

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in Statistica per l'Economia e l'Impresa



**RELAZIONE FINALE
UN'ANALISI SULLA POVERTÀ IN ITALIA NEL 2020**

Relatore Prof. Omar Paccagnella
Dipartimento di Scienze Statistiche

Laureando: Sofia Noemi Crobeddu
Matricola N. 1218766

Anno Accademico 2022/2023

Alla mia famiglia

INDICE

| | |
|--|-----------|
| Introduzione | 7 |
| 1. La povertà | 9 |
| 1.1. Origini storiche in Italia | 9 |
| 1.2. Esclusione sociale e disuguaglianza | 10 |
| 1.3. Tipi di povertà | 10 |
| 1.3.1. Povertà a lungo e breve termine | 11 |
| 1.3.2. Povertà ciclica, concentrata e individuale | 11 |
| 1.3.3. Povertà assoluta e relativa | 12 |
| 1.4. Povertà multidimensionale e il “Capability approach” | 12 |
| 2. Misurare la povertà | 15 |
| 2.1. Unità statistica | 15 |
| 2.2. Reddito e spesa per i consumi a confronto | 15 |
| 2.3. Soglia di povertà | 16 |
| 2.4. Scale di equivalenza | 17 |
| 2.5. Indicatori di povertà | 18 |
| 2.5.1. Curva di Lorenz e indice di Gini | 19 |
| 2.5.2. Head Count Ratio o indice di incidenza | 20 |
| 2.5.3. Income gap ratio o indice di intensità | 21 |
| 2.5.4. Poverty gap index | 21 |
| 2.5.5. Indice di Sen | 22 |
| 2.5.6. Poverty gap squared | 22 |
| 3. L’indagine della Banca d’Italia | 25 |
| 3.1. Il disegno campionario | 25 |
| 3.2. Rilevazione dei dati | 26 |
| 3.3. Le variabili aggregate | 27 |
| 3.4. Ulteriori modifiche dell’edizione 2020 | 28 |
| 4. Dati trasversali dell’indagine 2020 | 31 |
| 4.1. Analisi esplorativa | 31 |
| 4.2. Calcolo dei redditi equivalenti e soglia di povertà relativa | 36 |
| 4.3. Stima degli indicatori di povertà unidimensionali | 36 |
| 4.4. Confronti con Istat | 38 |
| 4.4.1. Incidenza di povertà relativa rispetto alle classi d’età nel 2020 | 38 |
| 4.4.2. Incidenza di povertà relativa rispetto al livello di istruzione nel 2020 | 38 |
| 4.4.3. Incidenza di povertà relativa rispetto alla condizione professionale nel 2020 | 39 |
| 4.4.4. Incidenza di povertà relativa rispetto all’area geografica nel 2020 | 39 |
| 4.5. Analisi specifica per area geografica | 40 |
| 5. Analisi multidimensionali | 43 |
| 5.1. Factor Analysis | 43 |
| 5.1.1. Variabili utilizzate | 43 |
| 5.1.2. Metodi per il calcolo dei fattori | 44 |
| 5.1.3. Analisi dei punteggi fattoriali per area geografica | 47 |
| 5.2. Modello di regressione lineare multipla | 49 |
| 5.2.1. Stima dei parametri | 49 |
| 5.2.2. Interpretazione del modello | 53 |
| 6. Conclusioni | 55 |

| | |
|---------------------------|-----------|
| Appendice A | 57 |
| Appendice B | 57 |
| Appendice C | 58 |
| Bibliografia | 59 |

INTRODUZIONE

L'obiettivo della presente relazione finale è quello di analizzare il fenomeno della povertà in Italia nel 2020, attraverso i dati dell'Indagine svolta dalla Banca d'Italia. In particolare, vengono utilizzati tre dataset tra quelli disponibili nell'archivio dati dello stesso istituto, concentrando l'attenzione su variabili di tipo demografico, sul reddito, che è anche la variabile di misurazione del benessere economico, e variabili di opinione sul reddito, spesa e consumi. Vengono quindi utilizzati strumenti statistici ed economici per analizzare il fenomeno sia da un punto di vista unidimensionale, sia da quello multidimensionale. Viene inoltre svolto un confronto tra la presente indagine di partenza e quella svolta da ISTAT sulle spese delle famiglie italiane, sempre nello stesso anno di riferimento, al fine di avere uno studio più approfondito del tema. Le stime e i risultati ottenuti o confrontati vengono poi interpretati contestualizzandoli al periodo storico critico che il mondo e l'Italia ha passato a causa della pandemia (si tratta proprio il primo anno di Covid-19). Nell'ultima parte del presente lavoro viene applicata un'Analisi fattoriale per studiare la povertà attraverso delle dimensioni non direttamente osservate. Viene infine applicato un modello di regressione lineare a ciascuna di queste dimensioni per capire quali siano le variabili demografiche che più le influenzano.

Tutte le analisi (stime, grafici, applicazioni di metodi, ...) sono state realizzate utilizzando il software R.

Capitolo 1

La povertà

La povertà è un fenomeno antico che non ha confini sociali poiché colpisce persone in tutto il mondo, e per sua natura, è una condizione che nega all'individuo il diritto di esercitare appieno il proprio potenziale (Aku, Ibrahim e Bulus, 1997). Tuttavia, non esiste un consenso universale sulla precisa definizione di povertà, in quanto per la sua concettualizzazione vengono utilizzati metodi diversi. Il problema della definizione è ulteriormente aggravato dalle connotazioni non economiche che la parola povertà ha acquisito. In ambito economico essa si riferisce alla carenza di risorse finanziarie di cui una famiglia dispone per soddisfare i propri bisogni fondamentali. In questo contesto, l'identificazione delle persone povere richiede innanzitutto la determinazione dei bisogni primari. Questi possono essere definiti in senso stretto come quelli necessari alla sopravvivenza o in senso lato come quelli che riflettono lo standard di vita prevalente nella comunità. Il primo criterio coprirebbe solo le persone vicine al limite della fame o della morte, il secondo si estenderebbe alle persone la cui alimentazione, abitazione e abbigliamento, sebbene adeguati a preservare la vita, non sono all'altezza di quelli della popolazione nel suo complesso. La povertà però non si limita solo alla mancanza di reddito e di risorse produttive per garantire mezzi di sussistenza dei bisogni primari, ma le sue manifestazioni includono anche la fame e la malnutrizione, l'accesso limitato all'istruzione e ad altri servizi di base, la discriminazione, l'esclusione sociale. Porre fine ad ogni forma di povertà nel mondo è il tema centrale dell'Obiettivo 1 dell'Agenda 2030 per lo Sviluppo Sostenibile, il programma sottoscritto nel 2015 dai Paesi membri dell'ONU e che ingloba 17 Obiettivi per lo Sviluppo Sostenibile (Sustainable Development Goals - SDGs) del nostro pianeta.

1.1. Origini storiche in Italia

Il dibattito sulla povertà in Italia ha inizio nel 1951, quando viene costituita la "Commissione parlamentare sulla miseria in Italia e sui mezzi per combatterla" con inizialmente 15 componenti, diventati poi 21 nel 1952. Essa aveva "il compito di condurre una indagine sullo stato attuale della miseria al fine di accertare le condizioni di vita delle classi povere ed il funzionamento delle istituzioni di assistenza sociale" (Archivio storico della Camera dei deputati, art. 1). Lo scopo dell'indagine era quello di misurare le condizioni di povertà della popolazione italiana e verificare le funzioni assistenziali degli organi statali. L'inchiesta comprendeva quindi:

- indagini svolte nelle zone depresse dell'Italia, cioè dove la povertà si presentava "in forme più penose o caratteristiche" (Atti Commissione, vol. I, p. 21);
- indagini statistiche svolte attraverso l'Istituto centrale di statistica (ISTAT) e l'Alto commissariato per l'igiene e la sanità (ACIS);
- indagini monografiche svolte da studiosi su aspetti economici e giuridici della miseria;
- un'indagine speciale svoltasi nel borgo di Grassano (provincia di Matera), "ai fini di una valutazione approfondita della situazione di una comunità rurale considerata nella sua struttura economica e sociale" (Atti Commissione, vol. I, p. 22).

L'indice sintetico risultante da questa indagine del 1952 in Italia aveva segnalato che l'11,8% delle famiglie viveva in condizioni di povertà relativa e sottolineava anche il divario di povertà a livello geografico: l'1,5% nel Nord, mentre il 28,3% nel Sud.

La Commissione venne sciolta nel 1954, e ne venne istituita un'altra nel 1984 con il presidente del consiglio Bettino Craxi, la Commissione d'indagine sulla povertà che durò fino al 2012. Nel 1985, come nell'indagine precedente, la stima era riferita alla povertà relativa e venne rilevato che l'11% delle famiglie italiane viveva in uno stato di povertà.

1.2. Esclusione sociale e disuguaglianza

Negli anni, sono emersi nuovi approcci al concetto stesso di povertà. Uno di questi riguarda la nozione di esclusione sociale, che è stata adottata dalla Commissione Europea come base per il suo programma sulla povertà negli anni Ottanta. I due economisti Chakravarty e D'Ambrosio (2006) ritengono che un uomo sia socialmente escluso se non è in grado di partecipare alle attività economiche e sociali di base della società in cui vive. In generale, l'esclusione sociale include la mancanza di beni, diritti, prodotti, amministrazioni e l'impotenza a partecipare alle normali relazioni e attività, accessibili a un numero più ampio degli individui in una società, sia a livello sociale, sia economico, politico o culturale. Sebbene sia vero che una persona può essere economicamente povera senza essere socialmente esclusa, le due esperienze sono fortemente correlate. L'esclusione sociale può portare alla povertà economica direttamente attraverso la mancanza di reti sociali che offrono opportunità o indirettamente attraverso la disparità nelle strutture della società stessa.

Anche la nozione di disuguaglianza si intreccia con quella di povertà. In statistica economica, il termine indica la diversità nel reddito fra i vari soggetti considerati. La disuguaglianza però non ha solo una componente 'quantitativa', ma riguarda anche le disparità di status, accesso e opportunità, che rappresentano la componente 'etica' del termine. In senso globale si può intendere come perdita di benessere sociale di cui la povertà è solo una delle tante conseguenze, più presente nelle società diseguali.

1.3. Tipi di povertà

Il significato della povertà è cambiato nel tempo. Nei modi di produzione economica tradizionali (cioè pre-industriali, nel periodo tra il 1750 e 1850), questo evento era accettato come inevitabile. La produzione totale di beni e servizi, anche se equamente distribuita, sarebbe stata comunque insufficiente a garantire all'intera popolazione un tenore di vita confortevole secondo gli standard prevalenti. Con la produttività economica derivante dall'industrializzazione (a partire dal 1870 circa), tuttavia, questa situazione è venuta meno, soprattutto nei Paesi più industrializzati del mondo, dove la produzione nazionale era sufficiente a portare l'intera popolazione a un livello confortevole, se si riusciva ad organizzare la necessaria redistribuzione senza incidere negativamente sulla produzione. Si possono distinguere diversi tipi di povertà in base a fattori quali il tempo o la durata (a lungo e a breve termine o ciclica) e la distribuzione (diffusa, concentrata, individuale). C'è poi la differenziazione tra

povertà assoluta e relativa, che si basa sulla diversa misurazione del fenomeno: a livello globale o circoscritto ad un determinato Paese.

1.3.1. Povertà a lungo e breve termine

La povertà ciclica si riferisce alla povertà che può essere diffusa in tutta la popolazione, ma che ha una durata limitata. Nelle società non industriali, questo tipo di incapacità di provvedere ai propri bisogni primari si basa principalmente su carenze alimentari temporanee causate da fenomeni naturali o da una cattiva pianificazione agricola, e conseguente aumento di prezzi. Nelle società industrializzate invece, la principale causa ciclica di povertà è rappresentata dalle fluttuazioni del ciclo economico (ovvero una conseguenza della regolamentazione del mercato), con disoccupazione di massa nei periodi di depressione o di grave recessione. I governi ricorrono quindi a politiche economiche aumentando la spesa pubblica, sia per assistenza diretta ai disoccupati, sia per il welfare e altri sussidi familiari. Sebbene le depressioni economiche colpiscano tutti i segmenti della società, l'impatto è più grave sulle persone degli strati socioeconomici più bassi, perché hanno meno risorse marginali rispetto a quelle degli strati più alti.

La povertà a lungo termine non verrà qui descritta nello specifico. Essa riguarda indagini statistiche mirate alla previsione della percentuale di povertà di un determinato individuo o nucleo familiare nell'arco di un intervallo di tempo lungo (ad esempio qualche anno), usando anche i dati longitudinali.

1.3.2. Povertà ciclica, concentrata e individuale

A differenza della povertà ciclica, che è temporanea, la povertà diffusa implica un'insufficienza relativamente permanente di mezzi per garantire i bisogni di base, una condizione che può essere così generale da descrivere il livello medio di vita di una società. Questo causa sottosviluppo economico, bassa aspettativa di vita, alti livelli di mortalità infantile e cattive condizioni di salute nella vita di queste società. Le risorse totali di molte nazioni in via di sviluppo che soffrono di questo problema, sarebbero insufficienti a sostenere adeguatamente la popolazione anche se fossero equamente divise tra tutti i cittadini.

Esiste anche la povertà concentrata, una tipologia presente soprattutto nei ghetti delle città, nelle regioni escluse o abbandonate dall'industria e nelle aree in cui l'agricoltura o l'industria sono inefficienti e non possono competere con il profitto delle zone centrali. Sono aree caratterizzate da disoccupazione, tassi di mortalità più elevati, cattive condizioni di salute, bassi livelli di istruzione rispetto ai segmenti più abbienti della società.

La povertà individuale invece si riferisce all'incapacità di un individuo o di una famiglia di assicurarsi i bisogni di base anche in un contesto sociale di generale prosperità. Questa incapacità è generalmente legata alla mancanza di qualche attributo di base che consenta all'individuo di mantenersi. Queste persone possono, ad esempio, essere cieche, disabili fisici o emotivi o malati cronici. Gli aiuti pubblici per migliorare la povertà dovuta a queste cause si

concentrano sull'istruzione, sul lavoro protetto e, se necessario, sul mantenimento economico da parte dello Stato.

1.3.3. Povertà assoluta e relativa

Esiste poi la differenziazione tra povertà assoluta e relativa. La povertà assoluta o estrema si riferisce all'insieme di risorse che una persona deve acquisire per mantenere uno standard di vita minimo per la sopravvivenza (fabbisogno nutrizionale), inteso a livello globale.

La povertà relativa è giudicata invece attraverso la posizione del povero rispetto alla società, ovvero riguarda il grado di peggioramento di un individuo o di una famiglia rispetto agli altri individui della stessa società (di uno stesso Stato, o di una stessa città ad esempio). Il concetto di povertà relativa è stato per la prima volta esplicitato dal sociologo Townsend¹ nel 1979, il quale l'ha definito come il livello di privazione in cui gli individui non sono in grado di "svolgere i ruoli, partecipare alle relazioni e seguire il comportamento abituale che ci si aspetta da loro in virtù della loro appartenenza alla società". La povertà relativa non riflette necessariamente la vulnerabilità alla mortalità o alle sofferenze acute, ma piuttosto il livello di divario in un determinato contesto e per questo motivo è un indicatore del grado di disuguaglianza. Per questo motivo necessita quindi della situazione media o mediana della società considerata.

1.4. Povertà multidimensionale e il "Capability approach"

Lo studio della Banca Mondiale "Le voci dei poveri", ha rilevato che piuttosto che il reddito, i poveri sono più propensi a descrivere la loro realtà in termini di beni fisici, umani, sociali e ambientali. Proprio per questo la povertà è un fenomeno che deve essere concepito anche multidimensionalmente. I beni si riferiscono alle risorse che i poveri possono utilizzare per gestire o ridurre il rischio di vulnerabilità agli shock esterni (come ad esempio il Covid-19), e possono essere tangibili o intangibili (per esempio, la salute, gli amici, l'istruzione).

Secondo Amartya Sen, economista e filosofo indiano e professore dell'Università di Harvard oltre che premio Nobel per l'economia nel 1998, la povertà è più di una condizione finanziaria, in cui mancano le necessità fondamentali della vita, come ad esempio il nutrimento, l'alloggio e l'abbigliamento (2001). Sen ha elaborato il "Capability approach" che riflette il concetto di povertà multidimensionale intesa come livello di benessere sociale. Prevede due fondamenti principali: in primo luogo, l'assunto che la libertà di raggiungere il benessere è di primaria importanza morale, e in secondo luogo, la libertà di raggiungere il benessere deve essere intesa in termini di persone dotate di capacità. In altre parole, l'approccio delle capacità sostiene che la libertà di raggiungere il benessere sia una questione di ciò che le persone possono fare ed essere, e quindi dalla qualità di vita che possono effettivamente condurre. Come sostiene Sen, i beni o la ricchezza delle persone o l'utilità sono un punto di vista limitato. Sen illustra il metodo con l'esempio di una bicicletta standard: essa ha

¹ Townsend P. (1979), *Poverty in the United Kingdom*, Harmondsworth, Penguin.

le caratteristiche di un "mezzo di trasporto", ma se effettivamente fornirà un mezzo di trasporto dipenderà dalle caratteristiche fisiche e mentali di coloro che cercheranno di usarla (condizioni di salute, etc). In questo senso la povertà viene identificata dalla deprivazione di capacità, in termini di "vita che le persone possono effettivamente condurre e di libertà che effettivamente hanno" (Sen, 1985), e quindi quanto più i poveri avvertono la condizione di povertà, tanto più il fenomeno è intenso.

Capitolo 2

Misurare la povertà

In questo capitolo si forniscono degli strumenti volti a misurare e quantificare la povertà vista in ottica unidimensionale attraverso indagini ed analisi statistiche e economiche. Si sceglie quindi l'unità statistica, la variabile per la misura del benessere economico, si individua una soglia di povertà e infine si rendono confrontabili le variabili usate per la misurazione del fenomeno.

2.1. Unità statistica

Nel misurare la povertà, l'unità di analisi può essere definita come famiglia, nucleo familiare o individuo. In genere, la povertà viene misurata a livello di famiglia o di nucleo familiare perché questa entità costituisce un'unità naturale di condivisione del reddito, e poiché la povertà è un fenomeno recepito a livello familiare piuttosto che individuale. Inoltre, la famiglia come unità di analisi può essere molto utile a fini politici, poiché i programmi di assistenza pubblica sono spesso basati su tipologie di famiglie. A fini pratici, tuttavia, la famiglia è generalmente considerata l'unità di analisi standard nella ricerca sulla povertà, poiché la maggior parte delle indagini internazionali raccoglie informazioni a livello familiare. In Italia le indagini statistiche sulla povertà sono ad esempio quella svolta da ISTAT sulle spese delle famiglie o quella della Banca d'Italia sui redditi delle famiglie. In entrambe l'unità statistica è la famiglia (anche se per le informazioni demografiche Banca d'Italia fa rilevazioni individuali), che ISTAT definisce come "l'insieme di persone legate da vincoli di matrimonio, unione civile, parentela, affinità, adozione, tutela, o da vincoli affettivi, coabitanti e aventi dimora abituale nello stesso comune (anche se non sono ancora iscritte nell'anagrafe della popolazione residente del comune medesimo). Una famiglia può essere costituita anche da una sola persona. L'assente temporaneo non cessa di appartenere alla propria famiglia sia che si trovi presso altro alloggio (o convivenza) dello stesso comune, sia che si trovi in un altro comune italiano o all'estero"².

2.2. Reddito e spesa dei consumi a confronto

Per la misura del benessere economico si usano solitamente due indicatori: il reddito disponibile e la spesa per i consumi. L'ISTAT definisce il reddito come "la remunerazione delle attività di cui dispone un individuo, reali, finanziarie e umane". Questa variabile è importante perché misura la capacità d'acquisto delle famiglie e genera ricchezza. La spesa per consumi è invece definita come "una stima diretta dei beni e servizi che concorrono a determinare le condizioni di vita di un individuo o di una famiglia, anche se nel confronto interpersonale risente delle diverse preferenze e abitudini degli individui"³. Stabilire quale variabile sia più adeguata nell'analisi. È una questione aperta. Le critiche rispetto alla variabile del reddito riguardano la temporaneità di questa dimensione, soggetta spesso a cambiamenti causati dalla vita professionale di un individuo (ad esempio passaggi di carriera, interruzioni del

² Dal Glossario statistico dell'Istituto Nazionale di Statistica.

³ *Le dimensioni del benessere economico*, Istituto Nazionale di Statistica (ISTAT).

rapporto di lavoro, etc..), e la relazione rispetto alle necessità di vita di ognuno da cui non si può prescindere (ad esempio da giovani non si è occupati ma, pur non percependo reddito, non possiamo consumare meno di quanto abbiamo bisogno per sopravvivere). Dall'altro lato, la spesa per il consumo si basa sulla teoria microeconomica fondata sull'ipotesi di razionalità dei consumatori di difficile riscontro empirico e in particolare essa riflette le preferenze degli individui rischiando quindi di sottostimare o sovrastimare la povertà (ad esempio si sovrastimerebbe la povertà di un nucleo familiare che sceglie di adottare uno stile di vita sobrio).

In questo dibattito e alla luce di queste argomentazioni, entrambe le variabili sono attendibili e la scelta finale dipende dalla disponibilità dei dati e dall'attendibilità delle fonti statistiche poiché le considerazioni teoriche non risultano sufficienti.

2.3. Soglia di povertà

La misurazione della povertà economica comporta la definizione di un limite minimo che distingua i poveri dagli altri poveri attraverso i livelli di reddito o di consumo delle persone in riferimento a tale soglia. Secondo l'approccio della metrica del denaro, quest'ultimo è il parametro migliore per misurare la povertà, e per distinguere tra poveri e non poveri è importante decidere il livello di salario che si prevede possa soddisfare i bisogni di base. In questo caso, si definisce povero chi non riesce a soddisfare i bisogni primari perché il suo salario è al di sotto del livello limite, e questo impedisce una partecipazione attiva nella società. Questo livello limite è noto come soglia di povertà. Per valutare però la povertà a livello globale bisogna, parallelamente, valutare anche le differenze tra i Paesi. I risultati delle stime utilizzando le soglie di povertà a livello mondiale, infatti, sono poco utili e creano distorsione, sia perché non è possibile mantenere un paniere di prodotti alimentari universale su cui basare la soglia standard, sia perché l'inflazione dilagante e l'incertezza dei prezzi rendono la determinazione della soglia di povertà ancora più disomogenea. La soglia di povertà può essere basata su dati di reddito o di spesa. Sebbene i dati sul reddito siano comunemente utilizzati per misurare il tenore di vita, essi soffrono di molti problemi di misurazione (ad esempio, la sottostima), soprattutto nei Paesi in via di sviluppo. I dati sulla spesa, invece, contengono i valori monetari dei beni consumati e possono quindi essere utilizzati per determinare il prezzo di un paniere di beni di prima necessità. D'altra parte però, nonostante spesso si ritenga che i dati sulla spesa stimino meglio il bisogno rispetto ai dati sul reddito, essi possono semplicemente riflettere scelte personali di allocazione delle risorse o di prudenza durante i periodi di basso reddito. La convenzione ideale sarebbe quella di utilizzare i dati sulla spesa per calcolare i bisogni nella costruzione della soglia di povertà e i dati sul reddito per esaminare quali famiglie soddisfino i loro bisogni fondamentali.

Nel 2018 la Banca Mondiale ha fissato la soglia di povertà estrema o assoluta a 1,90 dollari al giorno e quella di povertà moderata a 2 dollari giornalieri. In generale, la soglia di povertà può essere trattata in termini assoluti se basata su un paniere di consumo minimo, oppure in termini relativi se è percentuale del reddito medio o mediano. In questo caso, il limite di povertà viene comunemente fissato al 50% del reddito familiare medio del Paese oppure al 60% del reddito mediano familiare. A prescindere dal metodo che si sceglie di usare nel calcolo, il vero valore di una

qualsiasi soglia di povertà risiede nel mantenere una misura che consenta di effettuare confronti all'interno di gruppi e tra di essi nel corso del tempo.

2.4. Scale di equivalenza

Le soglie di povertà sono preziose per valutare la povertà e la disuguaglianza tra le famiglie, ma non sono efficaci nel cogliere le disparità all'interno delle famiglie. Le statistiche sulla povertà sono generalmente raccolte attraverso gli standard di vita delle famiglie. Come famiglia di riferimento viene solitamente considerata quella più rappresentata dal punto di vista demografico, ovvero la composizione familiare della maggior parte della popolazione. I modelli di spesa di questa famiglia di riferimento vengono quindi utilizzati come base per calcolare la soglia di povertà generale (basata sulla spesa per consumi appunto), come spiegato nel paragrafo 2.2., e si basano sull'idea che la famiglia non sia necessariamente uguale alla somma delle sue parti. In altre parole, gli individui all'interno della famiglia possono avere esigenze e consumi diversi. Ad esempio, mentre un adulto può richiedere una certa dieta energetica per la sussistenza, un bambino può averne bisogno in misura minore. In questo caso, non si può equiparare il costo di questi due individui nella famiglia. Allo stesso modo, mentre vivere da soli può costare un determinato importo, l'aggiunta di un altro adulto porta a una situazione di consumo comune in cui entrambi gli individui possono condividere l'uso di alcuni beni pubblici (ad esempio, la casa, il cibo, il riscaldamento). In questo caso, anche l'aggiunta di una persona alla famiglia non porta alla somma delle due parti, cioè i bisogni di una famiglia non crescono proporzionalmente all'aumento del numero di membri. Questo fenomeno viene definito economia di scala. La scala di equivalenza, ovvero un insieme di coefficienti, serve infatti ad assegnare a ogni tipo di famiglia nella popolazione un valore proporzionale ai propri bisogni (tiene conto dell'ampiezza familiare) e confronta il costo della vita di una particolare famiglia con quello della famiglia di riferimento quando entrambe raggiungono lo stesso standard di vita. Le dimensioni della famiglia e l'età dei suoi componenti sono i fattori che vengono comunemente presi in considerazione per assegnare i valori e a tutti i membri della famiglia viene attribuito un peso, che viene sommato per ottenere le dimensioni equivalenti della famiglia. Si trova una nuova variabile di reddito (o consumo) familiare equivalente, ottenuta dividendo il reddito (o consumo) di famiglie disomogenee tra loro per i coefficienti di equivalenza. Esistono 4 classificazioni delle scale di equivalenza:

- scale econometriche, cioè basate sulle spese delle famiglie per i consumi e fondate sulla teoria microeconomica del consumatore;
- scale "soggettive", cioè basate su informazioni date dalle famiglie alle quali viene chiesto il livello di soddisfazione rispetto al reddito disponibile nel loro nucleo;
- scale ottenute con il metodo dei "minimi calorici", cioè basate sul paniere di consumo di sussistenza per famiglie di differenti caratteristiche;
- scale pragmatiche, cioè scale che individuano un reddito familiare plausibile moltiplicato per un fattore di conversione che considera la numerosità del nucleo e l'età dei componenti.

In generale, viene spesso applicata a livello mondiale la scala di equivalenza dell'OCSE (scala pragmatica): quella originaria, anche definita "scala di Oxford",

assegna un valore di 1 al primo membro della famiglia (il capofamiglia), di 0.7 a ogni adulto aggiuntivo e di 0.5 a ogni bambino. Esiste poi la "Scala modificata dall'OCSE", usata anche in Italia e dall'Istituto di statistica dell'Unione Europea EUROSTAT, che assegna un valore di 1 al capofamiglia, di 0.5 a ogni altro membro adulto e di 0.3 a ogni bambino (con età inferiore a 14 anni).

La scala di equivalenza "Carbonaro" (scala econometrica) è invece quella tradizionalmente usata dall'ISTAT e dalla Commissione d'indagine per l'esclusione sociale e viene usata per l'analisi della povertà relativa. Essa trova la sua origine nella seconda metà del XX secolo quando l'economista e filosofo Engel propose la teoria secondo cui un uguale livello di benessere può essere raggiunto da due famiglie diverse se la loro quota di spesa destinata all'acquisto di generi alimentari rispetto alla spesa totale, è la stessa (Carbonaro, 1985). Questa scala tiene conto solo dell'ampiezza delle famiglie e ha due versioni: prende infatti come riferimento sia un nucleo familiare di 1 individuo, sia di 2 individui.

| Numero di componenti | Famiglia di rif. con 1 persona | Famiglia di rif. con 2 persone |
|----------------------|--------------------------------|--------------------------------|
| 1 | 1,00 | 0,60 |
| 2 | 1,67 | 1,00 |
| 3 | 2,22 | 1,33 |
| 4 | 2,72 | 1,63 |
| 5 | 3,17 | 1,90 |
| 6 | 3,6 | 2,15 |
| 7 o più | 4,00 | 2,40 |

Tabella 2.1: Coefficienti della scala di equivalenza "Carbonaro" con famiglia di riferimento sia di 1 individuo sia di 2. Fonti: Glossario statistico dell'ISTAT e Carbonaro G. (1985), "Note sulle scale di equivalenza", Istituto Poligrafico e Zecca dello Stato.

Un'altra usata nelle statistiche ufficiali è la scala di equivalenza dell'Indicatore della situazione economica (scala pragmatica), ovvero quella utilizzata per il calcolo dell'ISEE con impiego assistenziale. Tiene conto del numero di componenti e prevede maggiorazioni di 0.2 se il nucleo è monogenitoriale, di 0.5 se ci sono componenti con handicap psicofisico e di 0.2 se si tratta di un nucleo bireddito con minori a carico.

2.5. Indicatori di povertà

In questo paragrafo vengono riportati i principali indicatori di povertà. Tra questi vengono descritti degli indici sintetici che sintetizzano il fenomeno.

Si specifica che l'indice di Gini considera l'intera distribuzione del reddito o della spesa, mentre gli altri di seguito elencati considerano la parte povera relativamente, cioè sotto la soglia di povertà relativa.

2.5.1. Curva di Lorenz e indice di Gini

La curva di Lorenz è stata sviluppata dall'economista americano Max Lorenz nel 1905 ed è una rappresentazione grafica del livello di disuguaglianza di reddito o di ricchezza, e quindi della sua distribuzione, all'interno di una popolazione. Più precisamente viene definita curva di concentrazione poiché si tratta di una funzione di distribuzione cumulativa. Essa traccia i percentili della popolazione sull'asse orizzontale in base al reddito o alla ricchezza e sull'asse verticale i percentili del reddito o della ricchezza cumulati. La curva viene rappresentata al di sotto di una diagonale di pendenza 1 che indica la perfetta uguaglianza nella distribuzione del reddito o della ricchezza. L'area tra la curva e la diagonale è la distribuzione osservata o stimata. Più la curva di Lorenz è vicina alla linea di uguaglianza, meno disuguaglianza esiste. Più la curva di Lorenz è lontana dalla linea di uguaglianza, più esiste disuguaglianza. Un esempio di curva di Lorenz è riportato in Fig. 2.1. La misurazione empirica avviene attraverso le dichiarazioni dei redditi della popolazione.

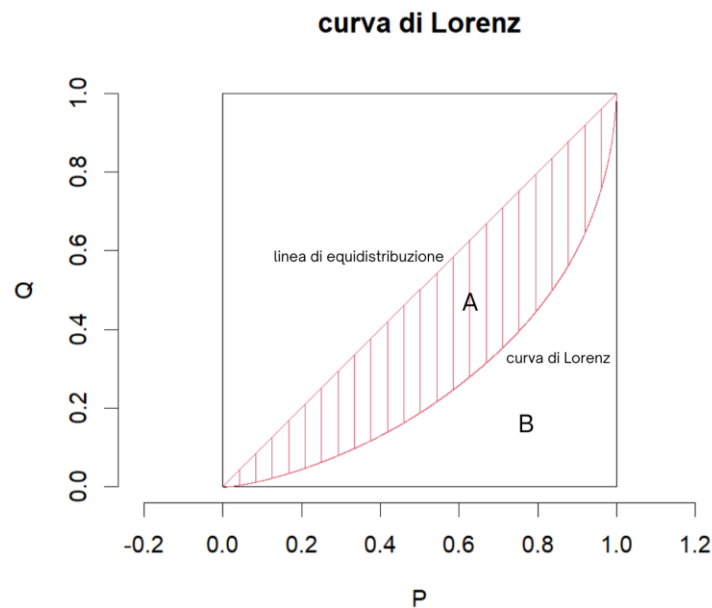


Fig. 2.1: Curva di Lorenz⁴ (disegnata con il software R). Asse x: Quota cumulata della popolazione (%). Asse y: Quota cumulata del reddito (%).

L'indice di Gini, proposto dallo statistico italiano Corrado Gini nel 1912, viene utilizzato per quantificare la disuguaglianza di una distribuzione e la curva di Lorenz è la sua interpretazione geometrica. È un numero compreso tra 0 (o 0%) e 1 (o 100%). L'equidistribuzione, ovvero quando tutti gli individui hanno esattamente lo stesso reddito o la stessa ricchezza, corrisponde a un coefficiente pari a 0. L'uguaglianza perfetta è una linea diagonale retta con una pendenza di 1 (infatti l'area tra questa curva e sé stessa è pari a 0, quindi l'indice di Gini è 0). Se invece l'indice è pari a 1, significa che nella popolazione è presente massima concentrazione di disuguaglianza, dove una persona

⁴ Funzione gini() del software R. È stata disegnata e calcolata attraverso i dati del reddito disponibile netto dell'indagine analizzata nei capitoli successivi 4 e 5.

guadagna tutto il reddito (o detiene tutta la ricchezza) e tutte le altre nulla. L'indice di Gini graficamente rappresenta il divario tra la linea di uguaglianza e la curva di Lorenz.

Matematicamente si ha:

$$G_q = \frac{2}{n^2 \bar{y}} \sum_{i=1}^n i |y_i - \bar{y}| \quad \text{con } i = 1, \dots, n$$

dove y_i rappresenta il reddito dell'individuo i -esimo, \bar{y} rappresenta il reddito medio della popolazione e n indica il numero totale delle unità statistiche che formano il campione.

La formula può essere espressa anche in riferimento al grafico della Fig. 2.1 in cui si mostra la curva di Lorenz:

$$G_q = \frac{A}{A + B}$$

dove A è l'area tra la linea di equidistribuzione e la curva di Lorenz e $A+B$ rappresenta l'area totale sotto la linea di perfetta uguaglianza.

L'indice di Gini può anche essere usato per confrontare due nazioni o Paesi diversi per vedere quale ha più disuguaglianza. Inoltre, rispetta il principio di simmetria per cui è irrilevante l'identità di soggetti con le stesse caratteristiche, ovvero se un ricco e un povero con uguali caratteristiche si scambiano il reddito, l'indice non cambia. Infine, rispetta anche il principio di indipendenza dalla media (l'indice non varia se si moltiplicano i redditi per una stessa costante) e il principio di trasferimento secondo cui un indice diminuisce quando avviene un trasferimento progressivo, cioè un trasferimento di reddito da un individuo povero ad un altro ancora più povero.

2.5.2. Head count ratio o indice di incidenza

L'Head Count Ratio (indicato con H) viene definito anche indice di incidenza o diffusione della povertà. Esso rappresenta la proporzione di individui di una popolazione che si trova al di sotto della soglia di povertà (ovvero l'incidenza del fenomeno in questa popolazione). Il suo valore è dato dal rapporto tra il numero di famiglie o individui al di sotto della soglia di povertà e il totale delle famiglie o individui della popolazione considerata.

Matematicamente può essere calcolato con la formula:

$$H = \frac{q}{n}$$

dove q è numero totale di individui poveri identificati e n è la popolazione totale. Esso può assumere valori tra 0, il quale indica che nella popolazione non ci sono individui poveri, e 1, che invece stabilisce il caso estremo in cui tutta la popolazione sia povera. Questo indice, nonostante sia semplice nel calcolo, ha la criticità di non riuscire a quantificare il livello di povertà degli

individui in situazioni economiche al di sotto della soglia. Esso, infatti, non è sensibile (ovvero l'indice non cambia) al principio di trasferimento.

2.5.3. Income gap ratio o indice di intensità

Un altro indice importante è quello che misura l'intensità della povertà, l'income gap ratio. Indicato con I , esso misura la percentuale del mancato reddito degli individui considerati poveri rispetto alla soglia di povertà. In questo caso, l'indice misura la gravità della povertà per gli individui poveri, ma non per l'intera collettività.

Matematicamente la formula risulta:

$$I = \frac{g}{q}$$

dove q è il numero di individui il cui reddito è al di sotto della linea di povertà, $g = \sum_{i=1}^q g_i$ è il divario di povertà aggregato con $i = 1, \dots, q$

di cui $g_i = \frac{(p - y_i)}{p}$, con p che rappresenta la soglia di povertà e y_i reddito dell'individuo o nucleo familiare i -esimo.

Anche questo indice va da 0, nel caso in cui nessuno è sotto la soglia di povertà o se ciascuno ha un reddito uguale a p , a 1 se tutti i redditi delle persone povere sono uguali a zero. Anche I non è sensibile al principio di trasferimento e ignora quindi il grado di povertà individuale.

2.5.4. Poverty gap index

Anche questo indice, indicato con PG , esprime il divario di povertà. Viene definito come il rapporto tra il totale dei 'poverty gap' individuali riferiti alla soglia di povertà, e il numero totale N degli individui della popolazione di riferimento.

Matematicamente risulta:

$$PG = \frac{g}{N}$$

dove $g = \sum_{i=1}^q g_i$ è il divario di povertà aggregato con $i = 1, \dots, q$

di cui q è il numero di individui il cui reddito è al di sotto della linea di povertà, $g_i = \frac{(p - y_i)}{p}$ con p che identifica la soglia di povertà e y_i il reddito dell'individuo o nucleo familiare i -esimo, come per l'indice precedente.

Può assumere valore 0 se nella popolazione non ci sono poveri o se ciascuno ha un reddito uguale a p , e valore 1 se l'intera popolazione ha reddito nullo.

Il limite di questo indice è che esso non descrive l'effetto della disuguaglianza tra i poveri stessi, ovvero di come sono distribuiti i redditi tra i poveri, ma solo la media dei divari di povertà individuali. Come i due indici precedenti, infatti, anche PG non considera i trasferimenti progressivi di denaro. Il suo vantaggio è che fornisce una stima di quante risorse siano necessarie per portare le famiglie sotto la soglia di povertà al livello minimo stabilito.

2.5.5. Indice di Sen

L'indice di Sen combina l'incidenza e l'intensità del rischio di povertà, cioè di superare la soglia di povertà. La sua importanza si rivela nella sintetizzazione tra diffusione, intensità e disuguaglianza.

Matematicamente risulta:

$$S = H \times \left[1 - (1 - I) \times \left(1 - \frac{q}{q+1} G_q \right) \right]$$

ma poiché nelle applicazioni empiriche il numero di famiglie povere è abbastanza grande per cui $\frac{q}{q+1} \rightarrow 1$, si arriva alla formula:

$$S = H \times [I + (1 - I) \times G_q]$$

dove H è l'indice di incidenza, I è l'indice di intensità e G_q è l'indice di Gini delle persone a rischio di povertà.

Se G_q assume valore 0, ovvero i redditi sono distribuiti in modo perfettamente uguale tra tutti i soggetti, S è uguale al prodotto tra H ed I . Si ha invece $S = H$ nel caso in cui G_q sia 1 (situazione di massima concentrazione di disuguaglianza). Inoltre, l'indice di Sen si avvicina all'indice di diffusione H tanto più basso è il reddito dei poveri e all'indice di intensità I quanto più è grande la percentuale di poveri.

In questo caso l'indice di Sen è sensibile ai trasferimenti progressivi tra i poveri, infatti, attraverso questi, la disuguaglianza si riduce e l'indice decresce. In questo modo l'indice riesce a cogliere le differenze nel reddito tra i poveri.

2.5.6. Poverty gap squared

Questo è l'indice di gravità della povertà e viene calcolato come media ponderata dei divari di povertà individuali relativi al quadrato. Confrontandolo con l'indice P2 definito nel paragrafo 2.5.4., è importante sottolineare che PG assegna lo stesso peso ad ogni poverty gap individuale (cioè di $1/n$), mentre P2 dà un peso più ampio ai poverty gap individuali più grandi. Quindi P2, a differenza di PG, considera la disuguaglianza tra i poveri e rispetta il principio di trasferimento.

Matematicamente risulta:

$$P2 = \frac{g_2}{N}$$

dove $g_2 = \sum_{i=1}^q g_i^2$ è il quadrato dei divari di povertà individuali con $i = 1, \dots, q$

di cui q è il numero di individui il cui reddito è al di sotto della linea di povertà, $g_i = \frac{(p - y_i)}{p}$ con p che identifica la soglia di povertà e y_i il reddito dell'individuo o nucleo familiare i -esimo, come per l'indice precedente.

Questo indicatore può essere espresso in forma additiva come scomposizione di sottogruppi, ad esempio in base alle classi d'età o la condizione professionale o l'area geografica. Per ogni gruppo può essere poi calcolato lo specifico indicatore P2.

Capitolo 3

L'indagine della Banca d'Italia

In Italia, una delle fonti principali per lo studio della povertà è l'indagine sui bilanci delle famiglie italiane condotta dalla Banca d'Italia. Viene svolta dal 1962 con l'obiettivo di analizzare la ricchezza e i comportamenti economici e finanziari delle famiglie attraverso i dati sui loro redditi e risparmi. Dal 1987 l'indagine aveva cadenza annuale, mentre successivamente si è passati a cadenza biennale, con eccezione del 1998 e del 2020 (a causa della pandemia che l'ha rimandata di un anno).

I risultati di questa indagine sono stati condivisi anche per ricerche internazionali. Dal 2010 infatti, l'indagine dei bilanci delle famiglie dell'Eurosistema gestita dalla Banca Centrale Europea (Household Finance and Consumption Survey o HFCS) riceve i dati (adeguatamente armonizzati) dell'Italia da quella effettuata dalla Banca d'Italia e lo stesso vale per i database LIS (Luxembourg Income Study) e LWS (Luxembourg Wealth Study) che contengono dati armonizzati del reddito e della ricchezza.

In questa relazione finale verranno analizzati i dati dell'edizione 2020 dell'indagine, che ha subito alcune modifiche rispetto a quelle degli anni precedenti.

3.1. Il disegno campionario

Il campione obiettivo dell'edizione del 2020 comprende 6239 famiglie residenti in Italia, distribuite tra circa 350 e 450 comuni (dato delle ultime edizioni d'indagine), per un totale di 15198 individui. A partire dal 1989 all'interno di esso è stata inclusa anche una quota di unità, definite famiglie panel, già intervistate nelle edizioni precedenti e che nelle ultime indagini rappresenta il 50% del campione. Questa quota panel comprende una parte di famiglie presenti da almeno due rilevazioni e un'altra che viene estratta casualmente tra le famiglie dell'indagine immediatamente precedente. Vengono esclusi nuclei familiari illegalmente residenti sul territorio nazionale oppure che abitano in istituti (ad esempio conventi, ospedali, carceri, ...).

Il disegno di campionamento utilizzato è quello a due stadi, dove nel primo stadio le unità sono rappresentate dai comuni, mentre nel secondo dalle famiglie. Prima di procedere con la selezione nel primo stadio, durante la prima fase viene effettuata una stratificazione nei comuni basata sulla regione e sulla classe di ampiezza demografica, di cui se ne possono distinguere tre: fino a 20.000 abitanti, tra 20.000 e 40.000 abitanti e infine oltre 40.000 abitanti. I comuni selezionati all'interno di ogni strato sono i cosiddetti comuni auto-rappresentativi, ovvero tutti quelli con popolazione superiore ai 40.000 abitanti, e tra i comuni rimanenti quelli selezionati con la metodologia probability proportional to size (PPS), ovvero tramite l'assegnazione di una elevata probabilità di inclusione nel campione ai comuni con maggiori dimensioni. Durante la seconda fase avviene invece l'estrazione casuale delle famiglie non panel per l'intervista, a partire dalle liste anagrafiche. A partire dall'edizione 2020, viene inoltre effettuata una stratificazione basata sul reddito e l'indebitamento economico nelle unità di secondo stadio (le famiglie).

Ogni famiglia selezionata comprende tutte le persone che condividono almeno una parte del reddito percepito e che abitano solitamente nella stessa casa fino al 31 dicembre dell'anno dell'indagine. Tra questi sono compresi anche gli assenti temporanei (ad esempio per ferie, per motivi di lavoro o di studio, etc...) ed eventuali non parenti che però vivono stabilmente nell'abitazione fino alla data specificata in precedenza. Non sono invece considerati i neonati dell'anno successivo rispetto a quello dell'indagine e gli studenti fuori sede.

All'interno di ciascuna famiglia viene individuata una persona di riferimento (PR), che rappresenta il capofamiglia, cioè il responsabile dell'economia familiare e che detiene più informazioni a riguardo tra tutti i componenti della famiglia nell'anno in cui si indaga. Durante l'intervista sarà compito del PR fornire le risposte relative alla famiglia e a ciascun membro qualora qualcuno di essi sia assente.

3.2. Rilevazione dei dati

Per la rilevazione vengono usati dei questionari che si possono trovare nella pagina web "*Documentazione per l'utilizzo dei microdati*" della sezione dell'indagine nel sito ufficiale della Banca d'Italia. Sono strutturati secondo una parte dei fenomeni comuni a tutte le famiglie e un'altra per specifici sottoinsiemi di famiglie attraverso diversi allegati. A partire dal 1998 viene utilizzata la metodologia CAPI (Computer Assisted Personal Interviewing), ovvero un questionario elettronico svolto dalle famiglie con il computer, per oltre il 90% delle rilevazioni, mentre è rimasto il questionario cartaceo per la restante parte (anche definita metodologia PAPI, Paper-And-pencil Personal Interviewing). Dal 2020 si è scelto di rilevare i dati solamente attraverso il metodo CAPI, il quale favorisce la raccolta di dati come la geo-localizzazione e la durata dell'intervista, che risulta in media 50 minuti circa, ed è inoltre un metodo che riduce la necessità di verificare la coerenza nella fase di controllo successiva all'acquisizione dei dati. Anche prima del 2020, le interviste provenienti dal questionario cartaceo venivano comunque controllate prima di immettere i dati nel database attraverso il programma CAPI, in modo da ridurre gli errori.

Le famiglie non panel inoltre, sono selezionate da ISTAT dalle liste anagrafiche, e questi dati vengono poi riportati alla Banca d'Italia. Il criterio di estrazione è casuale e la loro partecipazione è volontaria. Le loro interviste vengono analizzate per confermare la validazione e annullate nel caso in cui sia presente un numero cospicuo di domande senza risposta. I microdati risultanti dai questionari sono poi pubblicati in forma anonima nel sito web della Banca d'Italia.

Un problema per le stime sono la mancanza di risposte, anche rispetto ad una sola componente delle variabili aggregate, che in questo modo non possono essere calcolate. Per risolvere questo problema si ricorre all'imputazione dei dati attraverso modelli di regressione che stimano valori da attribuire ai dati mancanti mediante informazioni prese da altre risposte. A questo dato stimato viene inoltre aggiunta una componente d'errore casuale che si distribuisce normalmente con media uguale a zero e varianza uguale a quella dei residui del modello usato. Un'altra complicazione dell'indagine è rappresentata dalle mancate interviste, e questo può portare a stime distorte del campione. Si è cercato di ridurre tale fenomeno attraverso una più accurata selezione delle famiglie ed una sensibilizzazione preliminare rispetto

all'importanza di tale indagine. In generale si è visto che maggiori sono il reddito e il grado d'istruzione della PR, maggiore è la difficoltà nell'ottenere l'intervista. Questo ha portato anche alla necessità di valutare l'attendibilità delle risposte, in particolare delle domande più delicate sull'aspetto finanziario della vita della famiglia, attraverso la contestualizzazione del nucleo rispetto agli elementi oggettivi che l'intervistatore ha a sua disposizione (l'immobile di residenza, gli arredi, etc...).

3.3. Le variabili aggregate

Le tabelle da 3.1 a 3.3 riportano i principali aggregati economici costruiti attraverso il questionario. Le variabili sono state aggregate per tre piani di aggregazione: il conto del reddito, il conto dell'utilizzazione del reddito e il conto del patrimonio.

Si specifica che poi nell'analisi svolta nel seguente progetto, verranno considerati solo i dati prendendo l'aggregato Y interamente. Si riportano per completezza d'informazione gli aggregati.

| Conto del reddito |
|---|
| <ul style="list-style-type: none"> Y: Reddito disponibile netto <ul style="list-style-type: none"> YL: Reddito da lavoro dipendente <ul style="list-style-type: none"> YL1: Retribuzioni nette YL2: Integrazioni non monetarie YT: Pensioni e trasferimenti netti <ul style="list-style-type: none"> YTP: Pensioni e arretrati <ul style="list-style-type: none"> YTP1: Pensioni YTP2: Arretrati YTA: Altri trasferimenti <ul style="list-style-type: none"> YTA1: Assistenza economica YTA2: Borse di studio YTA3: Assegni alimenti e regali/contributi ricorrenti <ul style="list-style-type: none"> YTA31: in entrata YTA32: in uscita (-) YTA4: trasferimenti Covid-19 (per l'edizione 2020) YTA5: reddito e pensione di cittadinanza (per l'edizione 2020) YM: Reddito netto da lavoro autonomo <ul style="list-style-type: none"> YMA1: Reddito netto da lavoro autonomo YMA2: Utili e altri redditi da partecipazione YC: Reddito da capitale <ul style="list-style-type: none"> YCA: Reddito da fabbricati <ul style="list-style-type: none"> YCA1: Affitti effettivi YCA2: Affitti imputati⁵ YCF: Reddito da capitale finanziario⁶ <ul style="list-style-type: none"> YCF1: Interessi attivi su depositi YCF2: Interessi attivi su titoli di Stato YCF3: Rendimenti altri titoli YCF4: Interessi passivi (-) |

⁵ Sono esclusi gli immobili usati per attività da lavoro dipendente.

⁶ Ottenuto attraverso il prodotto del tasso di interesse e lo stock di capitale.

| |
|---|
| <p>totale: $Y = YL + YM + YM + YC$</p> |
|---|

Tab. 3.1: Variabili aggregate nel conto del reddito.

| |
|---|
| Conto dell'utilizzazione del reddito |
| <p>Y: Reddito disponibile netto C: Consumi CD: Consumi durevoli CD1: Spese per mezzi di trasporto CD2: Spese per mobili, ecc... CN: Consumi non durevoli S: Risparmio</p> |
| totale: $Y = C + S$ |

Tab. 3.2: Variabili aggregate nel conto dell'utilizzazione del reddito.

| |
|--|
| Conto del patrimonio |
| <p>W: Ricchezza netta AR: Attività reali AR1: Immobili AR2: Aziende AR3: Oggetti di valore AF: Attività finanziarie AF1: Depositi AF2: Titoli di Stato AF3: Altri titoli AF4: Crediti commerciali e da altre famiglie PF: Passività finanziarie (-) PF1: Debiti verso banche e società finanziarie PF2: Debiti commerciali PF3: Debiti verso altre famiglie</p> |
| totale: $W = AR + AF - PF$ |

Tab. 3.3: Variabili aggregate nel conto del patrimonio.

3.4. Ulteriori modifiche dell'edizione 2020

L'edizione del 2020 ha avuto una modifica rispetto alla stratificazione del secondo stadio, come descritto in precedenza. Le famiglie non panel, infatti, sono state stratificate in 10 strati di reddito familiare, sia per il Nord-Ovest, Nord-Est, Centro, sia per Sud e isole, affinché migliorasse la rappresentatività del campione rispetto alla distribuzione del reddito. Gli stimatori prodotti quindi sono migliori di quelli delle

edizioni passate, ma essendo cambiato il disegno campionario è difficile fare dei confronti storici con le edizioni passate. Per questo motivo sono stati introdotti dei pesi per i confronti storici: PESO ovvero il peso campionario, e PESOPOP cioè il peso di riporto all'universo (esprime quante unità della popolazione vengono rappresentate da ciascuna unità del campione statistico). Inoltre, solo per l'edizione del 2020, sono stati introdotti il peso campionario PESOFIT, assieme al peso di riporto all'universo PESOFIT2 (si trovano nell'archivio annuale Carcom20), che però non possono essere usati per confronti con le indagini degli anni precedenti.

Capitolo 4

Dati trasversali dell'indagine 2020

In questo lavoro, come specificato nel paragrafo precedente, i dati sono stati presi dal sito ufficiale della Banca d'Italia che riporta i dati dell'indagine in totale in 29 file in formato .csv riferiti alle diverse sezioni del questionario. Per la costruzione del dataset su cui sono state fatte le iniziali analisi esplorative e per il calcolo degli indicatori di povertà, sono stati utilizzati principalmente tre dei vari file: "carcom20.csv" (riporta informazioni demografiche), "q20b.csv" (informazioni su lavoro e reddito) e "q20e.csv" (informazioni sui consumi).

Inoltre, è necessario specificare che in queste analisi viene considerata la famiglia come unità statistica per studiare la povertà in Italia nel 2020. Nell'indagine della Banca d'Italia però, i dati rispetto alle caratteristiche demografiche sono riferiti all'individuo. Per questa ragione le successive analisi esplorative e i confronti che verranno esposti in questo capitolo e nel capitolo 5, fanno riferimento alle caratteristiche del capofamiglia (che di seguito verrà indicato con PR, ovvero persona di riferimento). Infine, la variabile usata da Banca d'Italia per misurare il benessere economico è il reddito disponibile netto, ovvero la somma del reddito netto da lavoro dipendente, autonomo e da pensione, oltre ad eventuali trasferimenti e redditi netti da capitale reale e finanziario, di tutti i componenti⁷.

4.1. Analisi esplorativa

Una prima analisi riguarda il sesso del capofamiglia, inteso come il maggior percettore di reddito della famiglia. Come mostra la Fig. 4.1, il 68.75% delle PR è di genere maschile, mentre il 31.25% delle PR è di genere femminile⁸. È importante sottolineare questo aspetto sia perché questi dati riflettono la disuguaglianza di genere presente nella nostra società (tema che però non sarà qui oggetto di discussione), sia perché il gender gap è presente anche all'interno dei salari.

Più nello specifico, uno dei fattori che causa disuguaglianza di genere all'interno del mercato del lavoro è dato dalla discriminazione statistica, teoria che fa riferimento alle caratteristiche che nella mentalità media/comune vengono conferite alla donna, come ad esempio la bassa produttività, la maternità o il lavoro di cura (casa, figli) che è attribuito maggiormente alle madri. Tutte queste concezioni vengono implicitamente e storicamente attribuite alle donne, le quali si trovano in una condizione di svantaggio, e provocano una scarsità di capitale umano femminile e il gender pay gap, ovvero la differenza di salario rispetto agli uomini.

Secondo i dati Istat, nel 2020 il tasso di disoccupazione maschile in Italia per la fascia d'età 15-74 anni (cioè le persone attive) è di 8.6%, mentre quello femminile sale a

⁷ Glossario Banca d'Italia.

⁸ Si precisa che si rispetta il tipo di rilevazione di dati basata sulla concezione binaria del genere nelle due forme mutualmente esclusive di maschio e femmina. Nel questionario dell'indagine statistica del 2020 non sono infatti presenti altre forme di genere (sezione A del questionario).

10.4%. Il Gender Pay Gap viene calcolato da Eurostat e nel 2020 risulta essere di 4.2% in Italia (a fronte del 13.0% in media in Europa)⁹.

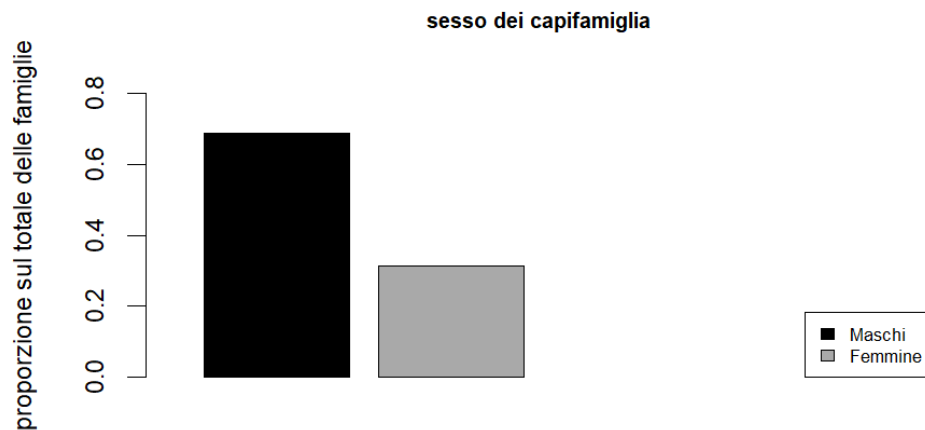


Fig. 4.1: Sesso dei capifamiglia.

Sulla base di queste considerazioni, una famiglia il cui reddito principale proviene da una PR donna sarà mediamente svantaggiata. Inoltre, come mostra la Fig. 4.2, il reddito medio netto disponibile è evidentemente minore per la PR femminile (stimato attraverso il dataset analizzato di 29358 euro annui) rispetto a quello maschile (stimato di 44961 euro annui).

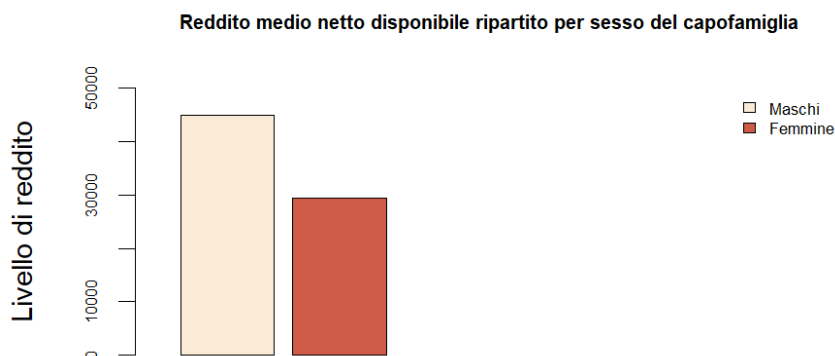


Fig. 4.2: Reddito medio netto disponibile per sesso del capofamiglia.

Per l'analisi rispetto all'età, viene usata come riferimento la variabile CLETA5 che indica le seguenti 5 classi di età: la prima fino a 34 anni, la seconda tra 35 e 44, la terza tra 45 e 54, la quarta tra 55 e 64, infine l'ultima oltre i 64 anni. I risultati riportati in Fig. 4.3 mostrano la classe 55-64 anni come quella con maggiore reddito medio disponibile netto, nello specifico stimato di 48590 euro nel 2020 in Italia, mentre quella inferiore risulta essere la classe più anziana over 64, che ha 31485 euro annui stimati nell'anno considerato. È interessante notare che anche la prima classe d'età, ovvero

⁹ Fonte: Eurostat, 2020.

quella più giovane, ha un reddito medio disponibile molto basso (si avvicina a quello dell'ultima classe che mediamente risulta la più povera), stimato di 32279 euro annui.

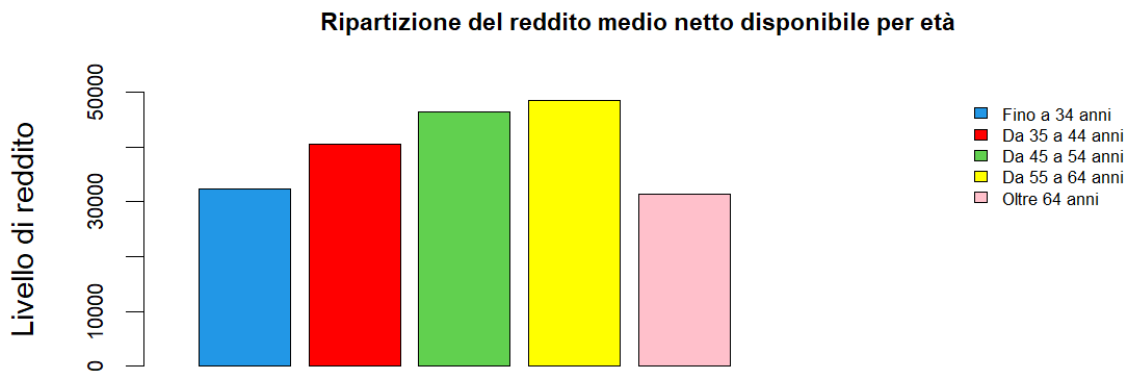


Fig. 4.3: Ripartizione del reddito medio netto disponibile per classe d'età.

Segue poi l'analisi rispetto al titolo di istruzione, in base alla variabile STUDIO che presenta le modalità: nessun titolo, licenza elementare, licenza media inferiore, raggruppate nel primo gruppo di modalità; diploma professionale (3 anni), diploma media superiore, che costituiscono il secondo gruppo; diploma universitario/laurea triennale, laurea/laurea magistrale e specializzazione post-laurea che formano il terzo gruppo. La Fig. 4.4 mostra come mediamente la variabile reddito e la variabile del titolo di studio abbiano legame positivo, ovvero ad un livello di istruzione superiore corrisponde un reddito mediamente maggiore.



Fig. 4.4: Ripartizione del reddito medio netto disponibile per titolo di studio.

Rispetto alla condizione professionale, quest'ultima variabile è costituita dalle modalità "dipendente", "autonomo", "pensionato" e "altro non occupato"¹⁰ e nella relazione con il reddito medio disponibile si hanno valori maggiori per i lavoratori autonomi (Fig. 4.5). Nello specifico vale 70962.79 euro annui stimati mediamente, mentre per il lavoratore dipendente 41278.64 euro annui mediamente. Per un pensionato invece risulta essere di 28961.2 euro.

Si specifica che per dipendente si intende operaio, impiegato o dirigente/direttivo; per lavoratore autonomo si intende imprenditore/libero professionista oppure altro autonomo. Infine, per altro non occupato si intende la parte della popolazione inattiva non pensionata oppure semplicemente senza occupazione (potrebbe essere anche in cerca di lavoro ad esempio).

Ripartizione del reddito medio netto disponibile per condizione professionale

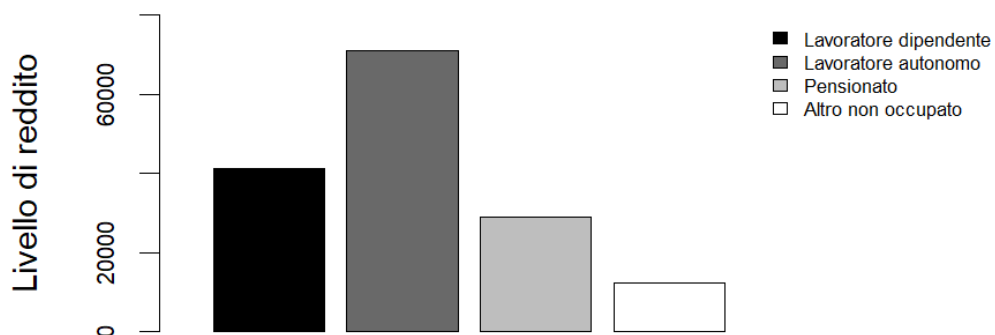


Fig. 4.5: Ripartizione del reddito medio netto disponibile per condizione professionale.

Infine, l'analisi si sposta sulla relazione tra il reddito medio disponibile e la ripartizione geografica facendo riferimento alla variabile AREA3 che divide il territorio italiano in tre zone diverse: Nord, Centro e Sud e Isole. Questi ultimi successivamente verranno indicati con Mezzogiorno. Come è evidente dalla Fig. 4.6, il Nord ha un reddito medio disponibile maggiore rispetto alle altre due zone.

Il divario economico e sociale tra Nord e Mezzogiorno rappresenta una questione delicata molto dibattuta in Italia, e che si è accentuata nell'ultimo decennio (2010-2020), secondo quanto scrive Banca d'Italia nel rapporto "*Divario Nord-Sud: sviluppo economico e intervento pubblico*"¹¹. In questo documento si analizza come il settore privato sia poco sviluppato nel Sud Italia, sia a causa di una minore digitalizzazione,

¹⁰ Sono state ricavate dalla variabile QUAL indicante lo status del lavoratore con le prime tre modalità (operaio, impiegato, dirigente/direttivo) raggruppante in "dipendente", le due successive (imprenditore/libero professionista, altro autonomo) raggruppate in "autonomo", infine "pensionato" e "altri non occupati" che coincidono con gli ultimi due raggruppamenti.

¹¹ Si specifica che questo documento non è oggetto di studio e di analisi nel presente progetto, ma si sono presi solo alcuni dati per una interpretazione migliore dei risultati trovati mediante l'elaborazione sui dati dell'Indagine sui bilanci delle famiglie italiane.

sia in termini di produttività. Per questo motivo, nel Mezzogiorno il settore pubblico ha un'importanza superiore rispetto al Nord e Centro. Allo stesso tempo però, la riduzione degli investimenti pubblici e le crescenti migrazioni dal Sud verso Nord, ostacolano il meridione nello stare al passo con il resto del nostro Paese. Inoltre, sempre secondo il rapporto della Banca d'Italia, tra il 2007 e il 2019 il PIL ha subito una caduta del 2% nel Centro-Nord e del 10% nel Mezzogiorno (anche causa della crisi finanziaria del 2008-2009).

In questo contesto, nel 2020 bisogna considerare l'avvento della pandemia che ha portato ad un calo del reddito del 5% nel Nord e del 8% nel Mezzogiorno, come riporta l'ISTAT nel rapporto *"Misure del Benessere equo e sostenibile dei territori"*¹². Questi dati vengono in parte confermati da quelli riportati da I.Stat riguardo la "Retribuzione lorda oraria per ora retribuita delle posizioni lavorative dipendenti in euro (mediana)"¹³ per i dipendenti del settore privato: si tratta di 12.45 euro/h nel Nord-ovest, 12.11 euro/h nel Nord-est, 11.58 euro/h nel Centro, 10.64 euro/h nel Sud e 10.88 euro/h nelle Isole.

Nella rielaborazione dei dati dell'indagine sui bilanci delle famiglie italiane qui oggetto di studio, si ha in particolare per il Nord un reddito medio familiare stimato di 47792.39 euro annui, per il Centro di 39948.3 euro annui, per il Mezzogiorno di 27447.65 euro annui (Fig. 4.6).

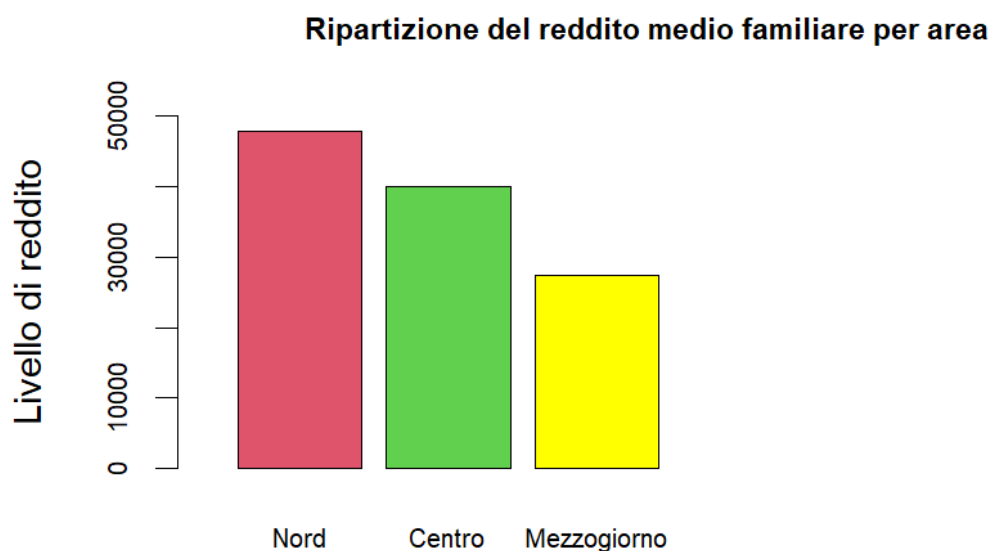


Fig. 4.6: Ripartizione del reddito medio netto disponibile per area geografica.

¹² Come nella nota 13, anche questo documento e i suoi contenuti non sono oggetto di analisi nel presente lavoro.

¹³ Dati presi dal sito ufficiale I.Stat, sezione "Lavoro e retribuzioni", più nello specifico "Retribuzioni orarie dei dipendenti del settore privato".

4.2. Calcolo dei redditi equivalenti e soglia di povertà relativa

Per proseguire con l'analisi bisogna considerare che ciascuna famiglia ha un numero di componenti specifico, e quindi a parità di reddito netto disponibile, non è possibile fare dei confronti senza considerare l'ampiezza familiare. È necessario quindi rendere equivalenti i redditi disponibili netti (la variabile Y) in base alla corrispondente dimensione familiare equivalente attraverso la scala OCSE-modificata (spiegata nel paragrafo 2.3), usata anche dalla Banca d'Italia stessa per la rielaborazione dei dati, e per questo motivo scelta anche nel presente progetto d'analisi. Si sono quindi ricavati prima i pesi di ciascuna famiglia (somma dei pesi individuali) passando per ogni individuo e riconoscendo i componenti di una stessa famiglia attraverso la variabile NQUEST (codice identificativo familiare). Si è poi trasformata la variabile Y in quella corrispondente del reddito equivalente dividendo ciascun reddito familiare (Y_i) per il corrispondente peso familiare.

La soglia di povertà relativa viene invece trovata calcolando il 60% della mediana del reddito disponibile netto equivalente¹⁴. Le famiglie relativamente povere sono quelle con il reddito equivalente inferiore a questa linea di povertà. Nello specifico, la soglia di povertà relativa nazionale stimata risulta essere di 7946.98 euro annui.

Questa soglia di povertà è moderatamente più alta rispetto alla soglia stimata da ISTAT sulla base dell'indagine sulle spese delle famiglie in Italia, pari a 7213.44 euro annui nel 2020. Bisogna tuttavia specificare che ISTAT utilizza metodologie diverse rispetto alla Banca d'Italia: l'unità di rilevazione è la famiglia e il benessere economico viene misurato con la spesa per i consumi. Inoltre, il reddito equivalente familiare viene ricavato con la scala Carbonaro con riferimento ad una famiglia di 2 persone (spiegata nel paragrafo 2.3.); infine, la soglia di povertà relativa è definita tramite l'ISPL, ovvero l'International Standard of Poverty Line, in base alla quale una famiglia di due componenti è povera se ha una spesa per consumi inferiore o uguale alla spesa media per consumi pro-capite¹⁵.

4.3. Stima degli indicatori di povertà unidimensionali

L'indice di Gini in Italia riportato da ISTAT per il 2020 è pari al 44.3%¹⁶, valore inferiore a quello ottenuto nella presente relazione, come riportato nella Tab. 4.1.

| Indici di povertà | Simbolo | Valore stimato |
|---------------------|---------|----------------|
| Indice di Gini | Gq | 46.8% |
| Indice di incidenza | H | 6.9% |
| Indice di intensità | I | 38.3% |

¹⁴ Si specifica che in questo calcolo si è presa la mediana del prodotto tra i redditi equivalenti di ciascuna famiglia e il peso campionario (indicato dalla variabile PESOFIT), affinché la soglia sia più realistica, perché riconducibile alla popolazione.

¹⁵ Analisi della povertà relativa, ISTAT.

¹⁶ La redistribuzione del reddito in Italia, ISTAT, 2020.

| | | |
|---------------------|----|------|
| Poverty gap index | PG | 2.6% |
| Indice di Sen | S | 4.6% |
| Poverty gap squared | P2 | 6.6% |

Tab. 4.1: Stima degli indicatori di povertà unidimensionali.

In entrambi i casi però, si tratta di un valore di disuguaglianza abbastanza alto (concentrazione di disuguaglianza di quasi il 50% tra i redditi disponibili netti delle famiglie del campione). Questa differenza è dovuta al fatto che ISTAT lo calcola in base al reddito primario, inteso come la somma del risultato di gestione, dei redditi da lavoro dipendente e dai redditi da capitale netti¹⁷, mentre Banca d'Italia usa il reddito disponibile netto.

L'incidenza di povertà è l'indicatore principale, poiché si misura la proporzione di famiglie povere sul totale. Sulla base dei dati di Banca d'Italia si ottiene un valore del 6.9%, cioè 7 famiglie ogni 100 sono povere, inferiore a quello stimato da ISTAT¹⁸ che riporta un'incidenza di povertà relativa familiare di 10.1% nel 2020, cioè circa una famiglia su 10 è povera. Anche per l'indice successivo, l'intensità di povertà relativa familiare, cioè la differenza media tra la soglia di povertà e il reddito o la spesa familiare, che misura il grado di povertà dei poveri, c'è questa discrepanza: le stime del presente lavoro riportano un valore di 38.3%, ovvero il reddito medio delle famiglie povere è 61.7% della soglia, mentre quelle di ISTAT riportano 21.4%, cioè la spesa media per i consumi delle famiglie povere è pari al 78.6% della soglia di povertà relativa ISTAT. Quello che sembra emergere da questi risultati e da questi confronti è che Banca d'Italia non riesca a cogliere le famiglie più povere: questo potrebbe avvenire perché misurare il benessere economico con la spesa dei consumi dà una misura di ciò che è realmente a disposizione della famiglia, tenendo conto anche dell'aumento dei prezzi. Banca d'Italia, utilizzando invece utilizzando il reddito monetario (o nominale, cioè la somma di denaro disponibile), non considera questa variazione.

L'indice PG è invece 2.6% ed esprime il divario di povertà nella popolazione, cioè la quantità di reddito che mediamente dovrebbe essere trasferita alle famiglie povere per arrivare almeno alla soglia di povertà.

L'indice di Sen ha valore 4.6%. Se tutte le famiglie povere fossero uguali, S sarebbe uguale a PG. Non essendo così, si può concludere che c'è disuguaglianza tra i redditi delle famiglie povere.

Infine, l'indice P2 vale 6.6% ed esprime la gravità della povertà.¹⁹

¹⁷ Glossario, ISTAT.

¹⁸ Indagine sulle spese delle famiglie 2020, ISTAT.

¹⁹ Per gli indici PG, S e P2 non ci sono le stime di ISTAT.

4.4. Confronti con Istat

In questo paragrafo si presentano dei confronti tra le stime dell'incidenza di povertà relativa (% di famiglie in povertà relativa) secondo ISTAT e quelle ricavate dai dati della Banca d'Italia del 2020, sulla base delle considerazioni emerse nei paragrafi precedenti.

4.4.1. Incidenza di povertà relativa rispetto alle classi d'età nel 2020

Le analisi evidenziate nel paragrafo 4.3. evidenziano una sottostima di indicatori di povertà sulla base dei dati di Banca d'Italia.

| Classi d'età | ISTAT | Presente relazione (dati Banca d'Italia) |
|----------------|-------|--|
| Fino a 34 anni | 12.9% | 9.3% |
| 35-44 anni | 14.8% | 11.4% |
| 45-54 anni | 11.9% | 9.6% |
| 55-64 anni | 8.2% | 7.5% |
| 65 anni e più | 7.6% | 4.3% |

Tab. 4.2: Confronto incidenza relativa rispetto all'età.

Come si può notare dalla Tab. 4.2, le stime percentuali di ISTAT sono superiori. La percentuale di povertà relativa più alta registrata in entrambe le indagini risulta essere quella della classe d'età 35-44 anni. Da notare inoltre, la differenza percentuale nella prima classe d'età, che per ISTAT risulta essere la seconda più povera relativamente, mentre per il presente lavoro risulta essere la classe 45-54 anni. In questo caso, i dati di ISTAT mostrano come il potere d'acquisto sia in realtà inferiore per la classe fino ai 34 anni, che ha redditi mediamente inferiori rispetto alle altre classi d'età (paragrafo 4.1), ma deve comunque sostenere le stesse spese per i consumi come tutti gli altri individui. La prima classe d'età è anche quella in cui le due stime differiscono di più (precisamente di 3.6 punti percentuali).

4.4.2. Incidenza di povertà relativa rispetto al livello di istruzione nel 2020

Come nel paragrafo 4.4.1., anche in questo caso le stime della presente relazione sono inferiori (Tab. 4.3).

| Titolo di studio | ISTAT | Presente relazione (dati Banca d'Italia) |
|--------------------------------------|-------|--|
| Nessun titolo di studio o licenza di | 14.9% | 13.7% |

| | | |
|-------------------------|-------|-------|
| scuola elementare | | |
| Licenza di scuola media | 14.4% | 13.1% |
| Diploma e oltre | 5.7% | 3.4% |

Tab. 4.3: Confronto incidenza relativa rispetto al titolo di studio.

In entrambe le indagini, ad un livello di istruzione inferiore corrisponde una percentuale di famiglie in povertà relativa maggiore. Infatti, come analizzato nel paragrafo 4.1, ad un inferiore titolo di studio corrisponde un reddito netto disponibile mediamente più basso. Interessante notare che la differenza percentuale nei primi due livelli tra le due indagini è circa uguale (nel primo livello si hanno 1.2 punti percentuali di differenza, mentre nel secondo livello si hanno 1.1 punti percentuali di differenza). Nel terzo livello invece, la differenza di punti percentuali è superiore (2.3 punti percentuali).

4.4.3. Incidenza di povertà relativa rispetto alla condizione professionale nel 2020

Come si evince dalla Tab. 4.4, le stime per la condizione professionale di dipendente differiscono maggiormente rispetto a quelle per i lavoratori autonomi o per i pensionati; infatti, nel primo caso si ha una differenza tra le stime di 4.9 punti percentuali, nel secondo caso di 2.4 punti percentuali e nel terzo caso di 2.6 punti percentuali. In questo confronto quindi, Banca d'Italia non riesce a cogliere appieno la povertà relativa dei lavoratori dipendenti italiani.

| Condizione professionale | ISTAT | Presente relazione (dati Banca d'Italia) |
|--------------------------|---|--|
| Indipendente /autonomo | 6.9% | 4.5% |
| Dipendente | 9.6% | 4.7% |
| Pensionato | 6.7% | 4.1% |
| Altri non occupati | 20.5% (inattivi non ritirati dal lavoro) 24.4% (in cerca di occupazione) | 57.1% |

Tab. 4.4: Confronto incidenza relativa rispetto alla condizione professionale.

Per quanto riguarda l'ultima condizione professionale, è difficile fare un confronto adeguato poiché ISTAT stima separatamente gli inattivi non pensionati e le persone in cerca di lavoro (oppure li stima assieme ai

pensionati). Si riportano comunque entrambe le stime di ISTAT per completezza di informazione.

4.4.4. Incidenza di povertà relativa rispetto all'area geografica nel 2020

Considerando la zona geografica (Tab. 4.5), la stima di incidenza di povertà relativa per il Nord è circa uguale tra le due fonti, mentre per il Centro differiscono anche se di poco (1.1 punti percentuali).

| Area geografica | ISTAT | Presente relazione (dati Banca d'Italia) |
|-----------------|-------|--|
| Nord | 2.9% | 3.0% |
| Centro | 2.1% | 3.2% |
| Mezzogiorno | 9.1% | 14.0% |

Tab. 4.5: Confronto incidenza relativa rispetto alla zona geografica.

Per il meridione invece la differenza è più ampia (4.9 punti percentuali). In questo caso le stime con i dati di Banca d'Italia sono maggiori rispetto a quelle di ISTAT. Questo aspetto riprende quanto descritto nel paragrafo 4.1, in cui si tratta il divario tra Nord e Sud, anche a livello salariale. In questo caso, infatti, Banca d'Italia riesce a cogliere maggiormente il divario poiché il benessere economico viene misurato con il reddito netto disponibile. Il reddito di Sud e Isole è mediamente inferiore rispetto a quello del Nord (par. 4.4.1) e quindi il calcolo dell'incidenza di povertà relativa ne risente, anche considerando il calo salariale avvenuto nel 2020 più alto nel meridione, piuttosto che nelle altre due aree. Inoltre, la vita nel Mezzogiorno costa meno rispetto al Nord; infatti, l'ISTAT riporta per il 2020 una spesa media mensile di 2523.38 euro nel Nord-ovest, di 2525.33 euro nel Nord-est, di 2510.51 euro nel Centro, di 1898.09 euro nel Sud e di 1949.00 euro nelle Isole²⁰. L'incidenza di povertà relativa minore nel Mezzogiorno secondo ISTAT potrebbe quindi, potrebbe essere dovuta al fatto che il benessere economico viene misurato con la spesa per i consumi.

4.5. Analisi specifica per area geografica

Per comprendere meglio i dati esposti nel paragrafo 4.4.4., sono state calcolate delle soglie di povertà relativa rispetto alle tre aree geografiche in cui è suddivisa l'Italia nell'indagine considerata. I risultati sono riportati in Tab. 4.6.

| Area geografica | Soglia di povertà relativa specifica | H stimato rispetto alla soglia specifica |
|-----------------|--------------------------------------|--|
| Nord | 10692.63 euro annui | 5.8% |

²⁰ Report di ISTAT sulla spesa per i consumi delle famiglie, 2020.

| | | |
|-------------|---------------------|------|
| Centro | 8153.352 euro annui | 3.7% |
| Mezzogiorno | 5623.536 euro annui | 6.0% |

Tab. 4.6: Stime della soglia di povertà relativa per ogni area e rispettiva incidenza di povertà relativa (H).

Questa analisi permette quindi di capire le differenze territoriali in termini di costo della vita, e allo stesso fornisce informazioni sull'incidenza di povertà relativa rispetto al tenore di vita specifico per quell'area. Si può notare innanzitutto che la soglia di povertà relativa del Nord Italia è molto più alta rispetto a quella nazionale (7946.98 euro annui), e questa differenza si nota anche rispetto al meridione che ha una soglia più bassa. Il Centro invece ha una soglia circa in linea con quella nazionale. Confrontando poi i dati con quelli riportati in Tab. 4.5, si può notare una lieve differenza di incidenza di povertà relativa per il Centro rispetto alla soglia specifica e a quella nazionale. Per il Nord invece l'incidenza di povertà calcolata rispetto alla soglia specifica aumenta di 2.8 punti percentuali, mentre per il Mezzogiorno la differenza aumenta ancora di più ma in termini opposti; infatti, H diminuisce di ben 8 punti percentuali. Si può quindi affermare che la povertà relativa del meridione vista in termini globali è accentuata in particolar modo dalle differenze reddituali con il Nord Italia, ma a livello locale non è così alta (come quella a livello nazionale), ed è addirittura equiparabile a quella specifica del Nord.

Capitolo 5

Analisi multidimensionali

Nel capitolo 4 si sono svolte delle analisi unidimensionali. Nel presente capitolo vengono invece proposte delle analisi per provare a concepire e interpretare il fenomeno della povertà in termini multidimensionali.

5.1. Analisi Fattoriale

In questo paragrafo si applica il metodo dell'Analisi Fattoriale per valutare la povertà in Italia nel 2020 mediante la percezione rispetto al livello di reddito, spesa e consumi riportata dai rispondenti nel questionario dell'indagine della Banca d'Italia. Si cerca quindi di quantificare la povertà attraverso delle dimensioni non direttamente osservate o latenti, cioè i fattori, interpretando delle variabili che hanno modalità qualitative ordinali essendo appunto riferite ad opinioni.

5.1.1. Variabili utilizzate

Per questa analisi, sono state utilizzate le seguenti variabili:

VARRED: vale 1 se, rispetto al reddito annuo che la PR pensava di percepire in un anno "normale", il reddito complessivo della sua famiglia nel 2020 è stato insolitamente alto, 2 se in linea, 3 se insolitamente basso, 4 se ha risposto "non so", 5 se non ha risposto.

VARREDFINE: vale 1 se, rispetto al reddito annuo che la PR percepirebbe in un anno "normale", il reddito complessivo della sua famiglia nel 2021 sarà insolitamente alto 1, 2 se in linea, 3 se insolitamente basso, 4 se ha risposto "non so", 5 se non ha risposto.

ASPRED: vale 1 se nel corso del 2021, la PR si aspetta che il reddito complessivo della sua famiglia cresca meno dei prezzi, 2 se circa quanto i prezzi, 3 se più dei prezzi, 4 se ha risposto "non so", 5 se non ha risposto.

VARCONS: vale 1 se la spesa nel 2020, rispetto a quanto la PR avrebbe pensato di spendere in un anno "normale", è stata insolitamente alta, 2 se normale, 3 se insolitamente bassa, 4 se ha risposto "non so", 5 se non ha risposto.

VARCONSALIM: vale 1 se per la PR, considerando la sola spesa per alimentari e altri beni essenziali sostenuta dalla sua famiglia al mese nel 2020, rispetto a quanto avrebbe pensato di spendere in un anno "normale", questa spesa è stata insolitamente alta, 2 se normale, 3 se insolitamente bassa, 4 se ha risposto "non so", 5 se non ha risposto.

Le successive 5 variabili sono poi riferite alla seguente domanda, sempre posta alla PR: "Nel corso del 2021 come si aspetta che varieranno i consumi della Sua famiglia nelle seguenti categorie di spesa rispetto ad un anno

normale?”. Ognuna di esse assume valore 1 se più bassi rispetto a un anno normale, 2 se uguali a un anno normale, 3 se più alti rispetto a un anno normale. Nello specifico sono le seguenti variabili:

ASP1: riferita a consumi di alimentari e altri beni essenziali,

ASP2: riferita ad acquisto di veicoli,

ASP3: riferita ad acquisto di abbigliamento e calzature,

ASP4: riferita a spese in alberghi e ristoranti,

ASP5: riferita ad acquisto di mobili ed elettrodomestici.

Infine, l'ultima variabile è CONDFIN: vale 1 se per la PR il reddito a disposizione della sua famiglia permette di arrivare alla fine del mese con molta difficoltà, 2 se con difficoltà, 3 se con qualche difficoltà, 4 se abbastanza facilmente, 5 se facilmente, 6 se molto facilmente.

Un problema riscontrato nella creazione del dataset è la mancanza di risposte. Per risolvere questo problema presente nelle variabili VARRED, VARREDFINE, ASPRED, VARCONS, VARCONSALIM, sono state eliminate tutte le righe del dataset con almeno un valore mancante (cioè di valore 4 o 5), in modo che non si abbiano stime distorte.

5.1.2. Metodi per il calcolo dei fattori

Per una prima analisi si guardano le correlazioni tra le variabili. Essendo variabili qualitative ordinali, la loro correlazione è stata calcolata con l'indice di Kendall. Un'alta correlazione è presente tra Varred e Varredfine (54%), e anche tra Varcons e Varconsalim (38%). Correlazioni più o meno importanti, come ci si può aspettare, sono presenti anche tra le variabili Asp1, ..., Asp5, in particolare tra Asp2 e Asp5 (55%), tra Asp3 e Asp4 (59%) e tra Asp4 e Asp5 (56%). Le variabili Aspred e Condfin invece non hanno una correlazione notevole con nessuna delle altre variabili.

Per determinare i fattori, si possono utilizzare due metodi: il metodo delle Componenti Principali, con cui si stimano tanti fattori quante sono le variabili, (quindi 11), e il metodo della Stima di Massima Verosimiglianza, con cui si possono stimare al massimo 6 fattori, poiché con 7 o più fattori si avrebbero gradi di libertà negativi della statistica test chi-quadro del test per l'adeguatezza del modello con m fattori.

Per quanto riguarda il primo metodo, per la scelta del numero di fattori si utilizzano tre criteri: lo screeplot, ovvero il grafico che indica il numero di fattori necessari guardando il punto di gomito (dove c'è il cambiamento di pendenza); la percentuale di varianza cumulata spiegata dai fattori e infine il numero degli autovalori della matrice di correlazione campionaria maggiori di uno.

Iniziando con il primo criterio²¹, si ha che il numero di fattori con una correlazione sufficientemente elevata con le variabili è 4 o 5, come mostra la Fig. 5.1.

Dal secondo criterio invece, emerge che 4 fattori sono sufficienti, infatti la percentuale di varianza spiegata risulta essere del 63%. Tuttavia, per una percentuale migliore, sarebbe più adeguato preferire un modello con 5 fattori (quasi 71% di varianza spiegata).

Con il terzo criterio la situazione cambia, infatti solamente i primi 4 fattori sono maggiori di 1. Il 5° non è maggiore di 1, ma è molto vicino (vale quasi 0.89).

In conclusione, con il primo metodo e basandosi sui risultati dei 3 criteri, un numero di fattori adeguato può essere 4. In ultima analisi, prima di passare al metodo della smv, viene calcolata la misura dell'errore totale. Per un modello con $m=4$ fattori, esso risulta essere di 0.78, mentre per un modello con $m=5$ fattori l'errore totale vale 0.58. Sulla base della misura dell'errore totale, si preferisce un modello con 5 fattori. Questo però andrebbe in contrasto con i risultati dei criteri precedenti, in particolare con il terzo che preferiva un modello con 4 fattori. Inoltre, per il principio del Rasoio di Occam, si preferisce in generale un modello più semplice e interpretabile. Di conseguenza, anche dopo questa analisi, viene confermato il modello con 4 fattori.

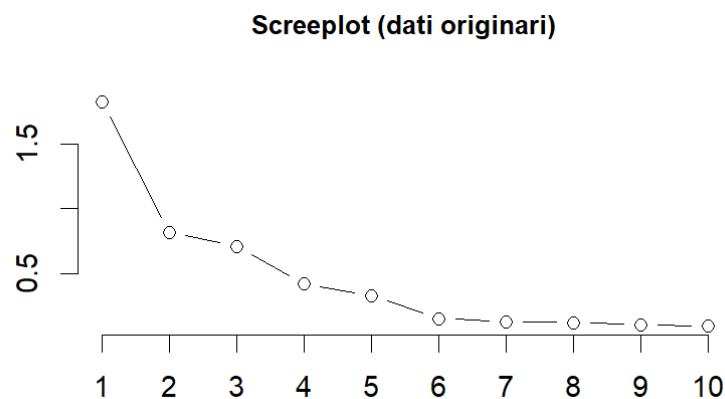


Fig. 5.1: Screepplot delle PCA ottenute con i dati originari.

Per quanto concerne il metodo della smv, in base al test di bontà di adattamento del modello, si andrebbe a preferire un modello con 5 fattori poiché il p-value risulta essere di 0.9%, che porterebbe a non rifiutare l'ipotesi nulla con livello di significatività di 0.005. Basandosi invece sulla proporzione della varianza cumulata spiegata, con un modello di 4 fattori risulta quasi del 50%, mentre con 5 fattori si raggiunge quasi il 52%. Tuttavia, la proporzione della varianza marginale del quinto fattore è circa del 2%, e se si scegliesse un modello con 5 fattori, si andrebbe in contraddizione con quanto concluso attraverso il metodo delle componenti principali. Quindi si conferma ancora

²¹ Non viene applicata la PCA standardizzata poiché le variabili sono qualitative e inoltre sono misurate su scale codificate con valori da 1 a 5.

una volta il modello con 4 fattori che, con il metodo smv, risulta avere una misura dell'errore totale di 0.18.

Senza rotazione dei loadings si ha la rappresentazione di figura 5.2:

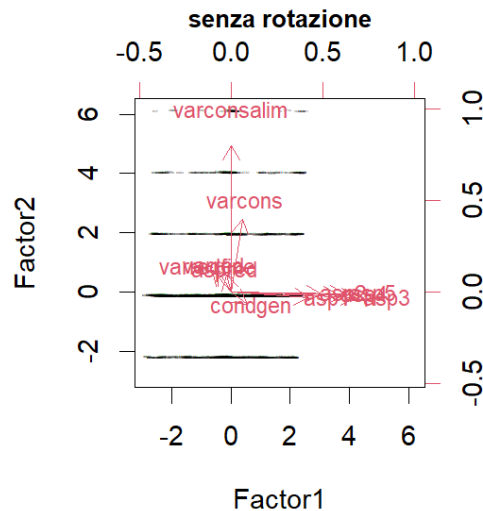


Fig. 5.2: Rappresentazione dei fattori senza rotazione.

Per una migliore interpretazione si determina la rotazione dei fattori, sia varimax (ortogonale), che promax (obliqua).

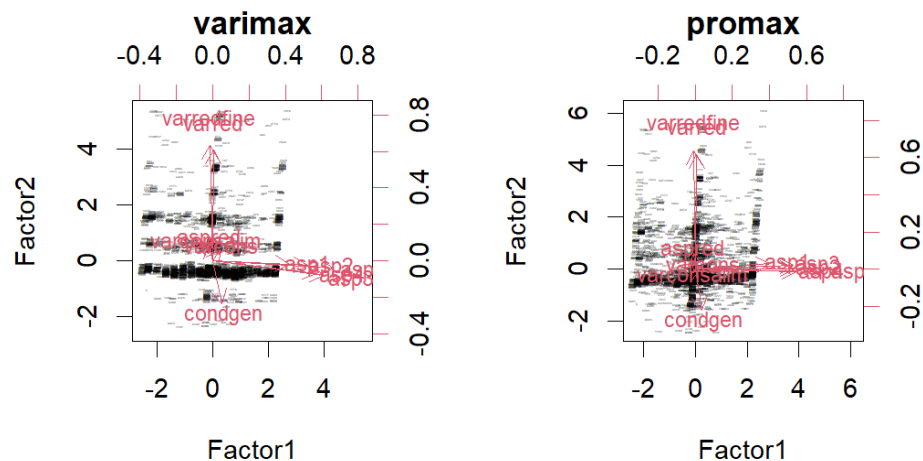


Fig. 5.3: Rappresentazione dei fattori con rotazione varimax e promax.

Con riferimento alle Fig. 5.2 e 5.3 (rotazione Varimax) e all'Appendice B, Varred e Varredfine risultano essere le variabili con i loadings più elevati nel secondo fattore con correlazione positiva, mentre il gruppo delle cinque variabili sui consumi (Asp1, ..., Asp5) è quello con i loadings più elevati nel primo fattore, sempre con correlazione positiva. Varcons e Varconsalim sono invece maggiormente rappresentate dal terzo fattore (a livello grafico non mostrato, ma risultante dall'analisi con il software R, si veda Appendice C) con correlazione positiva. La variabile Aspred è invece positivamente e poco correlata con il fattore 2 e il fattore 3. Non stupisce che Varcons, Varconsalim e Aspred siano correlate con il terzo fattore dal momento che tutte e tre sono

variabili riguardanti la spesa di una famiglia. Nel fattore 2 è inoltre presente una lieve correlazione positiva con Varconsalim e una negativa con Asp3. La variabile Condgen invece è spiegata solo dal fattore 2 con cui ha correlazione negativa (infatti nel grafico ha direzione negativa). Il quarto fattore spiega la restante parte delle variabili Asp2, Asp3, Asp4 e Asp5 e ha correlazione negativa con Asp2 e Asp5.

Si specifica inoltre che per il calcolo dei punteggi fattoriali (le stime campionarie dei fattori) è stato utilizzato il metodo della regressione.

In conclusione, questa applicazione ha permesso di trovare 4 fattori latenti, a cui, in base alle analisi svolte e alla correlazione tra le variabili e i fattori, si possono assegnare dei nomi. F1 è fattore di consumi1 (inteso come il primo tra i due fattori di consumo) poiché rappresenta una dimensione di percezione dei consumi che si avranno nel 2021 e più è alto il suo valore, più esprime una spesa alta per consumi familiari. F2 è fattore di percezione del livello di reddito del 2020 poiché è correlato principalmente con variabili che riguardano il livello di reddito in seguito alla pandemia. È importante notare che la variabile Condgen ha correlazione negativa con F2 perché le modalità seguono un andamento decrescente, ovvero sono ordinate dalla difficoltà più grave a quella meno grave. Quindi, ad un valore basso di Condgen, corrispondono valori alti delle variabili Varred e Varredfine che esprimono la percezione di avere un reddito basso rispetto al costo vita. Nel complesso quindi, un'interpretazione che si può dare è che nel 2020 ad una percezione di livello basso di reddito, corrisponde una grave difficoltà nell'arrivare a fine mese. Essendo poi correlato lievemente e positivamente anche con Varconsalim, ciò sembrerebbe indicare che ad una percezione di bassa spesa nel 2020 (mensile e annuale) corrisponde un basso reddito nel 2020 rispetto all'anno precedente la pandemia. F3 è fattore di percezione di spesa, poiché riflette soprattutto Varcons e Varconsalim. Infine, F4 è fattore di consumi2 (inteso come il secondo tra i due fattori di consumi), dato che è correlato solo con una parte del gruppo delle variabili sulla percezione di consumi su veicoli, abbigliamento e calzature, alberghi e ristoranti, mobili ed elettrodomestici. La relazione negativa con Aspred2 e Aspred5 indica che in seguito alla pandemia ad un maggiore acquisto di vestiti oppure spese turistiche, corrisponde un minore acquisto di veicoli oppure di mobili ed elettrodomestici.

5.1.3. Analisi dei punteggi fattoriali per area geografica

È interessante analizzare lo scatter-plot in due dimensioni delle stime dei fattori, ripartiti per area geografica (Nord, Centro, Sud e Isole).

Scatter-plot dei punteggi fattoriali

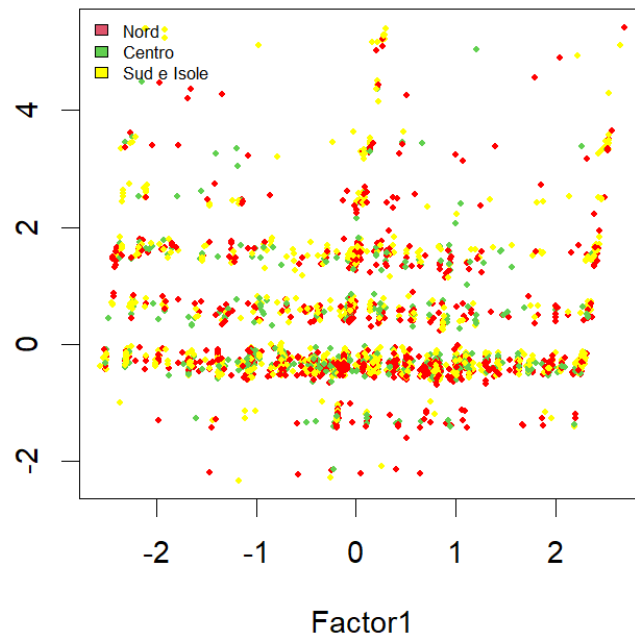


Fig. 5.4: Punteggi fattoriali distinti per area geografica.

Tra i gruppi definiti per area geografica sono presenti delle differenze relative ai punteggi fattoriali, come mostrato in Fig. 5.4. In particolare, si può notare come il Nord sia più concentrato nel fattore 1 che riguarda le variabili dei consumi ed esprime un livello di spesa alta. Una possibile spiegazione potrebbe essere collegata all'ammontare di reddito medio familiare che, come visto nel paragrafo 4.1, a riguardo l'analisi del reddito medio disponibile rispetto all'area geografica (Fig. 4.6), è presente in misura maggiore nel Nord Italia. Con una quantità di reddito più alta, si potrebbe essere portati a consumare di più e quindi il Nord Italia potrebbe avere maggiore sensibilità riguardo i consumi, ed aver notato quindi la differenza in questo tipo di spesa a causa della pandemia. Questa interpretazione si basa sia sulla teoria microeconomica, cioè ad un aumento di reddito corrisponde uno spostamento verso destra del vincolo di bilancio (quindi un aumento del potere d'acquisto), sia sulla teoria keynesiana del consumo, secondo cui ad un aumento del reddito corrisponde un aumento di consumo, infatti la funzione del consumo C è espressa con $C = C_0 + c(Y)$ in cui C_0 rappresenta il consumo di sussistenza (cioè quello che qualsiasi persona ha a prescindere dal reddito, il consumo che serve per "sopravvivere") e Y il reddito.

Anche il Centro ha un andamento simile al Nord, e risulta meno sensibile al fattore 2 (percezione del livello di reddito). La terza area geografica, ovvero il Mezzogiorno, risulta invece più sensibile al fattore 2 rispetto alle altre due aree geografiche. Sempre con riferimento alla Fig. 4.6, questa parte d'Italia ha un livello di reddito medio familiare inferiore; quindi, la percezione sul livello di reddito basso è più forte rispetto alle altre due aree.

Scatter-plot 3d dei punteggi fattoriali

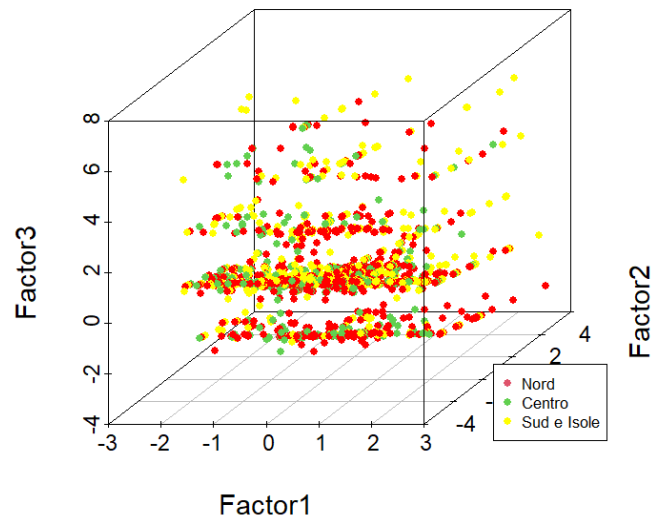


Fig. 5.5: Scatter-plot 3d dei punteggi fattoriali per area geografica.

Nella rappresentazione tridimensionale in riferimento alla Fig. 5.5, il fattore 3 sembra rilevante per tutte e tre le aree, anche se il Mezzogiorno risulta leggermente più sensibile. Una possibile interpretazione si può dare osservando che essendo questa parte d'Italia più sensibile al livello di reddito nel 2020 (fattore 2), lo è anche sul livello di spesa dello stesso anno (fattore 3), in quanto, come spiegato nel capitolo 4, ad un reddito basso corrisponde un livello più basso di spesa. Si tratta però di una differenza minima rispetto alle altre due aree, come mostrato in figura e quindi si può concludere che il fattore 3 rifletta tutte le aree.

5.2. Modello di regressione lineare multipla

Nel presente paragrafo viene stimato il modello di regressione lineare multipla ai punteggi fattoriali di ogni fattore del modello identificato nel paragrafo precedente. Le variabili indipendenti sono quelle utilizzate anche nell'analisi esplorativa del paragrafo 4.1 rispetto al reddito medio, ad eccezione della variabile "Condizione professionale" che non è stata inserita nel modello poiché si tratta di un'analisi sulla famiglia e questa è una variabile specifica dell'individuo (la PR). L'obiettivo è capire come le variabili (indipendenti) di caratteristiche demografiche del capofamiglia influenzano ciascun fattore di povertà (variabile dipendente), attraverso la stima dei loro coefficienti β_i .

Importante specificare che il valore dell'intercetta è da intendere come quello che gli scores assumono quando tutte le variabili indipendenti sono uguali a 0.

Per completezza di interpretazione viene riportata nell'Appendice C la legenda rispetto alle variabili indipendenti nel questionario somministrato dalla Banca d'Italia alle famiglie del campione.

5.2.1. Stima dei parametri

Si analizza quindi la significatività sul singolo coefficiente, ovvero si testa l'ipotesi per cui il coefficiente è uguale a zero.

Si specifica che per ognuna delle seguenti tabelle la significatività viene indicata con: '***' se $p\text{-value} < 0.001$, '**' se $p\text{-value} < 0.01$, '*' se $p\text{-value} < 0.05$, '.' se $p\text{-value} < 0.1$ e '' se $p\text{-value} < 1$.

Fissando $\alpha = 0.01$, si rifiuta H_0 , ovvero si rifiuta l'ipotesi che il coefficiente sia uguale a zero e quindi non significativo, per ogni α osservato minore di α fissato.

Scores del Fattore 1:

| Parametro (β) | Stima β | Std. Error (β) | Statistica test t | p-value | Sign. |
|-----------------------|---------------|------------------------|-------------------|----------|-------|
| Intercetta | -0.245604 | 0.073317 | -3.350 | 0.000813 | *** |
| Sesso | -0.013532 | 0.025694 | -0.527 | 0.598444 | |
| Classe d'età | 0.020643 | 0.010348 | 1.995 | 0.046114 | * |
| Titolo di studio | 0.026404 | 0.006573 | 4.017 | 5.95e-05 | *** |
| Area geog. | 0.030211 | 0.013554 | 2.229 | 0.025854 | * |

Tab. 5.1: Stime dei parametri del modello di regressione lineare multipla per i punteggi fattoriali del primo fattore.

Per i punteggi fattoriali del primo fattore risultano significative l'intercetta che entra negativamente nel modello, e la variabile del titolo di studio che entra invece positivamente nel modello, come mostrato in Tab. 5.1.

Concentrandosi solo su quella significativa si ha che la stima campionaria del fattore 1 teoricamente aumenta di circa 0.026 per ogni livello di istruzione. Vale poi, in termini teorici, -0.246 (stima dell'intercetta) quando tutte le altre variabili indipendenti sono uguali a 0.

Una possibile interpretazione, considerando il significato di F1 riportato nel paragrafo 5.1.2, potrebbe essere che i rispondenti con un livello di istruzione più alto pensino che nel 2021 avranno una spesa per i consumi mediamente più alta in seguito alla pandemia.

Scores del Fattore 2:

| Parametro (β) | Stima β | Std. Error (β) | Statistica test t | p-value | Sign. |
|-----------------------|---------------|------------------------|-------------------|---------|-------|
|-----------------------|---------------|------------------------|-------------------|---------|-------|

| | | | | | |
|------------------|---------------|----------|---------|----------|-----|
| Intercetta | 0.400278 | 0.067023 | 5.972 | 2.47e-09 | *** |
| Sesso | 0.100100 | 0.023488 | 4.262 | 2.06e-05 | *** |
| Classe d'età | - 0.096481 | 0.009460 | -10.199 | < 2e-16 | *** |
| Titolo di studio | - 0.064169 | 0.006008 | -10.680 | < 2e-16 | *** |
| Area geog. | 0.079120 | 0.012390 | 6.386 | 1.83e-10 | *** |

Tab. 5.2: Stime dei parametri del modello di regressione lineare multipla per i punteggi fattoriali del secondo fattore.

Per i punteggi fattoriali del secondo fattore risultano significative l'intercetta che entra positivamente nel modello, e in questo caso tutte le variabili indipendenti del modello, di cui sesso e area geografica entrano positivamente nel modello, mentre le altre (età e titolo di studio) negativamente, come mostrato in Tab. 5.2.

In termini teorici, se tutte le variabili indipendenti valgono 0, allora la stima campionaria del fattore 2 vale 0.400. Per l'interpretazione del modello rispetto alle altre variabili si ha che teoricamente aumenta di 0.100 se il capofamiglia è di genere maschile, del doppio di 0.100 se femminile; aumenta di 0.079 per ogni diversa area geografica; infine diminuisce di 0.096 ad ogni classe d'età e di 0.064 ad ogni livello di istruzione.

Una possibile interpretazione è la seguente: le persone più giovani hanno la percezione di avere un reddito basso, mentre le persone più adulte o anziane percepiscono di avere un reddito più alto; infatti, più aumenta la classe d'età più diminuisce lo score del F2 che a sua volta aumenta più la percezione è quella di avere un reddito basso. Stessa cosa per il titolo di studio: man mano che aumenta il livello di istruzione, la stima di F2 diminuisce. Per quanto riguarda le correlazioni positive, le famiglie del Sud e Isole e le PR donne piuttosto che uomini, sentono di avere un reddito basso a causa del Covid-19.

Scores del Fattore 3:

| Parametro (β) | Stima β | Std. Error (β) | Statistica test t | p-value | Sign. |
|-----------------------|---------------|------------------------|-------------------|----------|-------|
| Intercetta | - 0.176519 | 0.078011 | -2.263 | 0.023686 | * |
| Sesso | 0.088796 | 0.027339 | 3.248 | 0.001169 | ** |
| Classe d'età | 0.001310 | 0.011011 | 0.119 | 0.905321 | |

| | | | | | |
|------------------|---------------|----------|--------|----------|-----|
| Titolo di studio | - 0.011053 | 0.006993 | -1.580 | 0.114047 | |
| Area geog. | 0.055987 | 0.014422 | 3.882 | 0.000105 | *** |

Tab. 5.3: Stime dei parametri del modello di regressione lineare multipla per i punteggi fattoriali del terzo fattore.

Per i punteggi fattoriali del terzo fattore risultano significative le variabili del sesso e dell'area geografica che entrano positivamente nel modello, in riferimento a Tab. 5.3.

Come per le precedenti, si può dire che il valore teorico di F2 aumenta di 0.089 se il capofamiglia è di genere maschile, del doppio di 0.089 se femminile; aumenta di 0.056 per ogni diversa area geografica.

Quindi, le PR donne e le famiglie provenienti dal meridione percepiranno un livello di spesa bassa nel 2020 rispetto ad un anno "normale" (senza Covid-19).

Scores del Fattore 4:

| Parametro (β) | Stima β | Std. Error (β) | Statistica test t | p-value | Sign. |
|-----------------------|---------------|------------------------|-------------------|----------|-------|
| Intercetta | - 0.175330 | 0.057188 | -3.066 | 0.00218 | ** |
| Sesso | - 0.022565 | 0.020042 | -1.126 | 0.26026 | |
| Classe d'età | - 0.002074 | 0.008072 | -0.257 | 0.79723 | |
| Titolo di studio | 0.032875 | 0.005127 | 6.413 | 1.54e-10 | *** |
| Area geog. | 0.030613 | 0.010572 | 2.896 | 0.00380 | ** |

Tab. 5.4: Stime dei parametri del modello di regressione lineare multipla per i punteggi fattoriali del quarto fattore.

Per i punteggi fattoriali del quarto fattore risultano significative l'intercetta che entra negativamente nel modello, e le variabili del titolo di studio e dell'area geografica che entrano positivamente nel modello, come mostrato in Tab. 5.4.

Infine, la stima teorica del fattore 4 vale -0.175 se tutte le variabili indipendenti sono uguali a 0. Rispetto alle altre variabili significative si può dire che il suo

valore teorico aumenta di 0.033 per ogni livello di istruzione e di 0.031 per ogni diversa zona geografica.

Per l'interpretazione si ha che più aumenta il livello di titolo di studio, più si pensa che nel 2021 si spenderà maggiormente per i consumi familiari rispetto ad un anno precedente la pandemia. Stessa cosa vale per il Sud Italia e le Isole.

5.2.2. Interpretazione del modello

Di seguito vengono riportati i risultati del test F per la validità del modello, ovvero la bontà di adattamento del modello nel complesso. Questo test va a verificare l'assunzione di linearità del modello di regressione lineare multipla, per capire se esiste almeno una delle variabili indipendenti legata linearmente alla variabile dipendente, cioè se almeno uno dei coefficienti β_i è diverso da zero (ipotesi alternativa H1). Per testare l'ipotesi si ricorre all'analisi della varianza. In questo caso si ha nel p, ovvero il numero di parametri, è uguale a 6 (con l'intercetta), e n è il numero di famiglie (6239). Si rifiuta H0 se la F osservata è maggiore della F con $p - p_0 = 4$ (dove con p_0 si intende un parametro, cioè un modello con sola intercetta), e $n - p = 6176$, gradi di libertà. Ricordando che α fissato è 0.01, si ha che la F prima descritta vale 3.020181. Come si vede dalla Tab. 5.5, in tutti e quattro i modelli si rifiuta H0, ovvero si rifiuta l'ipotesi per cui tutti i coefficienti sono uguali a 0.

Viene riportato anche il coefficiente di determinazione R^2 che è un indice per misurare la capacità delle variabili indipendenti di spiegare la variabile risposta con il modello lineare e anche la forza del legame lineare tra la variabile dipendente e le variabili esplicative. Il suo valore varia tra 0 e 1. In questa analisi, in riferimento alla Tab. 5.5, i valori sono molto bassi: il modello di regressione lineare multipla spiega lo 0.34% di variabilità del fattore 1, il 4.1% di variabilità del fattore 2, lo 0.52% di variabilità del fattore 3 e infine lo 0.81% di variabilità del fattore 4.

| Punteggi fattoriali (scores) | R^2 | Stat. F osservata | g.d.l. | p-value |
|------------------------------|----------|-------------------|-----------|-----------|
| Factor 1 | 0.003381 | 5.239 | 4 6176 | 0.0003281 |
| Factor 2 | 0.04096 | 65.94 | 4 6176 | < 2.2e-16 |
| Factor 3 | 0.005233 | 8.122 | 4 6176 | 1.578e-06 |
| Factor 4 | 0.008079 | 12.58 | 4 6176 | 3.434e-10 |

Tab. 5.5: coefficiente di determinazione, test sull'analisi della varianza.

Si specifica che in questo caso l'interpretazione del modello in sé risulta poco utile, in quanto l'obiettivo non è vedere se il modello si adatta bene ai dati, ma piuttosto capire quali sono le variabili indipendenti rilevanti e importanti per ciascun fattore.

Capitolo 6

Conclusioni

I primi tre capitoli di questa relazione finale presentano una descrizione del fenomeno della povertà in letteratura, nello specifico dei tipi di povertà e della differenziazione tra povertà assoluta e relativa che qui è oggetto d'analisi. Segue poi una descrizione sugli strumenti per la misurazione della povertà: unità statistica, variabile di misurazione del benessere economico, scala di equivalenza, soglia di povertà relativa, indicatori di povertà. Nel capitolo 3 viene spiegata brevemente l'indagine sulle famiglie della Banca d'Italia da cui vengono presi i dati.

Nel capitolo 4 l'analisi si concentra sulla povertà relativa in Italia nel corso del 2020, ovvero sulla condizione di impoverimento di una famiglia rispetto al costo e agli standard di vita della popolazione italiana. Nelle analisi preliminari si è visto che mediamente la classe d'età più giovane e quella vicina al pensionamento o in pensione hanno un reddito più basso. Stessa cosa avviene per le famiglie con capofamiglia avente titolo di studio inferiore alla licenza media. In generale dalle analisi risulta che il reddito disponibile netto mediamente aumenta all'aumentare del titolo di studio e dell'età, fino poi a diminuire con la pensione. Per ciò che riguarda la condizione professionale, i lavoratori autonomi hanno un reddito mediamente più alto. Ci si è poi concentrati su divari reddituali dovuti in parte a delle problematiche strutturali della società: questa differenza è presente sia tra uomo e donna, sia tra Centro-Nord e Meridione. Si è visto che il reddito mediamente risulta minore per le donne (gender pay gap) e nel meridione (divario Nord-Sud). Si tratta quindi di differenze salariali non dovute alla quantità di esperienze lavorative o alle competenze tecniche che un determinato individuo possiede, ma piuttosto di criticità sociali che prescindono dall'individuo stesso (non si sceglie di nascere con un determinato genere oppure in una determinata area geografica).

Si è poi proceduto calcolando la soglia di povertà relativa di 7946.98 euro annui, i rispettivi redditi equivalenti per ciascuna famiglia e gli indicatori di povertà. In Italia nel 2020 risultano relativamente povere il 6.9% delle famiglie italiane, stima ricavata con i dati della Banca d'Italia. La stima è inferiore a quella di ISTAT perché il benessere economico viene misurato con una variabile diversa. In questo senso, anche per i successivi confronti tra le stime dell'indagine analizzata nel presente progetto e quelle di ISTAT, i risultati differiscono e quelli ottenuti da Banca d'Italia sono inferiori (rispetto alla classe d'età, al titolo di studio, alla condizione professionale), proprio perché ISTAT riesce a cogliere l'aumento dei prezzi avendo come unità di misura della povertà la spesa per i consumi, mentre Banca d'Italia tiene conto del reddito nominale o monetario e non di quello reale (cioè il reddito che include anche la variazione dei prezzi). Risultano più povere relativamente, come anticipato nelle analisi preliminari, le famiglie con basso livello di istruzione, con capofamiglia dipendente e della classe d'età 35-44 anni, seguita da quella fino ai 34 anni per ISTAT e 45-54 anni per Banca d'Italia. È interessante notare che rispetto all'analisi sul reddito medio annuo, la classe d'età più piccola risulta meno povera. Questo può essere dovuto anche alla pandemia che ha causato la perdita del lavoro e attraverso il lockdown la chiusura (temporanea o successiva alla pandemia) di molte attività lavorative, e la classe 35-44 che già partiva da un reddito medio stimato più basso delle due successive (si veda Fig. 4.3) può averne risentito.

Allo stesso tempo però ISTAT sottostima la povertà relativa rispetto all'area geografica. In questo caso Banca d'Italia porta a delle stime superiori proprio perché il reddito risulta un variabile di confronto (come si è visto nel capitolo 5) molto sensibile al divario tra le aree italiane, sia per il costo vita, sia perché i redditi a causa della pandemia hanno subito un calo generale, più accentuato nel meridione. Il Sud e le Isole risultano molto più povere relativamente a livello nazionale (11 punti percentuali in più rispetto al Nord e 10.8 punti percentuali in più rispetto al Centro) sulla base delle stime di questo lavoro. Se però si osserva la povertà relativa a livello locale, il Mezzogiorno risulta povero quasi quanto il Nord per gli standard di vita dell'area geografica considerata (e quindi la soglia di povertà relativa specifica). Ancora una volta, questi dati dimostrano quanto il divario tra Nord e

Mezzogiorno sia accentuato, e lo sia stato ancora di più con la pandemia. Questo implica che chi nasce in una zona svantaggiata, avrà, per causa naturale, un rischio maggiore di povertà.

Nel quinto capitolo viene svolta un'analisi multidimensionale sui dati, per provare a interpretare il fenomeno sotto un altro punto di vista. Dall'applicazione dell'analisi fattoriale risulta adatto un modello con 4 fattori latenti: il primo riguardante i consumi nel 2021, il secondo riguardante il livello di reddito percepito nel 2020, il terzo del livello di spesa percepita nel 2020 e infine il quarto ancora sui consumi del 2021. Questi fattori mettono in luce l'effetto della pandemia sulle opinioni e le vite delle famiglie considerate. In particolare, si è svolta un'analisi sulle differenze tra le tre aree geografiche: il Nord e il Centro sembrano essere più sensibili ai consumi e pensano che nel 2021 avranno una spesa per consumi più alta a causa della pandemia. Il Mezzogiorno invece è più sensibile al livello di reddito del 2020. Si può concludere quindi che il Nord e il Centro si percepiscono più poveri relativamente alla spesa per consumi nel 2021 dopo il Covid-19, mentre il meridione si sente povero relativamente al livello di reddito percepito nel 2020, sempre a causa della pandemia. Ritorna quindi il tema del divario tra Nord e Sud Italia.

Infine, nella seconda parte del capitolo 5, è stato applicato modello di regressione lineare multipla a ciascun punteggio fattoriale precedentemente identificato con variabili indipendenti genere, età, titolo di studio e area geografica: l'obiettivo è capire come queste variabili influenzino ciascun fattore. Ciò che emerge è che il titolo di studio e l'età influenzano molto la percezione sui consumi. Quindi, maggiore è il grado d'istruzione o maggiore è l'età, più si pensa che la spesa per i consumi aumenterà nel 2021, che si potrebbe interpretare come un timore sull'aumento dei prezzi a seguito del Covid-19. La variabile del sesso influenza invece la percezione sul livello di spesa nel 2020 (fattore 3). L'area geografica influenza il secondo, terzo e quarto fattore, quindi risulta rilevante.

Il risultato più rilevante di questa analisi riguarda il secondo fattore, ovvero la percezione sul livello di reddito nel 2020. Esso viene infatti influenzato molto da tutte le variabili. Ciò che sembra quindi preoccupare maggiormente le famiglie italiane in base a questa indagine, cioè il fattore che fa sentire maggiormente poveri, è il livello di reddito percepito. Questo è un risultato importante se contestualizzato nel periodo in cui stiamo vivendo, e potrebbe essere utile per possibili applicazioni di politiche pubbliche. La pandemia innanzitutto ha causato una crisi economica, oltre che sanitaria, accentuando così le disuguaglianze. A fronte di queste difficoltà, nel 2020 era stato inserito il Reddito di Emergenza come forma di assistenza verso 425 mila famiglie, le più bisognose colpite dalla crisi. Dal 2019 è inoltre in vigore il Reddito di Cittadinanza, che nel 2020 ha interessato circa 1 milione e 600 mila famiglie italiane²². Queste forme di assistenzialismo hanno sicuramente ridotto la povertà durante il periodo di pandemia e quello successivo. Il Rdc verrà tolto dal 1° gennaio 2024 e questa è stata una questione molto dibattuta negli scorsi mesi a causa delle problematiche e dei controversi aspetti che lo hanno riguardato (ad esempio l'esistenza di beneficiari che in realtà non rispecchiavano i requisiti e i mancati controlli, il mancato aiuto nel cercare un lavoro, etc...), considerando però che questa forma di sostentamento sia stata fondamentale per molte famiglie. Anche il salario minimo è una delle proposte presenti nel dibattito attuale italiano: potrebbe diventare una sicurezza rispetto a lavori sottopagati (ad esempio molti lavori stagionali), e in generale potrebbe portare ad una riduzione delle disuguaglianze reddituali trattandosi di una paga (oraria, giornaliera o mensile) fissa per tutti.

²² Report ISTAT, "Le diverse forme della disuguaglianza".

Appendice A

Correlazioni tra le variabili qualitative ordinali del dataset attraverso l'indice di correlazione di Kendall.

| | | | | | | |
|-------------|---------------|-------------|---------------|--------------|-------------|-------------|
| | varred | varredfine | aspred | varcons | varconsalim | |
| varred | 1.0000000000 | 0.54068085 | 0.0004530544 | 0.081545977 | 0.08932599 | |
| varredfine | 0.5406808452 | 1.00000000 | -0.0205988459 | 0.037279449 | 0.08893834 | |
| aspred | 0.0004530544 | -0.02059885 | 1.0000000000 | 0.002704308 | 0.06433520 | |
| varcons | 0.0815459773 | 0.03727945 | 0.0027043081 | 1.0000000000 | 0.38062345 | |
| varconsalim | 0.0893259945 | 0.08893834 | 0.0643351957 | 0.380623453 | 1.00000000 | |
| asp1 | -0.0069177535 | -0.02893850 | -0.0445647281 | -0.011972787 | -0.05257426 | |
| asp2 | -0.0367252797 | -0.05328312 | -0.0187636097 | 0.020207909 | -0.02280819 | |
| asp3 | -0.0444159295 | -0.08847617 | -0.0297555784 | 0.074681256 | -0.04702663 | |
| asp4 | -0.0558277935 | -0.07476553 | -0.0146000315 | 0.077667744 | -0.02396705 | |
| asp5 | -0.0530180539 | -0.07285042 | -0.0169603971 | 0.030934628 | -0.02541184 | |
| condgen | -0.1936990785 | -0.21601231 | 0.0569765882 | 0.063195241 | -0.04300245 | |
| | asp1 | asp2 | asp3 | asp4 | asp5 | condgen |
| varred | -0.006917753 | -0.03672528 | -0.04441593 | -0.05582779 | -0.05301805 | -0.19369908 |
| varredfine | -0.028938495 | -0.05328312 | -0.08847617 | -0.07476553 | -0.07285042 | -0.21601231 |
| aspred | -0.044564728 | -0.01876361 | -0.02975558 | -0.01460003 | -0.01696040 | 0.05697659 |
| varcons | -0.011972787 | 0.02020791 | 0.07468126 | 0.07766774 | 0.03093463 | 0.06319524 |
| varconsalim | -0.052574263 | -0.02280819 | -0.04702663 | -0.02396705 | -0.02541184 | -0.04300245 |
| asp1 | 1.0000000000 | 0.32640363 | 0.44674877 | 0.36226343 | 0.41106572 | -0.00318029 |
| asp2 | 0.326403629 | 1.00000000 | 0.43837232 | 0.44398850 | 0.54792950 | 0.03666886 |
| asp3 | 0.446748768 | 0.43837232 | 1.00000000 | 0.59285062 | 0.53586440 | 0.07472943 |
| asp4 | 0.362263432 | 0.44398850 | 0.59285062 | 1.00000000 | 0.55625286 | 0.12340081 |
| asp5 | 0.411065719 | 0.54792950 | 0.53586440 | 0.55625286 | 1.00000000 | 0.04045146 |
| condgen | -0.003180290 | 0.03666886 | 0.07472943 | 0.12340081 | 0.04045146 | 1.00000000 |

Appendice B

Analisi fattoriale con il metodo della smv, rotazione varimax, numero di fattori uguale a 4 (per le motivazioni espresse nel paragrafo 5.2.1.).

```
> fa4_var
```

```
Call:
factanal(x = dati_fatt, factors = 4, scores = "regression", rotation = "varimax")
```

```
Uniquenesses:
```

| | | | | |
|---------|------------|--------|---------|-------------|
| varred | varredfine | aspred | varcons | varconsalim |
| 0.417 | 0.373 | 0.974 | 0.745 | 0.005 |
| asp1 | asp2 | asp3 | asp4 | asp5 |
| 0.715 | 0.564 | 0.188 | 0.469 | 0.271 |
| condgen | | | | |
| 0.907 | | | | |

```
Loadings:
```

| | Factor1 | Factor2 | Factor3 | Factor4 |
|-------------|---------|---------|---------|---------|
| varred | | 0.759 | | |
| varredfine | | 0.789 | | |
| aspred | | 0.123 | | |
| varcons | | | 0.495 | |
| varconsalim | | 0.117 | 0.988 | |
| asp1 | 0.525 | | | |
| asp2 | 0.652 | | | -0.101 |
| asp3 | 0.764 | -0.102 | | 0.466 |
| asp4 | 0.713 | | | 0.118 |
| asp5 | 0.831 | | | -0.190 |
| condgen | | -0.292 | | |

| | Factor1 | Factor2 | Factor3 | Factor4 |
|----------------|---------|---------|---------|---------|
| SS loadings | 2.492 | 1.340 | 1.240 | 0.301 |
| Proportion Var | 0.227 | 0.122 | 0.113 | 0.027 |
| Cumulative Var | 0.227 | 0.348 | 0.461 | 0.488 |

```
Test of the hypothesis that 4 factors are sufficient.
The chi square statistic is 163.6 on 17 degrees of freedom.
The p-value is 5.18e-26
```

Appendice C

Legenda per le variabili indipendenti utilizzate nell'applicazione del modello di regressione lineare, paragrafo 5.1.2. Si specifica che queste caratteristiche demografiche sono riferite al capofamiglia.

SEX = sesso. Vale 1 se maschio, 2 se femmina.

CLETA5 = classe di età. Vale: 1 fino a 34 anni, 2 da 35 a 44 anni, 3 da 45 a 54 anni, 4 da 55 a 64 anni, 5 oltre 64 anni.

STUDIO = titolo di studio. Vale 1 se nessun titolo di istruzione, 2 se licenza elementare, 3 se licenza media inferiore, 4 se diploma professionale (3 anni), 5 se diploma media superiore, 6 se dipl. universitario/laurea triennale, 7 se laurea/laurea magistrale, 8 se specializzazione post-laurea.

QUAL = status del lavoratore. Vale: 1 se operaio, 2 se impiegato, 3 se dirigente/direttivo, 4 se imprenditore/libero professionista, 5 se altro autonomo, 6 se pensionato, 7 se altri non occupati.

AREA3 = area geografica. Vale: 1 se Nord, 2 se Centro, 3 se Sud o Isole.

BIBLIOGRAFIA

- Levitas, R. et al. (2007) The multi-dimensional analysis of Social Exclusion. Department of Sociology and School for Social Policy Townsend Centre for the International Study of Poverty and Bristol Institute for Public Affairs University of Bristol.
- Aku, P. S., Ibrahim, M. J., & Bulus, Y. D. (1997). Perspective on poverty alleviation strategies in Nigeria. In Proceedings of the Nigeria Economic Society Annual Conference on Poverty Alleviation in Nigeria (pp. 41-54)
- Chakravarty S., D'Ambrosio C. (2006), The measurement of social exclusion, *Review of Income and Wealth*, 52, 3, pp. 377-398.
- Mahembe, E. and M. Odhiambo, N. (2019) Foreign aid and poverty reduction: A Review of International Literature, Taylor & Francis. Edited by R. Read. *Cogent social sciences*.
- Lok-Dessallien, R. (1999) Review of Poverty Concepts and Indicators.
- Goedhart T., V. Halberstadt, A. Kapteyn, B. M. S. Van Praag (1977), The Poverty Line: Concept and Measurement, *Journal of Human Resources*, Vol. 12, pp. 503-520.
- Banca d'Italia - Documentazione per l'utilizzo dei microdati (bancaditalia.it). Banca d'Italia.
- Banca d'Italia - Distribuzione dei microdati (bancaditalia.it). Banca d'Italia.
- Uno (2021) Obiettivo 1: Porre fine ad ogni forma di povertà Nel Mondo, ONU Italia.
- Archivio della Commissione parlamentare d'inchiesta sulla miseria in Italia e sui mezzi per combatterla (1951 - 1954). Archivio della Commissione parlamentare d'inchiesta sulla miseria in Italia e sui mezzi per combatterla - L'Archivio storico della Camera dei deputati.
- Lee, Peter, and Peter Townsend. 1994. "A Study of Inequality, Low Incomes and Unemployment in London, 1985–92". *International Labour Review* 133 (5–6): 579–595.
- Cmi (2021) Amartya Sen's capability theory approach, CMI Business School.
- Sen A. *Commodities and capabilities*. New York: Oxford University Press, 1985.
- Carbonaro G. (1985), "Nota sulle scale di equivalenza", in Commissione di indagine sulla povertà e sull'emarginazione (a cura di), *Primo rapporto sulla povertà in Italia*, 153-159. Roma, Istituto Poligrafico e Zecca dello Stato.
- Metadata Glossary, Databank.
- Inequality - poverty gap - OECD data, the OECD.
- Eurostat, "European social statistics. Income, poverty and social exclusion: 2nd report. data 1994-1997", (2002 Edition), Office for Official Publications of the European Communities, 2002, Luxembourg
- d'Italia, B. (2022) N. 25 - il Divario Nord-Sud: Sviluppo Economico e Intervento Pubblico, Banca d'Italia.

Misure del Benessere equo e sostenibile dei territori (2022) Istat - Istituto Nazionale di Statistica.
Bes dei territori.

La redistribuzione del reddito in Italia - istat.it (2021) Istat - Istituto Nazionale di Statistica.

Le Spese per i consumi delle famiglie | Anno 2020 (2021) Istat - Istituto Nazionale di Statistica.

Johnson, R.A. and Wichern, D.W. Applied Multivariate Statistical Analysis. Pearson New International Edition.

Grigoletto, M., Ventura, L. and Pauli, F. (2017) Modello lineare: Teoria e applicazioni con r. Torino: Giappichelli.

RINGRAZIAMENTI

Vorrei dedicare qualche parola a tutti coloro che mi sono stati vicini in questo percorso di triennale e alle persone che ho incontrato nel mio cammino.

Ringrazio il mio relatore Omar Paccagnella per avermi guidata nell'analisi e nella scrittura di questo elaborato con molta elasticità, disponibilità, e suggerimenti pratici importanti.

Ringrazio i miei genitori e mio fratello per darmi la possibilità e il privilegio di studiare sostenendomi nelle mie scelte.

Ringrazio il Buddismo e i miei compagni di fede per aiutarmi a disegnare la mia vita.

Ringrazio l'Asu e tutte le persone che ne fanno parte, per avermi fornito una visione critica e cosciente del mondo e delle sue problematiche politiche, sociali ed economiche.

Ringrazio Angelo, Francesca, Geneviève, Gioia, Giuliano, Margherita, Marius, Martina, Sara, Samuele per esserci sempre, da anni, anche quando viviamo in parti diverse dell'Europa e del mondo.

Ringrazio casa Falloppio e in particolare Camilla e Ludovica, per avermi accolta con genuinità e amore nel mio anno da (quasi) fuorisede a Padova.

Infine, ringrazio la mia seconda famiglia, nata a lezione nei primi giorni di università, Alessia, Cecilia, Elisa, Erni, Gianmarco, Leonardo, Riccardo per continuare a crescere assieme a me a prescindere dai contesti e le città differenti che attraversiamo.