5,8%	100%
Basilicata	32,6%	7,6%	0,0%	56,9%	2,9%	100%
Calabria	36,4%	10,5%	0,1%	51,3%	1,7%	100%
Sicilia	33,4%	12,8%	0,2%	50,1%	3,4%	100%
Sardegna	32,9%	19,4%	0,1%	40,3%	7,2%	100%
ITALIA	23,2%	25,5%	7,6%	27,8%	15,9%	100%

Tabella della distribuzione degli assistiti Caritas per cluster all'interno di ogni Regione geografica.

Di seguito si presenta un cartogramma dei dati costruito con il software Tableau, dal quale si evince ad occhio nudo sia la particolarità della distribuzione per *cluster* della Liguria, sia l'elevata incidenza del *cluster* GENITORI FRAGILI in tutta la fascia adriatica e nel sud del Paese.



Cartogramma della distribuzione degli assistiti Caritas per cluster all'interno di ogni Regione geografica.

3.3. Commento ai risultati

I profili degli assistiti Caritas 2022 emersi dall'analisi multivariata dei dati ci indicano che ci sono due grandi dimensioni che contribuiscono a differenziare le persone assistite:

- Caratteristiche del nucleo al quale è legata la persona assistita e in particolare le due grandi tipologie
 - Persone sole (in prevalenza uomini)
 - Genitori di minori e altre persone inserite in un nucleo familiare (in prevalenza donne)
- La tipologia di bisogni rilevati e in particolare la dicotomia tra:
 - Solo bisogno di povertà



- Bisogni multipli, a volte associati con fragilità familiari, sociali e/o di tipo psicologico.

In merito ai bisogni rilevati, si osserva che essi, molto più degli interventi effettivamente fruiti e delle richieste avanzate/registrate, sono uno specchio della dimensione a tutto tondo della persona. Premesso che la povertà è la condizione di fondo della grande maggioranza degli assistiti, coloro per i quali viene rilevato il solo bisogno di povertà sono intrinsecamente più resilienti rispetto a chi invece presenta bisogni multipli. Se poi questa condizione di “sola” povertà si innesta in un contesto di appartenenza familiare, la persona ha maggiori possibilità di stabilizzare la propria esistenza, pur continuando ad essere assistita, anche per più anni.

Quanto delineato si sposa quasi alla perfezione con i gruppi emersi dall'analisi multivariata dei dati. Quattro di essi sono infatti il risultato di un prodotto cartesiano 2x2 delle grandi dimensioni descritte nelle righe precedenti. I cluster i nn. 1 (SOLI VULNERABILI) e 5 (POVERI SOLI) sono costituiti da persone sole, mentre i nn. 2 (FAMIGLIE POVERE) e 4 (GENITORI FRAGILI) da persone inserite in un contesto familiare e molto spesso con figli minori. I *cluster* FAMIGLIE POVERE e POVERI SOLI evidenziano una forte incidenza del solo bisogno di povertà, mentre i *cluster* SOLI VULNERABILI e GENITORI FRAGILI sono dominati da persone associate a bisogni multipli, a volte con fragilità. In questo incrocio, ovviamente, agiscono sottotraccia anche altre variabili (dato che le dimensioni principali assunte a base delle procedure di *clustering* sono state quattro), tra le quali la dimensione territoriale, le tipologie di interventi fruiti, la durata dell'assistenza, la dimensione demografica delle località di residenza, il livello territoriale dei CdA Caritas che erogano l'assistenza, in parte anche l'età. Ciascuna di queste variabili contribuisce ad arricchire i profili e certamente integrano il ruolo trainante delle due grandi dimensioni descritte.

A una prima analisi superficiale, i *cluster* FAMIGLIE POVERE e GENITORI FRAGILI sembravano poter essere ulteriormente aggregati tra loro, ma analisi più approfondite hanno evidenziato ulteriori differenze che invece vale la pena sottolineare. La prima è la durata dell'assistenza. Il *cluster* FAMIGLIE POVERE presenta infatti una forte



incidenza di persone assistite ormai da più di 5 anni, molte di loro da oltre 10. La seconda riguarda l'ambito del lavoro: la percentuale di persone non occupate del *cluster* FAMIGLIE POVERE non differisce da quella del *cluster* GENITORI FRAGILI e la differenza con la media generale non è abissale. Eppure le persone appartenenti al *cluster* FAMIGLIE POVERE presentano un bisogno di occupazione che è circa la metà delle persone del *cluster* GENITORI FRAGILI. Disaggregando questi dati per gruppi di nazionalità si è accertato che le persone di nazionalità africana, che caratterizzano al 30% il *cluster* FAMIGLIE POVERE, sono molto più spesso occupate (35% contro il 23% degli italiani), quindi un sottogruppo più stabile da questo punto di vista. Tra l'altro, affermazione non certificata da dati ma da osservazioni indirette di tipo qualitativo, le persone di nazionalità africana attivano spesso reti di assistenza che legano l'un l'altro interi nuclei familiari. Di fatto possono quindi contare quindi su un'ulteriore assistenza parallela e questo spiega in parte come queste persone non “chiedano” a Caritas altre prestazioni che beni e servizi. La relativa stabilità assicurata dalla rete inter-familiare le porta, forse, a cristallizzare una situazione nella quale accettano da Caritas un'assistenza materiale in termini di beni e servizi, ma al contempo non sentono il bisogno di attivarsi per un'occupazione stabile. Tutte ipotesi di lavoro che potranno essere approfondite in seguito.

In questo contesto abbiamo finora trascurato il *cluster* GIOVANI STRANIERI, ma non a caso. Questo gruppo è costituito infatti da nuovi arrivati che in qualche modo considerano l'Italia come tappa di passaggio verso il Nord- Europa. Non sono motivati a mettere radici, ma solo ad attraversare prima possibile la frontiera. Essendo per lo più giovani, hanno speranze, aspirazioni ed energie che li hanno portati a viaggiare da sud a nord, o in parte da est ad ovest, accumulandosi nei pressi della frontiera di Ventimiglia. Sono sul nostro Paese, hanno bisogno di assistenza e non possono essere ignorati. Ma allo stesso tempo costituiscono un segmento fluido, scivoloso, che difficilmente produrrà relazioni di medio-lungo periodo. In questo caso tra bisogni rilevati e assistenza fruita c'è un *gap* visibile. Essi non chiedono, se possibile rimangono nascosti, sono intercettabili solo quando si rendono visibili nel premere alla frontiera. Di conseguenza



sono potenzialmente soggetti a un forte tasso di sostituzione. Ciò non significa che questo segmento possa dissolversi, anzi potrebbe viceversa rafforzarsi con l'aumento degli sbarchi che si è verificato nel corrente anno.