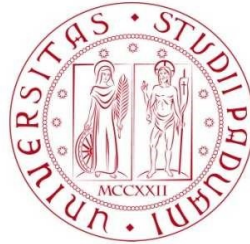


Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Triennale in  
Statistica per l'Economia e l'Impresa



Relazione finale

**SCOMPOSIZIONE DELLA DISUGUAGLIANZA DI MORTALITÀ:  
UN'APPLICAZIONE AI DATI ITALIANI A LIVELLO LOCALE**

Relatore Prof. Stefano Mazzuco  
Dipartimento di Scienze Statistiche

Laureando Andrea Gaspari  
Matricola 1198789

Anno Accademico 2021/2022



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Evoluzione della speranza di vita alla nascita.....</b>	<b>3</b>
<b>2 Scomposizione della disuguaglianza di mortalità .....</b>	<b>7</b>
<b>3 Risultati dell'analisi per le province italiane.....</b>	<b>17</b>
<b>3.1 Disuguaglianze nel sesso maschile .....</b>	<b>17</b>
<b>3.2 Disuguaglianze nel sesso femminile.....</b>	<b>24</b>
<b>Conclusioni</b>	<b>31</b>
<b>Bibliografia</b>	<b>33</b>
<b>Appendice</b>	<b>35</b>



# Introduzione

Il tema della mortalità è un elemento cardine nella demografia, in quanto gioca un ruolo determinante nella crescita di una popolazione: la sua descrizione rappresenta un mezzo per conoscere lo stato di salute di una popolazione.

L'aspettativa di vita alla nascita funge da buon indicatore per monitorare l'evoluzione della mortalità e dello sviluppo umano; infatti, con l'avvento del COVID-19, che ha inevitabilmente influenzato negativamente le dinamiche della mortalità, si sono registrati degli importanti declini in termini di speranza di vita, in maniera eterogenea anche all'interno di una stessa popolazione (Islam 2021).

La mortalità si presta bene a confronti temporali e geografici: negli ultimi anni tra i demografi c'è stato un consistente aumento dello studio della disuguaglianza di mortalità tra le popolazioni, ponendo meno attenzione sull'analisi del progresso generale.

In Italia il fenomeno della mortalità è stato oggetto di approfondite ricerche statistiche, condotte in modo continuativo e sistematico. Il primo report nazionale sulle disuguaglianze di mortalità per condizione socioeconomica fu pubblicato dall'Istat nel 1990 (Istat 1990), considerando fattori tra cui il sesso, l'età e l'istruzione. La stessa analisi venne condotta anche dieci anni dopo (Istat 2001). Altri studi hanno ricercato possibili relazioni tra la mortalità e aspetti socioeconomici, ad esempio la condizione professionale (Agabiti et al.) e la classe sociale (Caselli et al. 2011). Anche la relazione tra fattori sociali ed epidemiologici è stato oggetto di studio, focalizzandosi sulle cause di morte (Alicandro et al.).

Un obiettivo della statistica e della demografia rappresenta la descrizione di un fenomeno attraverso indicatori di sintesi. Può dunque risultare di interesse la possibilità di riassumere le disuguaglianze di mortalità. Nell'elaborato *'Decomposing dimensions of mortality inequality'* di Monica Alexander (2022) viene utilizzata una tecnica di scomposizione, basata sulla decomposizione a valori singolari, per estrarre delle ridotte dimensioni chiave utili per descrivere le diversità nella mortalità negli Stati Uniti d'America, distintamente per maschi e femmine.

L'obiettivo di questa relazione è svolgere la stessa analisi per le province italiane negli anni 2019, 2020 e 2021, in modo da comprendere dove risiedono le principali differenze di mortalità e capire come la pandemia ha influito su queste.

Di seguito si presentano brevemente i capitoli di cui si compone la relazione. Il primo introduce l'argomento di studio fornendo una visione generale sull'evoluzione dell'aspettativa di vita nel corso degli ultimi trent'anni, dove emergono differenze sostanziali tra le province, anche in relazione alla diffusione del Covid-19. Il secondo capitolo riguarda la presentazione teorica della tecnica utilizzata per l'analisi, partendo dalla decomposizione a valori singolari: si riportano i dati impiegati, la spiegazione del modello e le scelte eseguite. Infine, nell'ultimo capitolo si mostrano ed interpretano i risultati ottenuti, separatamente per maschi e femmine, analizzando anche possibili relazioni con la speranza di vita.

# Capitolo 1

## Evoluzione della speranza di vita alla nascita

L'aspettativa, o speranza, di vita alla nascita esprime il numero medio di anni che un neonato ha la probabilità di vivere se esposto ai tassi di mortalità specifici per sesso ed età al momento della sua nascita.

Negli ultimi secoli questo indicatore ha presentato un continuo aumento a livello mondiale, caratterizzato da una rapida crescita durante il secolo scorso, grazie ad una molteplicità di fattori, come il miglioramento dell'accesso e della qualità dei servizi sanitari, dell'educazione e dello stile di vita. Per questo motivo rappresenta un buon strumento per valutare lo stato di sviluppo di una popolazione e inoltre, riflette il livello complessivo della mortalità.

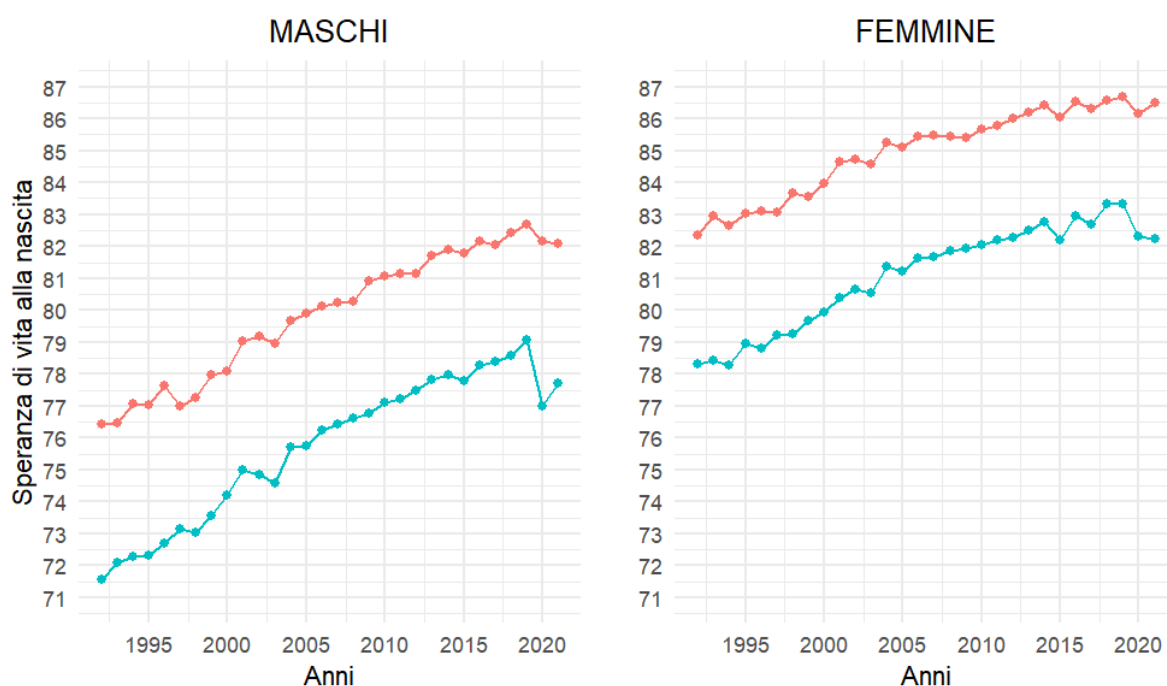
Anche in Italia vi è stata una progressiva crescita dell'aspettativa di vita alla nascita, per entrambi i sessi. Quanto appena espresso è visibile in Figura 1.1, dove sono raffigurati, dal 1992 al 2021, per maschi e femmine, i valori della *'best-practice life expectancy'* tra le province italiane, replicando l'idea del grafico di Oeppen-Vaupel (2002), aggiungendo a questa anche la *'worst-practice life expectancy'*; in altre parole, i valori della più alta e più bassa speranza di vita alla nascita tra le province italiane per ogni singolo anno.

Per quanto riguarda il sesso maschile, le province con valore massimo sono nella quasi totalità situate nel Centro Italia, in particolare in Toscana e nelle Marche, mentre la posizione geografica delle province con valore minimo ha subito una variazione nel tempo: fino al 2004 al Nord Italia, principalmente in Lombardia, e successivamente Napoli per quasi tutti gli anni. È da sottolineare il fatto che nel 2020 risulta Bergamo, una delle province più colpite dal Covid-19, ad avere il valore più basso.

Passando all'analisi del grafico si può notare che per entrambe le serie temporali maschili c'è stato un graduale incremento della speranza di vita alla nascita con un trend lineare; la *best-practice* passa da circa 76 a 82 anni, mentre la *worst-practice* da circa 71 a quasi 78 anni. Per tutto il periodo considerato quindi, la loro distanza si aggira approssimativamente sui 4-5 anni. Appare evidente la differenza delle conseguenze della pandemia sulla speranza di vita: la provincia con il valore minimo subisce una diminuzione di due anni tra il 2019 ed il 2020, arrivando ad un valore (77) perfino inferiore rispetto al 2010. Da sottolineare, però, che a questa forte decrescita è susseguito un significativo incremento l'anno successivo. Decisamente meno critico risulta invece l'impatto del Coronavirus per la provincia con l'aspettativa di vita maggiore, la quale ha però una variazione negativa anche nel 2021.

In merito al sesso femminile, le province che presentano i valori massimi sono perlopiù collocate nel Centro e nel Triveneto, mentre il caso opposto è costituito principalmente da Napoli; l'unica provincia del Nord presente è Vercelli, nel 2020, anche qui come chiara ripercussione del virus. Prima di approfondire il grafico, è interessante notare che i livelli della speranza di vita alla nascita nel 1992, sia per la migliore che per la peggiore provincia, sono all'incirca gli stessi raggiunti dai maschi nel 2021, dunque un divario di trent'anni. L'andamento dell'aspettativa di vita presenta dei tratti comuni con quello maschile, con la differenza che dal 2014 in poi le serie risultano più stazionarie. Anche in questa circostanza si nota un effetto negativo maggiore della pandemia nei confronti della provincia con la speranza di vita minore, con una decrescita di un anno rispetto al 2019. Una nota particolare rappresenta l'evoluzione dal 2020 al 2021, in cui i versi degli scostamenti sono opposti rispetto a quanto accadeva nell'altro sesso; infatti, la *best-practice* evidenzia una crescita, la *worst-practice* una lieve diminuzione.





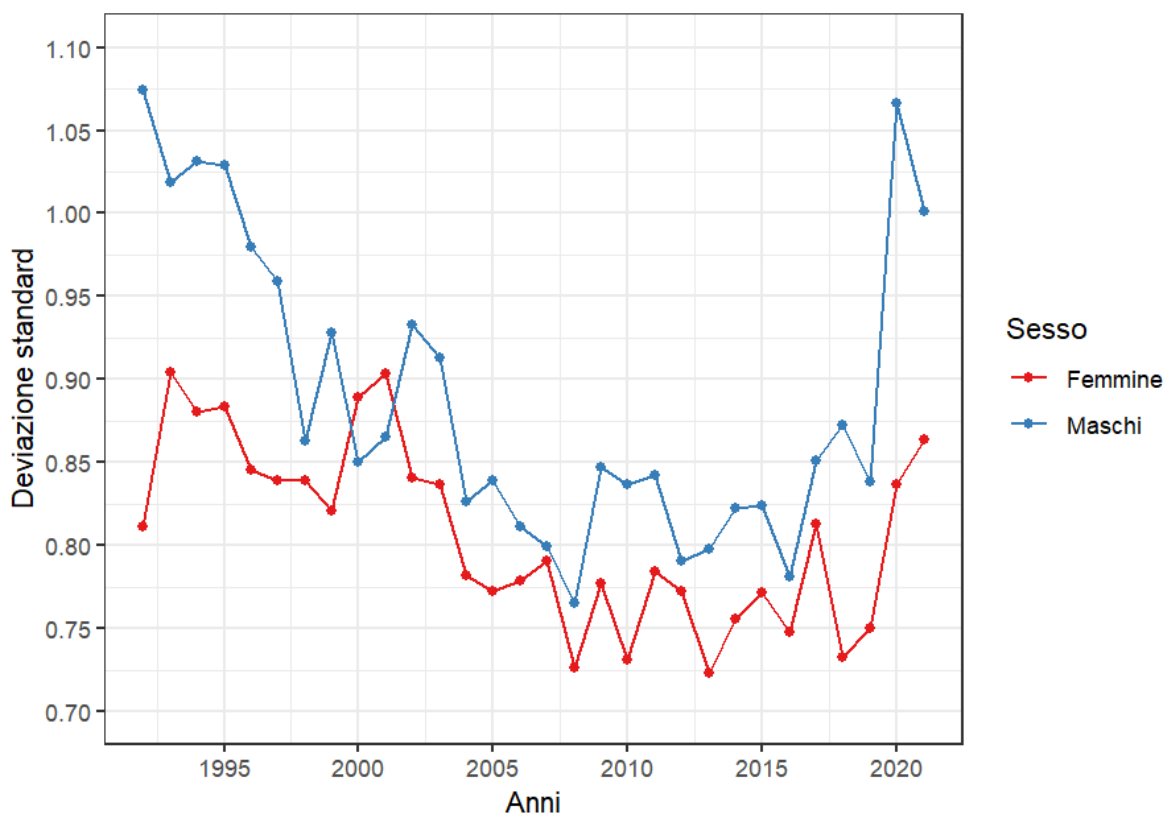
**Figura 1.1:** *Speranza di vita alla nascita più alta e più bassa a livello provinciale, maschi e femmine, 1992-2011*

Per integrare e comprendere più approfonditamente come è cambiata negli anni la variabilità di questo fenomeno, si mostra in Figura 1.2 la serie delle deviazioni standard della speranza di vita alla nascita tra le varie province e per i due sessi come indice riassuntivo.

Dal grafico emerge in primo piano che la deviazione standard dei maschi è maggiore rispetto a quella delle femmine in tutti gli anni eccetto nel 2000 e 2001. Si nota inoltre, che lo scarto quadratico medio si è evoluto in maniera differente negli anni '90, infatti nei maschi è avvenuta una netta diminuzione, causata, osservando attentamente la Figura 1.1, dalla crescita più veloce della *worst-practice*, mentre per le femmine risulta stazionario. Dopodiché per entrambe le serie i valori calano negli anni 2000 fino a fine decennio, riduzione determinata da un rallentamento nella crescita di entrambe le *best-practice*, per poi stabilizzarsi fino al brusco picco del 2020, anno in cui la diffusione disomogenea del Covid-19 ha causato un aumento significativo della variabilità della speranza di vita tra le province, particolarmente spiccato nei maschi, con una variazione di 0.25 rispetto all'anno precedente, raggiungendo così i livelli degli anni '90.

Infine, nel 2021 i segni delle variazioni nei due sessi sono opposti, negativo per i maschi, e positivo per le femmine, le quali ritornano al grado di variabilità esistente prima della diminuzione avvenuta negli anni 2000.

Combinando i grafici in Figura 1.1 e Figura 1.2, si può comprendere che il marcato aumento della variazione nel 2020 rispetto al 2019 è provocato dalla maggiore riduzione della speranza di vita alla nascita della *worst-practice* rispetto alla *best-practice*.



**Figura 1.2:** deviazione standard dell'aspettativa di vita alla nascita tra le province, maschi e femmine 1992-2021

Si può dunque concludere che il Covid-19 ha provocato delle disparità nella mortalità, oltre che per diverse caratteristiche come il sesso, anche tra le province.

Per comprendere meglio la natura di queste differenze, è più informativo esaminare i modelli di mortalità attraverso le età. Nel prossimo capitolo si introduce dunque una tecnica di scomposizione matematica utile comprendere le principali componenti che determinano la disuguaglianza nella mortalità tra le province.

# Capitolo 2

## Scomposizione della disuguaglianza di mortalità

In questo capitolo si presenta una tecnica di scomposizione utile per individuare le dimensioni chiave della disuguaglianza di mortalità che esiste all'interno di una popolazione.

La base di questo modello è la Decomposizione a Valori Singolari (SVD), una fattorizzazione matriciale che permette di rappresentare una qualunque matrice  $N \times P$  come il prodotto di tre matrici, come segue:

$$\mathbf{X} = \mathbf{U} \mathbf{D} \mathbf{V}^T \quad (2.1)$$

dove:

- $\mathbf{U}$  è una matrice ortonormale  $N \times N$  e contiene nelle colonne i cosiddetti vettori singolari sinistri di  $\mathbf{X}$ , ovvero gli autovettori di  $\mathbf{X}\mathbf{X}^T$
- $\mathbf{V}$  è una matrice ortonormale  $P \times P$  e contiene nelle colonne i cosiddetti vettori singolari destri di  $\mathbf{X}$ , ovvero gli autovettori di  $\mathbf{X}^T\mathbf{X}$
- $\mathbf{D}$  è una matrice rettangolare diagonale i cui elementi sono i cosiddetti valori singolari di  $\mathbf{X}$ , ovvero la radice quadrata degli autovalori di  $\mathbf{X}\mathbf{X}^T$

La Decomposizione a valori singolari, grazie alle sue proprietà, trova numerose applicazioni in diversi settori, compreso quello demografico, consentendo di analizzare fenomeni quali mortalità, fertilità e migrazione.

Risulta uno strumento utile per ridurre la dimensionalità di un fenomeno in un numero ristretto di componenti grazie all'analisi delle componenti principali (PCA): queste sono combinazioni lineari delle variabili di partenza di una matrice (nel caso in esame si tratta della matrice dei tassi di mortalità) tali per cui siano ad una ad una incorrelate e la varianza risulti massima nelle componenti. In relazione all'Equazione 2.1, le componenti principali coincidono con gli autovettori di  $\mathbf{XX}^T$ , ovvero i vettori singolari sinistri della matrice  $\mathbf{X}$  (il numero di componenti è quindi  $N$ ). Date le proprietà della PCA appena esposte, è possibile rappresentare un fenomeno complesso attraverso le prime componenti limitando la perdita di informazione, in quanto racchiudono una buona quota di varianza. Negli studi concernenti la popolazione ciò è possibile grazie alla presenza di forti regolarità nei fenomeni demografici. Specificamente alla mortalità, si tratta di estrapolare le principali fonti di variazione all'interno di una popolazione o attraverso gli anni.

Per ulteriori approfondimenti riguardo la SVD e l'analisi delle componenti principali si veda Johnson et al. (2014).

Nell'ambito demografico ci sono numerose applicazioni che derivano dalla SVD: la più popolare è il modello di Lee-Carter (1992), il quale permette di modellare i tassi di mortalità specifici per età utilizzando una sola componente principale, con l'obiettivo di ottenere stime e fare previsioni; per questo motivo tiene in considerazione i dati ordinati nel tempo. La formula è la seguente:

$$\log (m_{x,t}) = a_x + b_x k_t \quad (2.2)$$

dove:

- $a_x$  si stima con la media del logaritmo dei tassi di mortalità specifici per una certa età  $x$
- $b_x$  rappresenta gli scarti dalla media del logaritmo dei tassi di mortalità specifici per una data età  $x$
- $k_t$  esprime la grandezza di questi scarti per un determinato anno  $t$

Si noti che il numero di età  $x$  e il numero di anni  $t$  è diverso, quindi si ha a che fare con una matrice non quadrata. La decomposizione a valori singolari permette di scomporla in due matrici ortonormali ed una diagonale.

Dalla formulazione del modello di Lee-Carter (1992) sono susseguite numerose varianti ed estensioni, perlopiù modelli con fini predittivi. La quasi totalità di queste applicazioni tratta le componenti età e tempo, mentre lo spazio solitamente non rappresenta il fulcro principale delle analisi; infatti, pochi studi sfruttano la SVD per ottenere misure di sintesi della mortalità all'interno di una popolazione.

Questo elaborato si sofferma proprio su questa variabile, trasferendo dunque il focus primario sullo spazio, ma sfruttando comunque il tempo come veicolo per esaminare l'evoluzione del fenomeno.

L'idea trae ispirazione dalla visione dell'articolo 'Decomposing dimensions of mortality inequality' di Alexander (2022), in cui viene discusso l'utilizzo di un'applicazione della Decomposizione a Valori Singolari per ricavare le dimensioni chiave della disuguaglianza di mortalità per età tra gli stati degli Stati Uniti per l'anno 2019, separatamente per maschi e femmine.

Quest'applicazione costituirà il corpo centrale di questa relazione e verrà affrontata nel prossimo capitolo.

Si procede ora con la presentazione dei dati reperiti per la ricerca e la spiegazione della procedura analitica del modello.

I dati presi in esame riguardano le province italiane, che ammontano a 107. Questi sono ricavati dalle tavole di mortalità complete a livello provinciale disponibili dal 1992 al 2021 presenti in Demo.Istat, il database dell'istituto nazionale di statistica in cui vengono diffuse e rese disponibili gratuitamente statistiche, indicatori e rilevazioni demografiche italiane.

Per l'obiettivo dell'analisi sono stati utilizzati, distintamente per maschi e femmine, i tassi di mortalità specifici per età, indicati con  $m_x$  e l'aspettativa di vita alla nascita, indicata con  $e_0$ , dove  $x$  si riferisce all'età. I tassi di mortalità specifici per età si stimano a partire dai decessi reali all'età  $x$  ( $D_x$ ) e la popolazione reale media nell'età  $x$  ( $P_x$ ) come segue:

$$m_x = D_x / P_x \quad (2.3)$$

È possibile, inoltre, ottenerli con semplicità sfruttando le funzioni della tavola di mortalità, stime ricavate proprio dai tassi di mortalità. In particolare, dati i decessi all'età  $x$  ( $d_x$ ) e gli anni vissuti all'età  $x$  ( $L_x$ ):

$$m_x = d_x / L_x \quad (2.4)$$

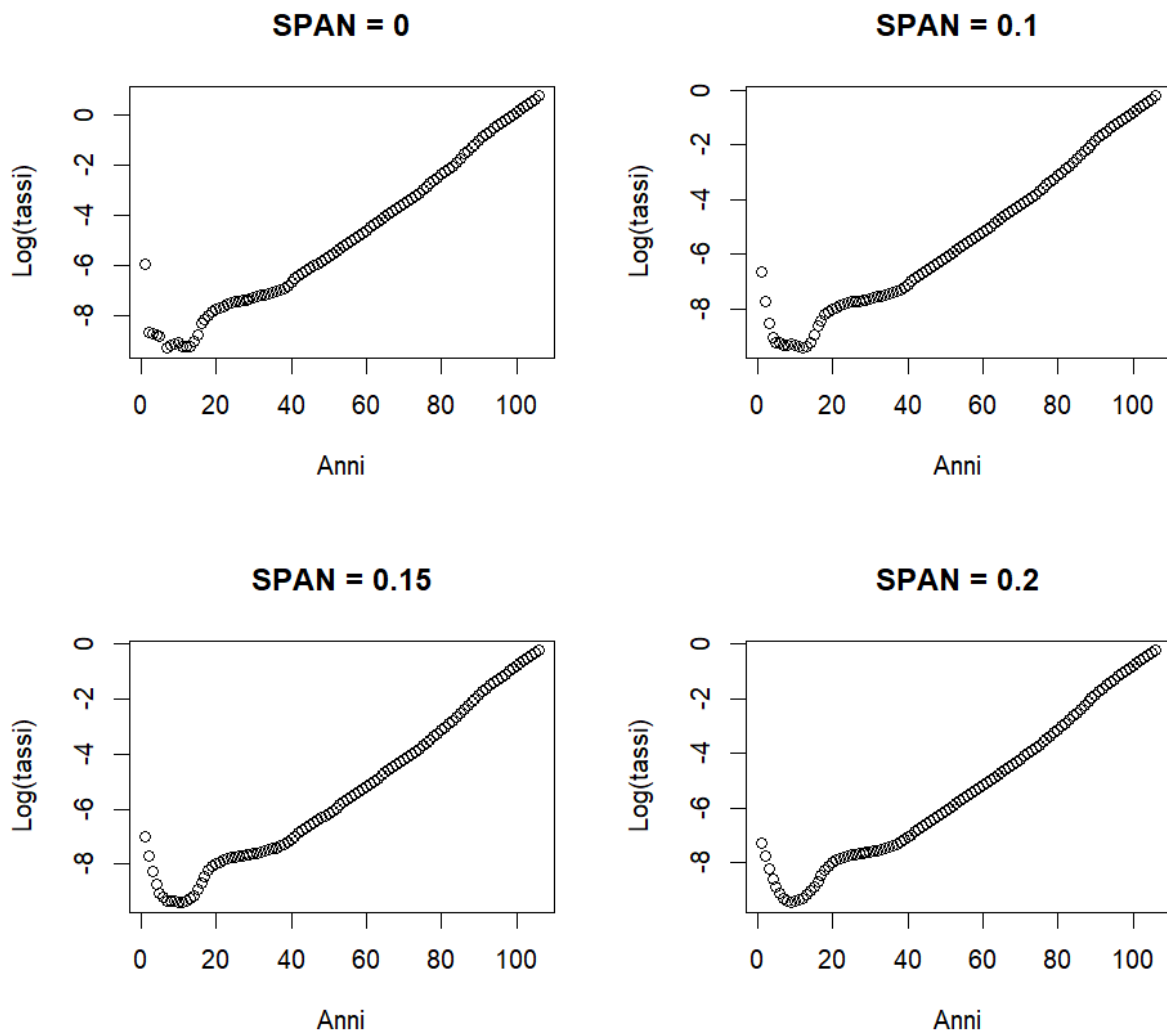
La speranza di vita alla nascita si ricava anch'essa come funzione della tavola di mortalità ed è espressa come:

$$e_0 = T_0 / l_0 \quad (2.5)$$

dove  $T_0$  rappresenta la somma degli anni vissuti per ogni età ( $L_x$ ) e  $l_0$  rappresenta i sopravvissuti all'età 0. Nelle tavole di mortalità sono disponibili i dati fino all'età 119; per convenienza e semplicità si andrà a considerare per i più anziani la fascia d'età aperta 105+.

Prima di proseguire con la Decomposizione a Valori Singolari, seguendo quanto svolto da Alexander, ai tassi di mortalità viene applicata la trasformazione logaritmica come tipicamente accade, successivamente viene applicata loro una funzione di lisciamento per garantire che le componenti estratte siano altrettanto 'lisce', ed infine ad ogni osservazione viene sottratta la rispettiva media specifica per età, con il risultato di trattare gli scarti dalla media per ogni provincia ed ogni singola età.

Per quanto riguarda la funzione di lisciamento, il metodo impiegato è la regressione locale, un modello non parametrico che applica la regressione a sottoinsiemi dei dati corrispondenti ad intervalli regolari dell'asse delle ascisse. Questa funzione attribuisce peso maggiore ai punti più vicini al centro dell'intervallo. Il modello è caratterizzato dallo *span*, ovvero il parametro di livellamento, un numero reale positivo che specifica la dimensione dell'intorno di dati. Alexander opta per uno span, ovvero il parametro di livellamento, pari a 0.2, mentre con questi dati si è deciso di applicare uno span di 0.1. La spiegazione di questa scelta è visibile in Figura 2.1, dove vengono rappresentati, come esempio, le medie dei tassi di mortalità per età provinciali dei maschi nel 2019, ed emerge che uno span pari a 0.2 rende il calo della mortalità infantile molto più graduale rispetto a quanto avviene nella realtà. Si noti che risultano grafici simili per gli anni 2020 e 2021 e anche per il sesso femminile.



**Figura 2.1:** *Logaritmo dei tassi di mortalità con diversi valori di lisciamento, maschi, 2019*

Ricapitolando, per un determinato anno risulta una matrice  $N \times G$  contenente il logaritmo dei tassi di mortalità specifici per età delle province italiane, lisciati e a cui è stata sottratta la media per ogni età.  $N$  rappresenta il numero di province, quindi  $N = 107$ , mentre  $G$  rappresenta il numero di fasce d'età, dunque  $G = 106$ , avendo a che fare con singole età e la fascia aperta 105+.

Applicando la Decomposizione a Valori Singolari a questa matrice, è possibile comprendere quali sono le principali differenze nella mortalità tra le province nelle diverse età. La SVD permette dunque di rappresentare il logaritmo del tasso di mortalità specifico per età in un determinato anno  $x$  e una determinata provincia  $p$  come:

$$\log (m_{x,p}) = a_x + b_{x,1} k_{p,1} + b_{x,2} k_{p,2} + \dots + b_{x,p} k_{p,p} \quad (2.6)$$

dove:

- $a_x$  viene stimato con la media dei tassi di mortalità specifici di tutte le province per l'età  $x$
- $b_{x,i}$  rappresenta lo scarto dalla media per l'età  $x$  espresso dall' $i$ -esima componente
- $k_{p,i}$  rappresenta il peso della provincia  $p$  in riferimento all' $i$ -esima componente

$a_x$  viene stimato con la media così da applicare la SVD ad una matrice a media nulla; in questo modo avremo una serie di componenti  $b_{x,i}$ , ovvero i vettori singolari destri ( $\mathbf{V}$ ), che forniscono informazioni riguardo alle principali differenze tra le province, e una serie di pesi  $k_{p,i}$ , cioè i vettori singolari sinistri ( $\mathbf{U}$ ) moltiplicati per i valori singolari di  $\mathbf{D}$ , che definiscono il contributo di ogni provincia per ogni componente.

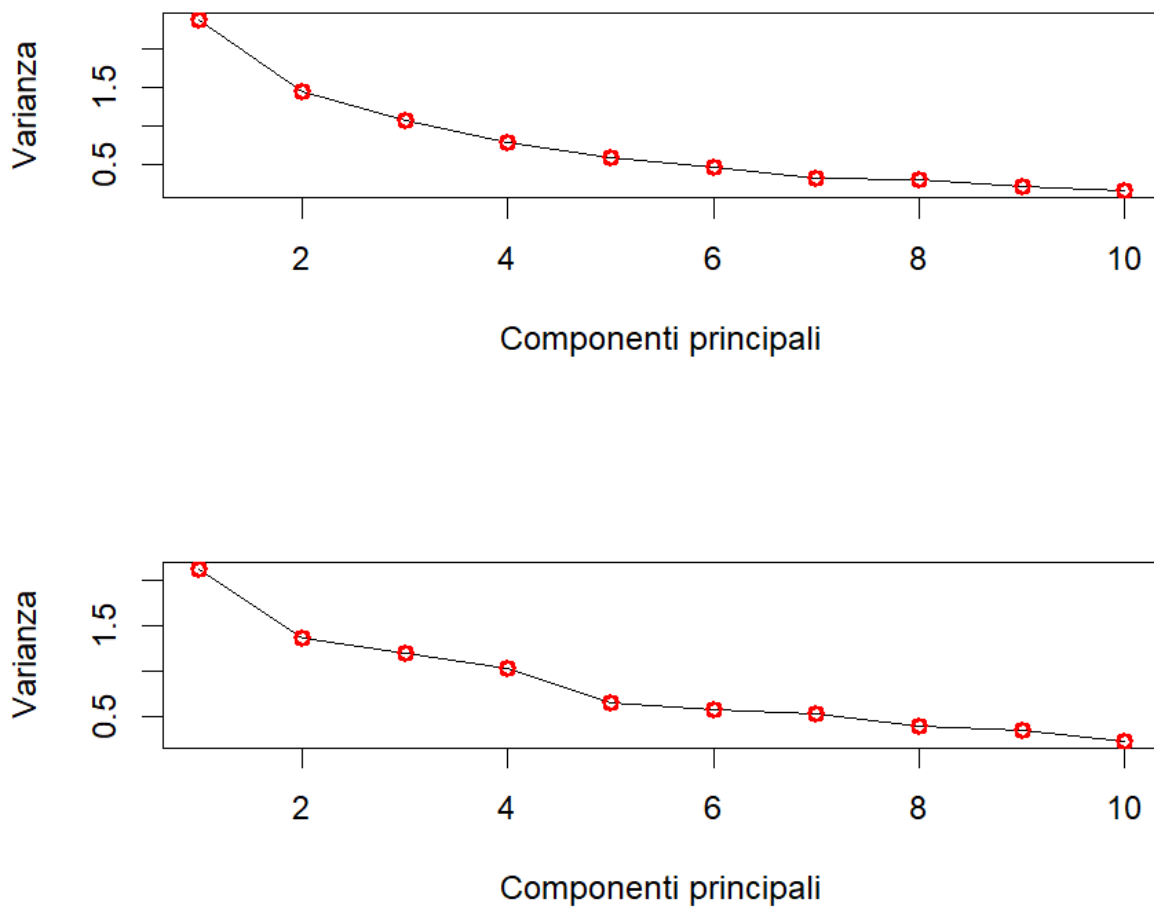
L'obiettivo dell'analisi è ottenere misure di sintesi riguardo la disuguaglianza di mortalità tra le province attraverso poche dimensioni chiave confrontandole tra l'anno immediatamente precedente alla pandemia e gli anni colpiti dalla stessa; per questo motivo il numero di vettori singolari, che costituiranno le componenti principali, da tenere in esame è stato selezionato secondo i criteri dell'analisi delle componenti principali.



Alexander nella sua relazione opta l'utilizzo di due componenti per spiegare il fenomeno. È importante considerare che per molteplici motivi, gli stati degli Stati Uniti presentano una maggiore variabilità tra loro rispetto alla variabilità che esiste tra le province italiane. In termini di varianza totale spiegata, alla luce dei dati considerati, due componenti spiegano in media circa il 40-45%, il che rappresenta un limite, in virtù del fatto che sarebbe auspicabile spiegare almeno approssimativamente il 70% della varianza; per ottenere un tale risultato, però, bisognerebbe tenere in media cinque componenti, complicando notevolmente l'interpretazione delle componenti, soprattutto in combinazione tra loro.

Attraverso gli scree-plot, un'esplorazione grafica della varianza delle componenti principali nei vari anni e per sesso, in alcuni casi sono emersi dei 'gomiti' (non particolarmente spiccati) nei pressi della seconda componente, mentre in altri casi non risulta alcuna brusca variazione di pendenza. In Figura 2.2, come esempio, si illustrano gli scree-plot delle componenti principali nei maschi per gli anni 2019 e 2020.

Seguendo il criterio di parsimonia e tenendo a mente l'obiettivo dell'analisi, ovvero individuare le principali fonti di variazione, si è dunque concluso di tenere le prime due componenti per favorire un'interpretazione più diretta ed intuitiva a discapito di una minore percentuale di varianza spiegata.



**Figura 2.2:** scree-plot delle componenti principali, maschi, rispettivamente 2019 e 2020

In conclusione, il logaritmo dei tassi di mortalità per un'età  $x$  e una provincia  $p$  sarà dunque espressa in tal modo:

$$\log(m_{x,p}) = a_x + b_{x,1} k_{p,1} + b_{x,2} k_{p,2} \quad (2.7)$$

dove  $b_{x,1}$  e  $b_{x,2}$  sono i primi due vettori singolari destri, mentre  $k_{p,1}$  e  $k_{p,2}$  sono il prodotto dei primi due vettori singolari sinistri moltiplicati rispettivamente per il primo ed il secondo autovalore della matrice diagonale  $\mathbf{D}$ .

Nel prossimo capitolo si presenteranno i risultati ottenuti da questo modello nell'analisi delle province italiane, separatamente per maschi e femmine, negli anni 2019, 2020 e 2021, analizzando le prime due componenti principali e fornendo qualche esempio specifico tramite i loadings associati alle componenti.



# Capitolo 3

## Risultati dell'analisi per le province italiane

In questo capitolo vengono illustrati i risultati ottenuti elaborando i tassi di mortalità, distintamente per sesso, nei tre anni presi in esame: 2019, 2020 e 2021. Tutti i calcoli sono stati svolti con il software R e i grafici con la funzione `ggplot`, contenuta nel pacchetto `ggplot2` (Wickham 2009).

### 3.1 Disuguaglianze nel sesso maschile

Questa sezione presenta le componenti chiave che descrivono la disuguaglianza di mortalità, la loro evoluzione negli anni e la loro relazione con le province, per il sesso maschile.

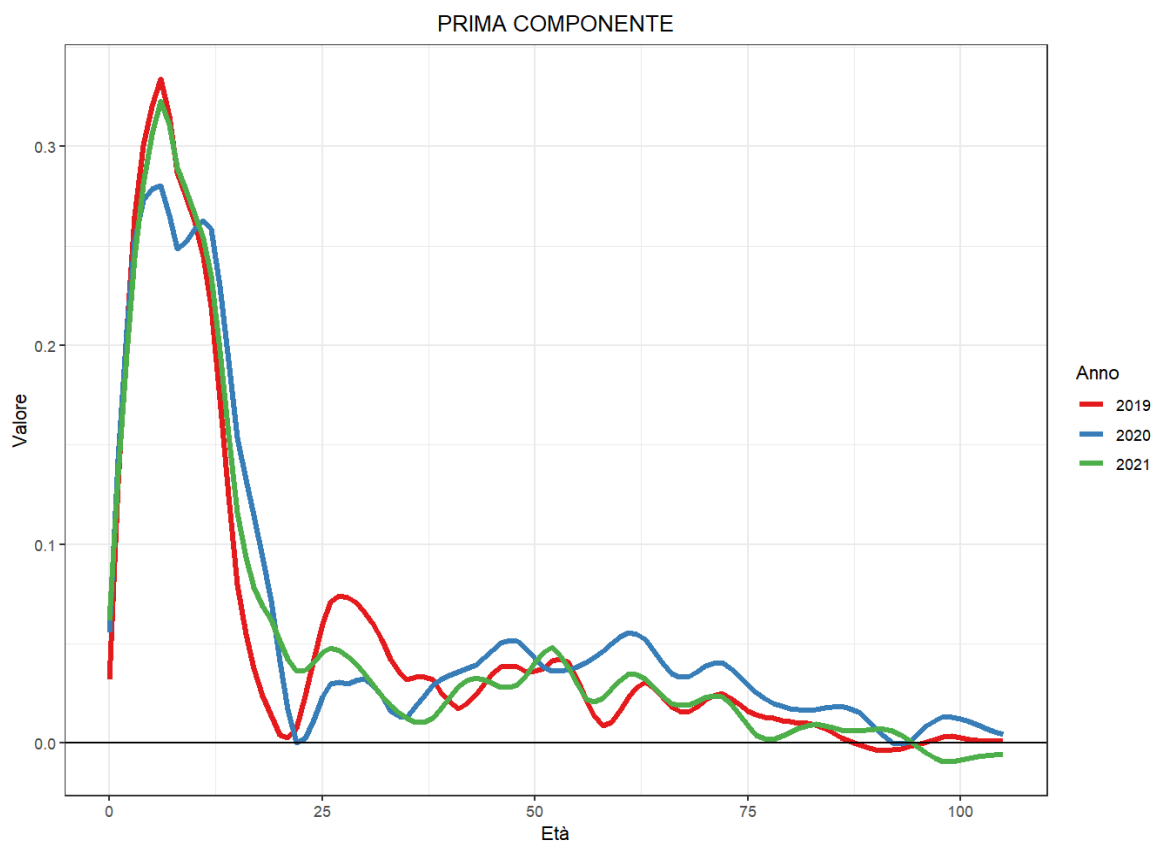
In figura 3.1.1 è rappresentata la prima componente principale per i tre anni considerati. Nella formula del modello (Equazione 2.7) corrisponde al primo vettore singolare destro  $b_{x,1}$  e qui indica la principale fonte di variazione nella mortalità tra le province. Essa fornisce informazioni riguardo la disuguaglianza di mortalità nel suo complesso. Per tutti e tre gli anni si evidenziano le differenze maggiori nei bambini, esclusi i neonati, e nei giovani fino alla maggiore età, con l'apice della diversità intorno ai 5-6 anni d'età; dai 20 anni in poi avviene una decrescita repentina e le differenze diminuiscono gradualmente all'aumentare dell'età.

L'avvento del Covid-19, nel 2020, ha causato dei cambiamenti in questi termini; infatti, rispetto all'anno precedente si nota una riduzione della variabilità nei bambini di età compresa tra i 5 e i 12 anni e nei giovani adulti tra i 20 e i 30 anni. Al contrario aumenta la disparità nelle fasce d'età più adulte, anziani compresi.

Quanto detto è coerente con le note conseguenze della pandemia, ossia l'aver colpito principalmente le persone oltre i 50 anni e, data l'eterogeneità con cui il virus si è diffuso nelle varie province, pare comprensibile l'effetto di generare disparità nella mortalità in questa classe d'età.

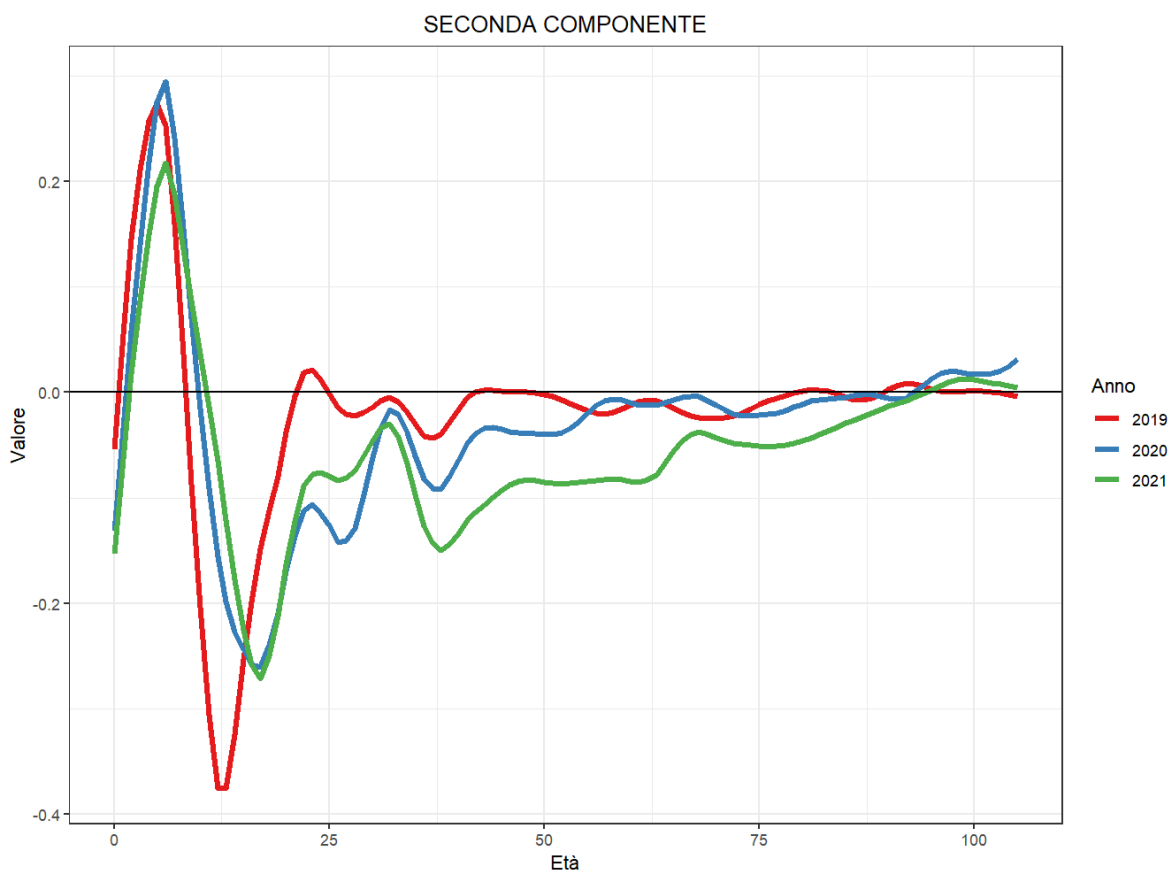
Nel 2021, anno ancora delicato in termini di propagazione del virus, le differenze tornano in realtà ad assumere un comportamento simile a quanto visto per il 2019, ad eccezione della fascia d'età 15-34 anni. È interessante osservare che negli anni è avvenuto un incremento, seppur contenuto, della diversità nella mortalità infantile tra province.

Riassumendo, per tutti e tre gli anni le differenze più lampanti si verificano nei bambini e negli adolescenti fino alla maggiore età, con una progressiva attenuazione al crescere dell'età. L'arrivo della pandemia ha provocato un modesto abbassamento della disparità tra i giovani a discapito di un aumento nelle fasce adulte, mentre già l'anno successivo si ritrovano dei pattern presenti nell'anno precedente alla circolazione del virus.



**Figura 3.1.1:** *prima componente principale, maschi, 2019-2021*

In Figura 3.1.2 è raffigurato il secondo vettore singolare destro, ovvero  $b_{x,2}$  nell'Equazione 2.7 e rappresenta la seconda principale fonte di disuguaglianza nella mortalità. Si denota che, rispetto alla prima componente, aggiunge informazioni riguardo alle differenze nella forma delle curve di mortalità, in particolare mette in contrapposizione la mortalità tra i bambini e la mortalità tra gli adolescenti. Questa affermazione vale soprattutto per il 2019, mentre negli anni successivi queste variazioni tendono ad attenuarsi per le suddette classi d'età. Inoltre, gli anni 2020 e 2021 mostrano un aumento della diversità tra le province nei giovani adulti e negli adulti, oltre che nella mortalità infantile, come già spiegato dalla prima componente, il tutto in contrapposizione con la mortalità tra i bambini.



**Figura 3.1.2:** seconda componente principale, maschi, 2019-2021

È possibile avere un'interpretazione più chiara delle curve di mortalità specifiche delle province combinando i primi due vettori singolari destri con i pesi, chiamati anche loadings, riferiti alle prime due componenti per ogni provincia, ovvero i coefficienti lineari applicati alle variabili originarie (tassi di mortalità) che definiscono le componenti principali. Nell'Equazione 2.7 corrispondono a  $k_{p,1}$  e  $k_{p,2}$ . Essi descrivono il contributo di ogni provincia alla costruzione della prima e seconda componente principale rispettivamente.

Valori alti dei loadings (in valore assoluto) esprimono una forte relazione tra la provincia in questione e la componente principale a cui si riferisce il peso, mentre il segno dei loadings determina se la correlazione con la componente principale è positiva o negativa. In Figura 3.1.3 vengono presentati i loadings per le prime due componenti principali dell'anno 2019; l'ordine delle province è tale da avere i loadings riferiti alla prima componente ordinati in senso decrescente.



Ad esempio, la Valle d'Aosta è correlata positivamente con la prima componente principale, per cui la curva di mortalità sarà ben descritta da questa componente, deducendo che avrà dunque un livello di mortalità giovanile più alto della media. La provincia di Rimini, invece, è correlata positivamente con la seconda componente, per cui si può affermare che la curva di mortalità è tale da avere una mortalità tra i bambini relativamente alta mentre un livello di mortalità tra gli adolescenti relativamente basso. Per approfondimenti dettagliati, con la possibilità di effettuare confronti per ogni provincia, in Figura A.1 e in Figura A.2 (vedi Appendice) vengono mostrati i loadings per il 2020 e 2021. Per eseguire confronti temporali nei diversi anni, è bene sottolineare che il segno ed il valore dei loadings si riferiscono esclusivamente alle componenti dello stesso anno, per cui è fondamentale soffermarsi sull'interpretazione di questi due elementi.

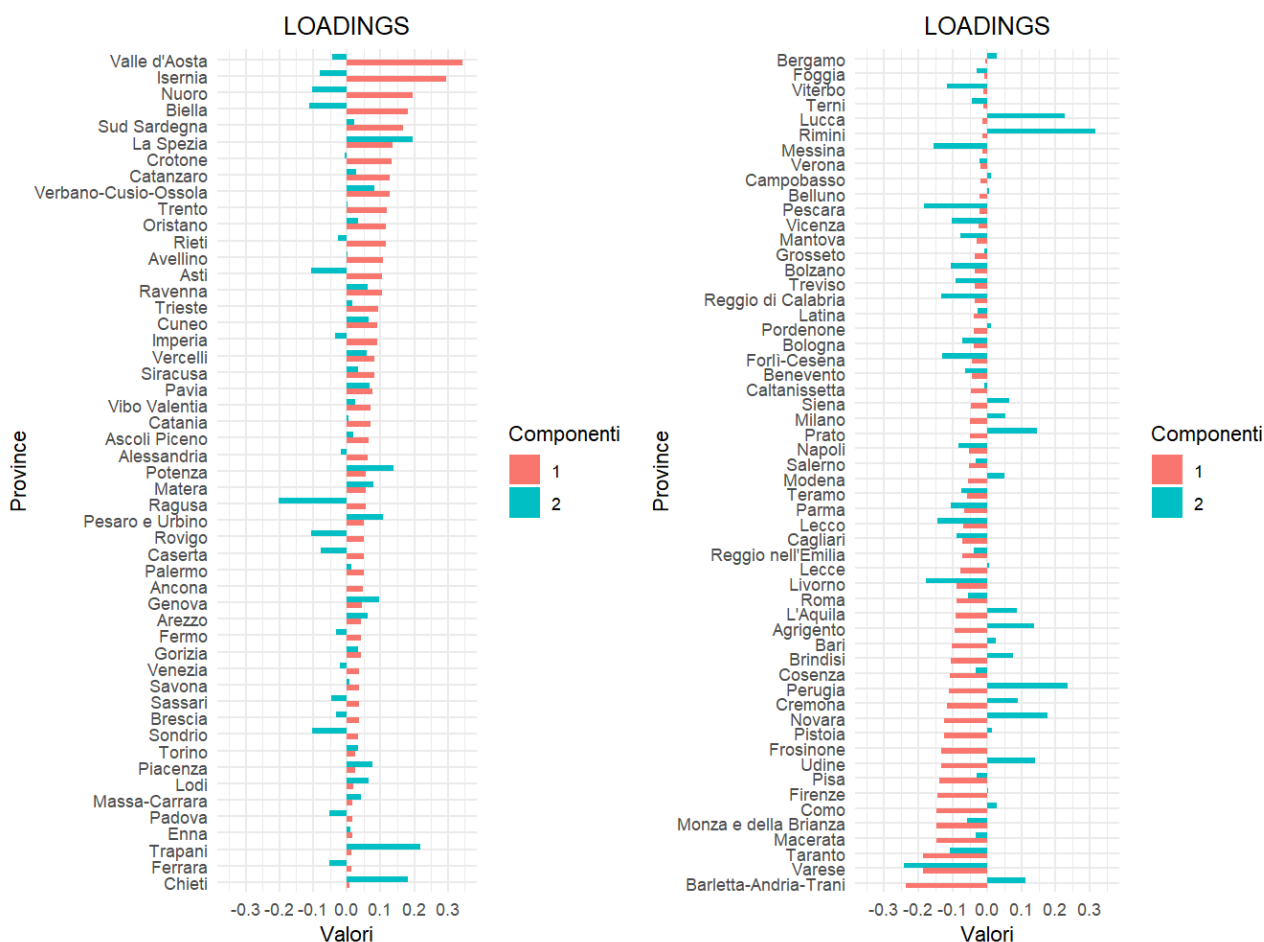


Figura 3.1.3: loadings riferiti alla prima e seconda componente principale, maschi, 2019

In Tabella 3.1.1 sono raccolti i loadings di alcune province che hanno subito cambiamenti negli anni dal punto di vista del livello di mortalità e della forma della curva di mortalità.

La provincia di Ancona ha conseguito dei miglioramenti notevoli in termini di mortalità, infatti se al 2019 il loading della prima componente suggerisce un livello di mortalità complessivo superiore alla media, il segno del peso nel 2021 indica una correlazione negativa con la prima componente e di conseguenza una mortalità generale sotto la media. Si segnala inoltre una correlazione positiva del loading riferito alla seconda componente per il 2020, evidenziando un grado di mortalità sotto la media per la fascia d'età adolescenziale e sopra la media per i bambini, seppur di poco considerando l'interpretazione del loading della prima componente.

Per quanto riguarda la provincia di Cremona, diminuisce la correlazione con la prima componente e aumenta, con segno negativo, la correlazione con la seconda componente: ne risulta una curva con livelli di mortalità relativamente bassi tra i bambini e relativamente alti tra gli adolescenti.

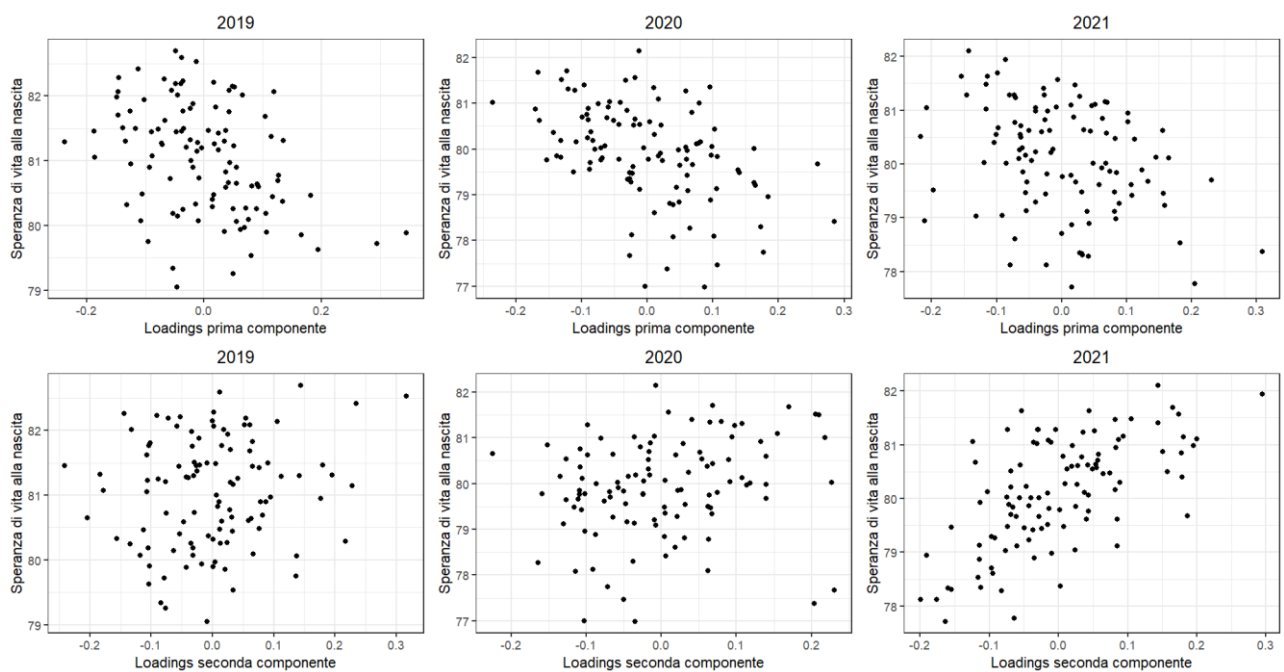
Infine, per la provincia di Enna si nota un aumento del livello di mortalità complessivo. La correlazione negativa con la seconda componente, combinata con quanto appena descritto, suggerisce alta mortalità tra gli adolescenti.

	Prima componente			Seconda componente		
	2019	2020	2021	2019	2020	2021
<b>Ancona</b>	0.049	-0.12	-0.208	0.0003	0.107	-0.011
<b>Cremona</b>	-0.117	-0.03	-0.098	0.09	-0.103	-0.12
<b>Enna</b>	0.017	0.106	0.206	0.012	-0.036	-0.064

**Tabella 3.1.1:** *loadings delle prime due componenti per alcune province, maschi, 2019-2021*

Per concludere la sezione maschile, si analizza la relazione tra i loadings delle due componenti principali e la speranza di vita alla nascita attraverso i grafici di dispersione nei tre anni, in Figura 3.1.4.

È possibile notare che l'aspettativa di vita alla nascita è correlata negativamente con i loadings riferiti alla prima componente principale, mentre è correlata positivamente con i loadings della seconda componente. Ciò suggerisce che le province con una minore speranza di vita tendono ad avere un livello di mortalità complessivo maggiore ed è ciò che intuitivamente ci si può aspettare, dal momento che l'aspettativa di vita alla nascita è un indicatore che riflette il grado di mortalità generale. Riguardo alla relazione con la seconda componente, combinata con la conclusione precedente, risulta che le province con la speranza di vita più alta si caratterizzano con un livello di mortalità particolarmente basso nella fascia d'età adolescenziale più che per le altre classi d'età.

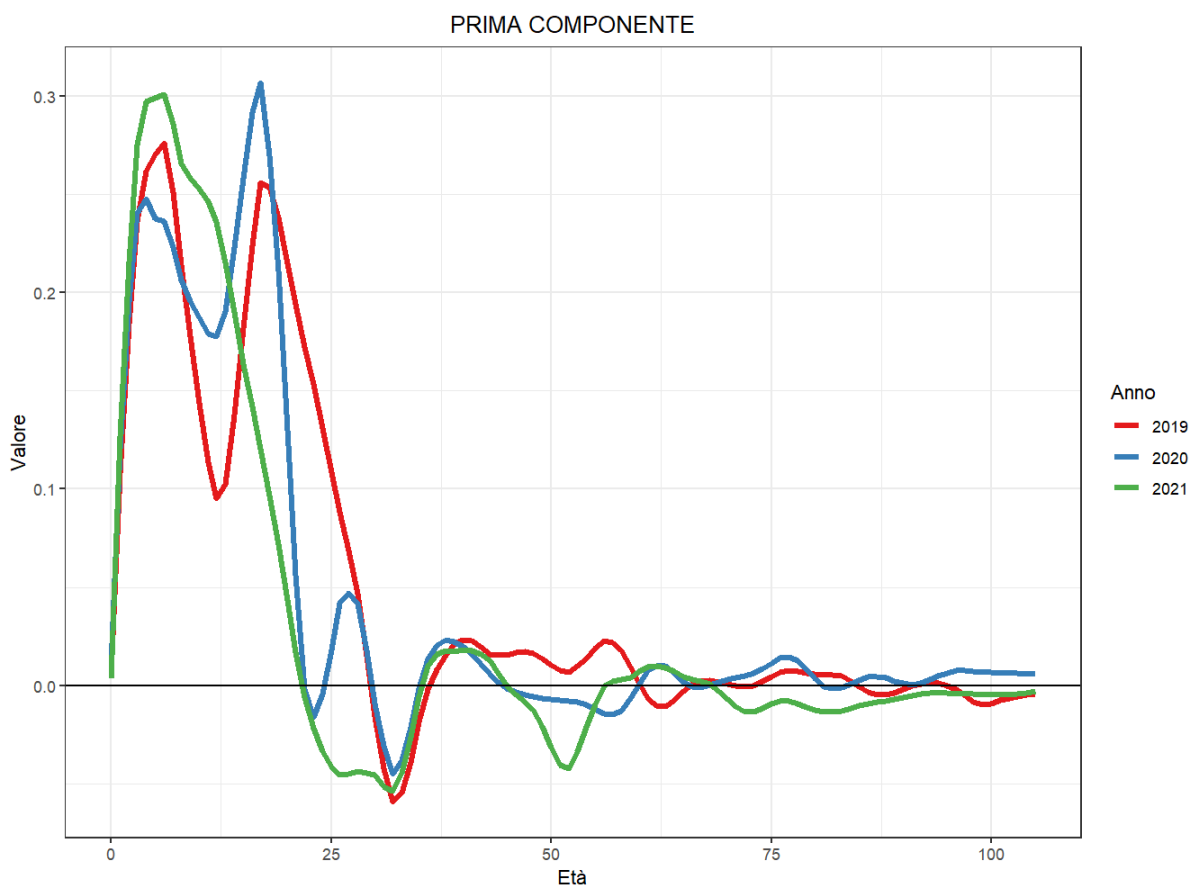


**Figura 3.1.4:** relazione tra la speranza di vita alla nascita e i loadings riferiti alle prime due componenti principali, maschi, 2019-2021

## 3.2 Disuguaglianze nel sesso femminile

Questa sezione presenta le componenti principali che definiscono le differenze di mortalità nelle varie età, il loro cambiamento negli anni e la loro relazione con le province, per il sesso femminile.

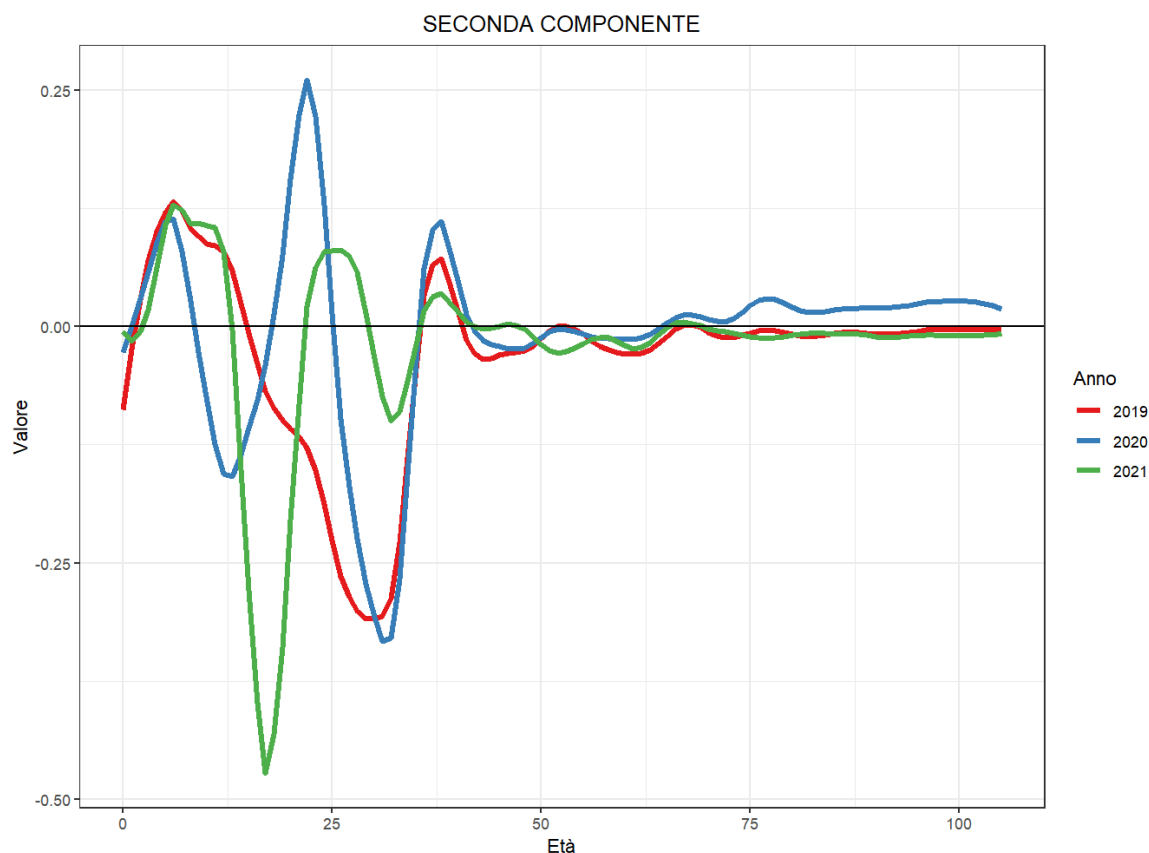
In figura 3.2.1 è rappresentata la prima componente principale per i tre anni considerati. Come per i maschi, anche in questo caso le differenze più rilevanti si evidenziano nelle bambine, escluse le neonate, e nelle giovani ragazze fino alla maggiore età, con un graduale annullamento della disuguaglianza all'aumentare dell'età, per tutti e tre gli anni. Appare chiaro però, che tra i 5 ed i 30 anni la principale fonte di variazione subisce dei cambiamenti negli anni molto più di quanto osservato nei maschi, in particolar modo dal 2020 al 2021. Sinteticamente, nel 2020 aumenta la disuguaglianza tra le adolescenti fino ai 18 anni rispetto all'anno precedente, mentre avviene un brusco annullamento intorno ai 25 anni; nel 2019 avviene qualche anno più tardi. Nel 2021 si osserva un aumento della diversità di mortalità per le bambine fino ai 12 anni, una successiva brusca decrescita intorno ai 20 anni ed un aumento intorno ai 50 anni. È una componente un poco più difficile da interpretare rispetto a quella dei maschi in termini di evoluzione nel tempo, ma rimane comunque il fatto che anche per il sesso femminile la principale disuguaglianza è presente nelle fasce d'età giovanili (bambine e adolescenti). Non compaiono evidenti pattern nel 2020 riconducibili ad un effetto della pandemia; infatti, si registra un aumento della variazione intorno ai 16 anni ed una diminuzione nella classe d'età successiva (18-24 anni). Infine, è importante sottolineare che tra il 2019 e il 2021, per alcune classi d'età i versi delle variazioni sono opposti, in particolare nella fascia d'età 20-30 anni per l'anno 2021 rispetto ai due precedenti.



**Figura 3.2.1:** prima componente principale, femmine, 2019-2021

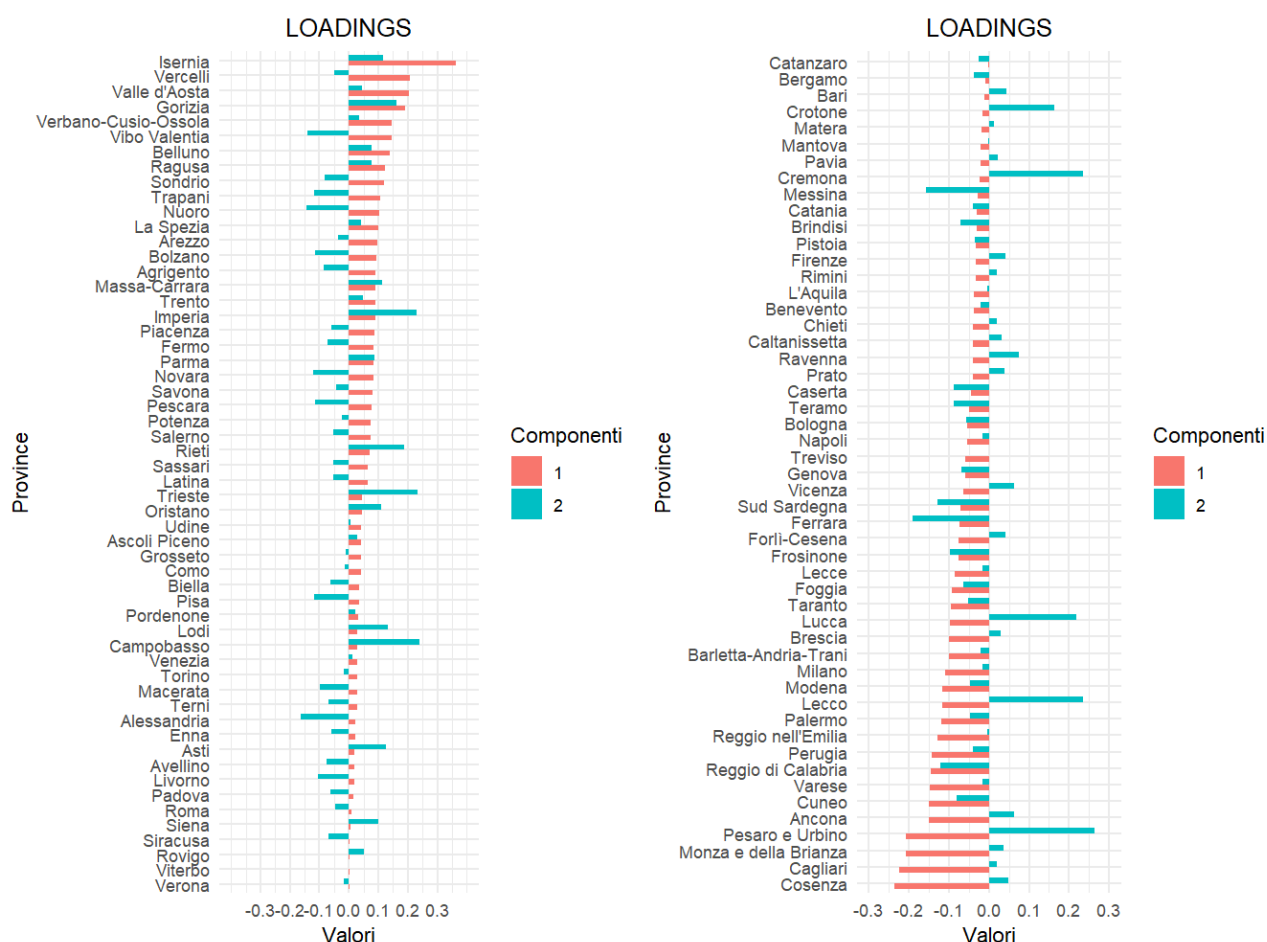
In Figura 3.2.2 viene illustrata la seconda componente principale dal 2019 al 2021. Un forte impatto visivo è dato dalla notevole differenza nell'andamento della disuguaglianza di mortalità, tra il 2019 e il 2021, nelle classi d'età comprendenti le adolescenti e le giovani adulte fino ai 30 anni. Questa componente, come per i maschi, fornisce informazioni sulla forma delle curve di mortalità, sebbene qui sia più complicata e meno intuitiva l'interpretazione, specialmente se combinata con la prima. Per il 2019 la componente esprime il contrasto generale tra il livello di mortalità nella classe d'età 1-14 anni e quello nella fascia d'età 15-34 anni. Per il 2020 e per il 2021 le disuguaglianze sono più disparate anche all'interno di una stessa classe d'età, con cambi di direzione repentini: nel 2020 si mettono in contrapposizione la mortalità nelle età 0-8, 18-25 e 35-40 con la mortalità nelle età 9-17 e 26-34 anni, mentre nel 2021 si contrappongono la mortalità nell'età 3-12 e 20-30 con la mortalità nelle età 13-20 (da sottolineare l'alto valore, soprattutto in relazione ai due anni precedenti) e 30-35 anni.

Inoltre, in questa componente si nota un possibile effetto del Covid-19, ovvero l'aumento delle differenze di mortalità nella fascia anziana nel 2020 contro una diminuzione nei neonati.



**Figura 3.2.2:** seconda componente principale, femmine, 2019-2021

La Figura 3.2.3 raffigura i loadings delle province, riferiti alle prime due componenti principali, per il 2019. Si sceglie come esempio esplicativo dell'interpretazione dei loadings la provincia di Pesaro e Urbino: è correlata negativamente con la prima componente e positivamente con la seconda. Si deduce una quota di mortalità generale relativamente bassa (prima componente), in particolar modo nella classe d'età 15-34 anni (seconda componente). Anche qui, per confrontare ogni provincia per gli anni successivi, in Figura A.3 e in Figura A.4 (vedi Appendice) sono presenti i loadings per il 2020 e il 2021.



**Figura 3.2.3:** loadings riferiti alla prima e seconda componente principale, femmine, 2019

Analogamente a quanto svolto nel sesso maschile, si presentano in Tabella 3.1.2 alcune province che hanno sperimentato modifiche nella mortalità.

La provincia di Cagliari mostra un aumento del livello generale di mortalità, permanendo comunque al di sotto della media provinciale. Per quanto riguarda la curva di mortalità, osservando i valori per il 2020 e il 2021 con i relativi grafici delle due componenti, si deduce una quota di mortalità relativamente bassa per la maggior parte delle età, eccetto per la classe 25-34 anni.

Per la provincia di Ferrara si può notare una parziale stabilità della mortalità sotto la media per tutti gli anni, anche se i loadings della seconda componente suggeriscono tassi di mortalità relativamente alti nella fascia tra i 20 e i 30 anni.

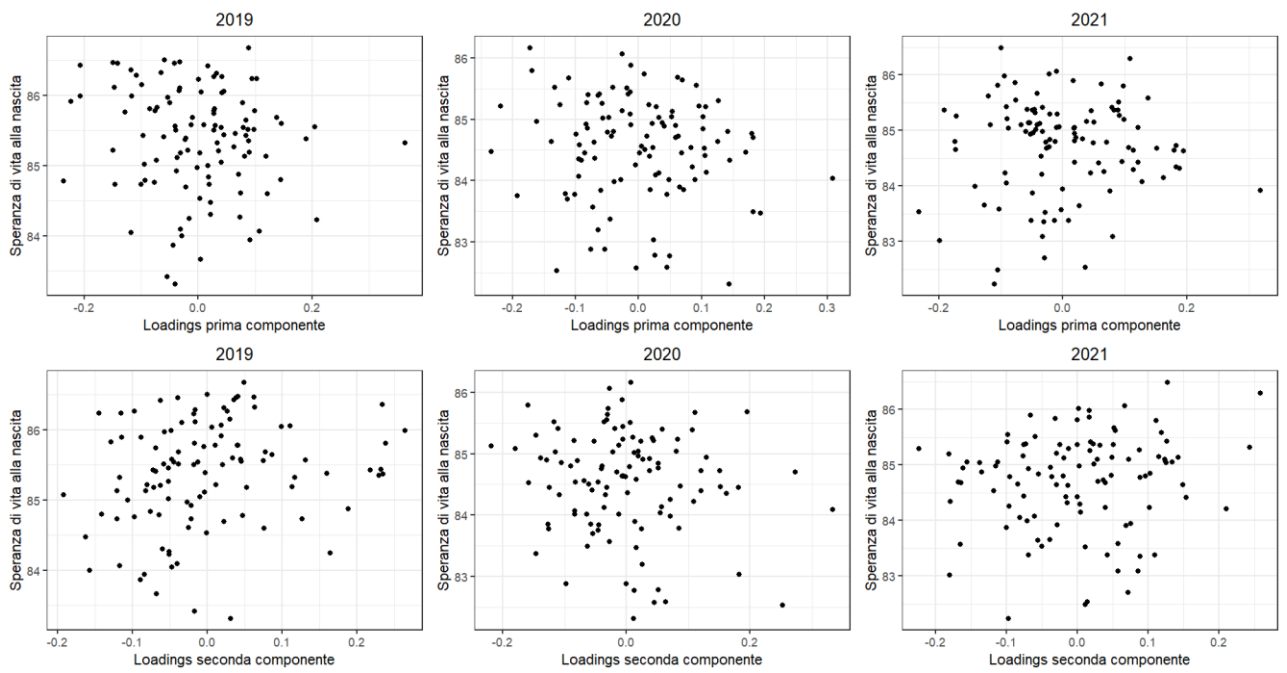
Per Rieti vale il contrario di quanto appena dichiarato, infatti i valori dei loadings hanno tutti segno opposto rispetto a quelli di Ferrara: ne consegue una mortalità complessiva superiore alla media e un grado di mortalità relativamente basso per le ragazze sui 20 anni.

	Prima componente			Seconda componente		
	2019	2020	2021	2019	2020	2021
<b>Cagliari</b>	-0.223	-0.169	-0.099	0.019	-0.16	0.127
<b>Ferrara</b>	-0.073	-0.07	-0.093	-0.192	0.002	0.102
<b>Rieti</b>	0.071	0.17	0.18	0.189	0.152	-0.029

**Tabella 3.2.1:** *loadings delle prime due componenti per alcune province, femmine, 2019-2021*

Infine, in Figura 3.2.4 si valuta la relazione tra la speranza di vita alla nascita e i loadings delle componenti principali. A differenza dei risultati osservati nel sesso maschile, non sembra esserci una chiara correlazione lineare tra le due variabili, rendendo di fatto complicato trarre delle conclusioni rilevanti, considerando che almeno per la prima componente ci si poteva aspettare una correlazione positiva.





**Figura 3.2.4:** relazione tra la speranza di vita alla nascita e i loadings riferiti alle prime due componenti principali, femmine, 2019-2021



# Conclusioni

In questa relazione si è discusso l'utilizzo di una tecnica di scomposizione utile per ridurre la dimensionalità del fenomeno della mortalità ed ottenere delle misure di sintesi in grado di descrivere le principali fonti di variazione della mortalità nelle varie età tra le province italiane, singolarmente per maschi e femmine, tecnica proposta da Alexander (2022) nel suo articolo.

Il punto di partenza sottostante a questa relazione è il fatto che la mortalità presenta caratteristiche differenti nel tempo e nello spazio in quanto influenzata da innumerevoli fattori. Questo avviene ovviamente anche all'interno di una stessa popolazione. Il Covid-19 è un elemento che ha evidentemente influenzato la mortalità, ma è fattuale che il virus non ha colpito in modo omogeneo, acuendo delle differenze che già preesistevano; questo viene dimostrato nel primo capitolo mostrando graficamente il diverso condizionamento del Covid-19 sull'aspettativa di vita alla nascita per le province con la migliore e la peggiore speranza di vita. Ovviamente questo risultato vale anche per tutte le province italiane.

La scomposizione della disuguaglianza di mortalità è stata resa possibile dalla decomposizione a valori singolari, una fattorizzazione matriciale che permette di catturare dei pattern complessi e non lineari, come nel caso della curva di mortalità, spiegandoli attraverso poche componenti. Per questo motivo la SVD è già uno strumento molto diffuso in demografia ma usato principalmente per ottenere stime e previsioni, mentre in questo elaborato ci si è focalizzati sullo spazio. La SVD è stata eseguita sulla matrice del logaritmo dei tassi di mortalità specifici per età a cui vengono sottratte le medie per ogni età. L'analisi è stata svolta negli anni 2019, 2020 e 2021, in modo tale da pescare l'impatto della pandemia sulla diversità nella mortalità.

Il risultato è stato la possibilità di riassumere la disparità di mortalità attraverso due componenti e i loadings associati, ovvero il contributo di ogni provincia alla creazione delle componenti principali. È da sottolineare un limite riscontrato nell'utilizzo di due sole componenti, ovvero la quota di varianza spiegata relativamente bassa che rimane sotto il 50%.

È emerso che nei maschi la prima componente, la quale rappresenta la principale fonte di variazione, esprime il livello di mortalità complessivo, evidenziando una forte disuguaglianza nelle fasce più giovani, nello specifico bambini e adolescenti, mentre la seconda componente fornisce informazioni riguardo alla forma della curva di mortalità, suggerendo importanti differenze anche tra le due fasce d'età menzionate precedentemente. Per entrambe le componenti risulta vero che l'impatto del Covid-19 è stato diminuire le disparità nelle fasce più giovani a discapito di un aumento nelle classi più adulte e negli anziani. Per le femmine i risultati ottenuti sono più difficili da interpretare, specialmente in termini di evoluzione nel tempo, comunque in generale sono simili ma con qualche differenza, in particolare per la seconda componente. Anche in suddetto caso valgono le conseguenze del Covid-19 esposte per il sesso maschile. Infine, per il sesso maschile è emersa una correlazione tra la speranza di vita alla nascita e i pesi delle province associati alle componenti principali, notificando che le province con la speranza di vita più alta tendono ad avere un grado di mortalità complessivo relativamente basso e sono inclini ad avere una mortalità tra i bambini sopra la media, ma relativamente bassa tra gli adolescenti. Non è stato possibile invece ottenere esiti in questo senso per il sesso femminile.

Negli ultimi anni i recenti studi hanno spostato il focus dall'analisi del progresso della mortalità generale alla ricerca di disuguaglianze che esistono tra le popolazioni. Questo comporta la necessità di nuove metodologie e modelli che permettano il confronto nel tempo e nello spazio, come quello svolto in questa relazione. Dato che questo modello può essere applicato per qualsiasi confronto territoriale, potrebbe essere interessante valutare dei paragoni anche a livello internazionale.

# Bibliografia

- Agabiti, N., Angelici, L., Cesaroni, G., Davoli, M. & Paglione, L. (2020). Mortality inequalities by occupational status and type of job in men and women: results from the Rome Longitudinal Study. *BMJ open*, **10(6)**.
- Alexander, M. (2022). Decomposing dimensions of mortality inequality. *Unpublished*.
- Alicandro, G., Costa, G., Demuru, E., Di Napoli, A., Frova, L., Marchetti, S., Marmot, M., Petrelli, A., Rossi, A., Rossi, P. G., Sebastiani, G. & Zengarini, N. (2019). Italian atlas of mortality inequalities by education level. *Epidemiologia e prevenzione*, **43(1)**.
- Carter, L. R. & Lee, R. D. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association* **87(419)**, 659–71.
- Caselli, G., Di Giulio, P. & Luy, M. (2011). Differences in life expectancy by education and occupation in Italy, 1980-1994: Indirect estimates from maternal and paternal orphanhood. *Population Studies* **65(2)**, 137-155.
- Islam, N., Jdanov, D. A., Kawachi, I., Khunti, K., Lacey, B., Lewington, S., Shkolnikov, V. M. & White, M. (2021). Effects of covid-19 pandemic on life expectancy and premature mortality in 2020: time series analysis in 37 countries. *BMJ*, **375**.
- Istituto Nazionale di Statistica. (1990). La mortalità differenziale secondo alcuni fattori socio-demografici, anni 1981-1982. *Istituto Nazionale di Statistica*, Roma.
- Istituto Nazionale di Statistica. (2001). La mortalità differenziale secondo alcuni fattori socio-demografici, anni 1991-1992. *Istituto Nazionale di Statistica*, Roma.
- Johnson, R. A. & Wichern, D. W. (2014). Applied multivariate statistical analysis (6° ed.). *Edinburgh: Pearson Education Limited*, 100-102, 430-469.
- Oeppen, J. & Vaupel, J. W. (2002). Broken Limits to Life Expectancy. *Science*, **296(5570)**, 1029-1031.
- Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. *New York: Springer*.



# Appendice

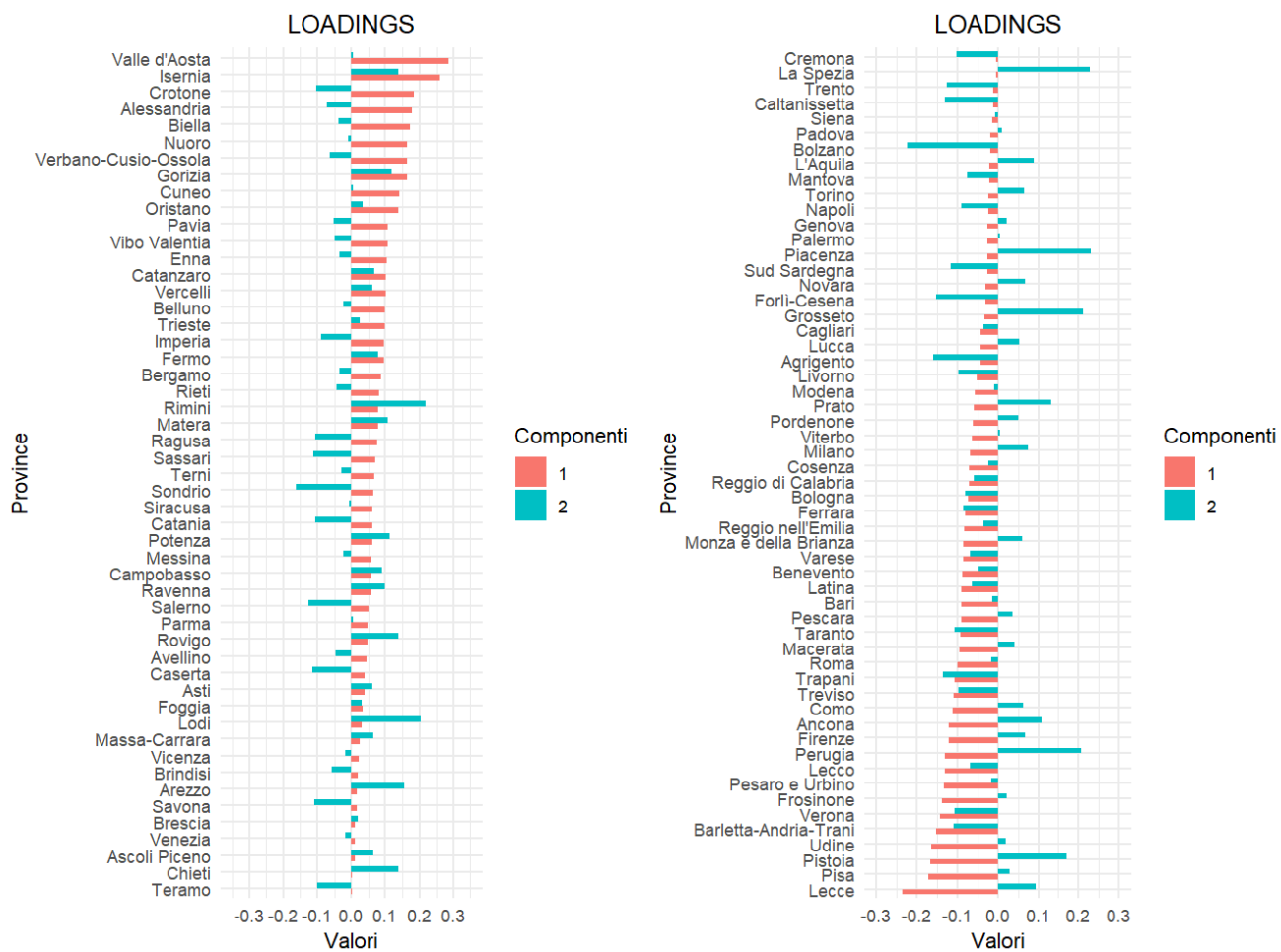
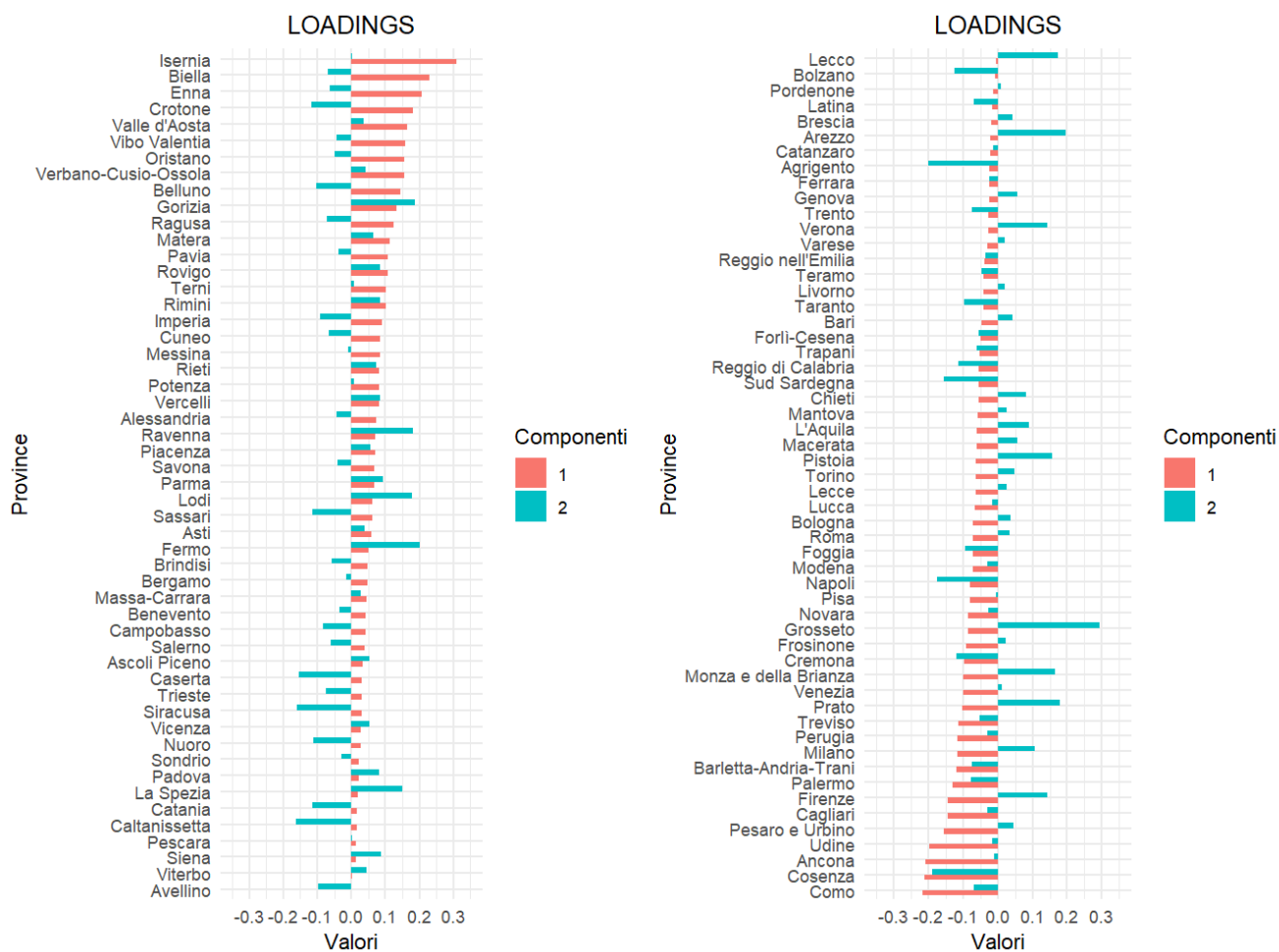
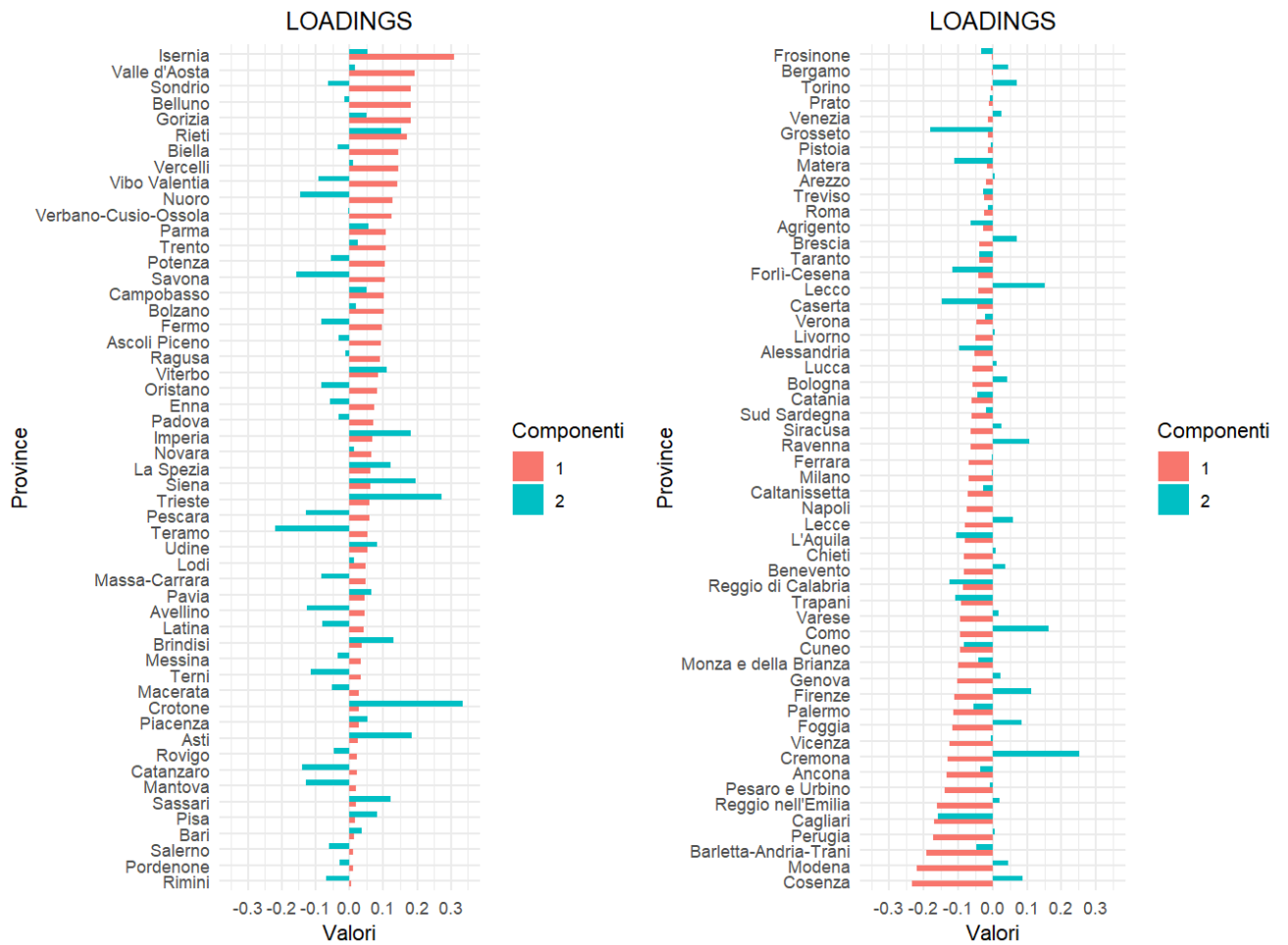


Figura A.1: loadings riferiti alla prima e seconda componente principale, maschi, 2020

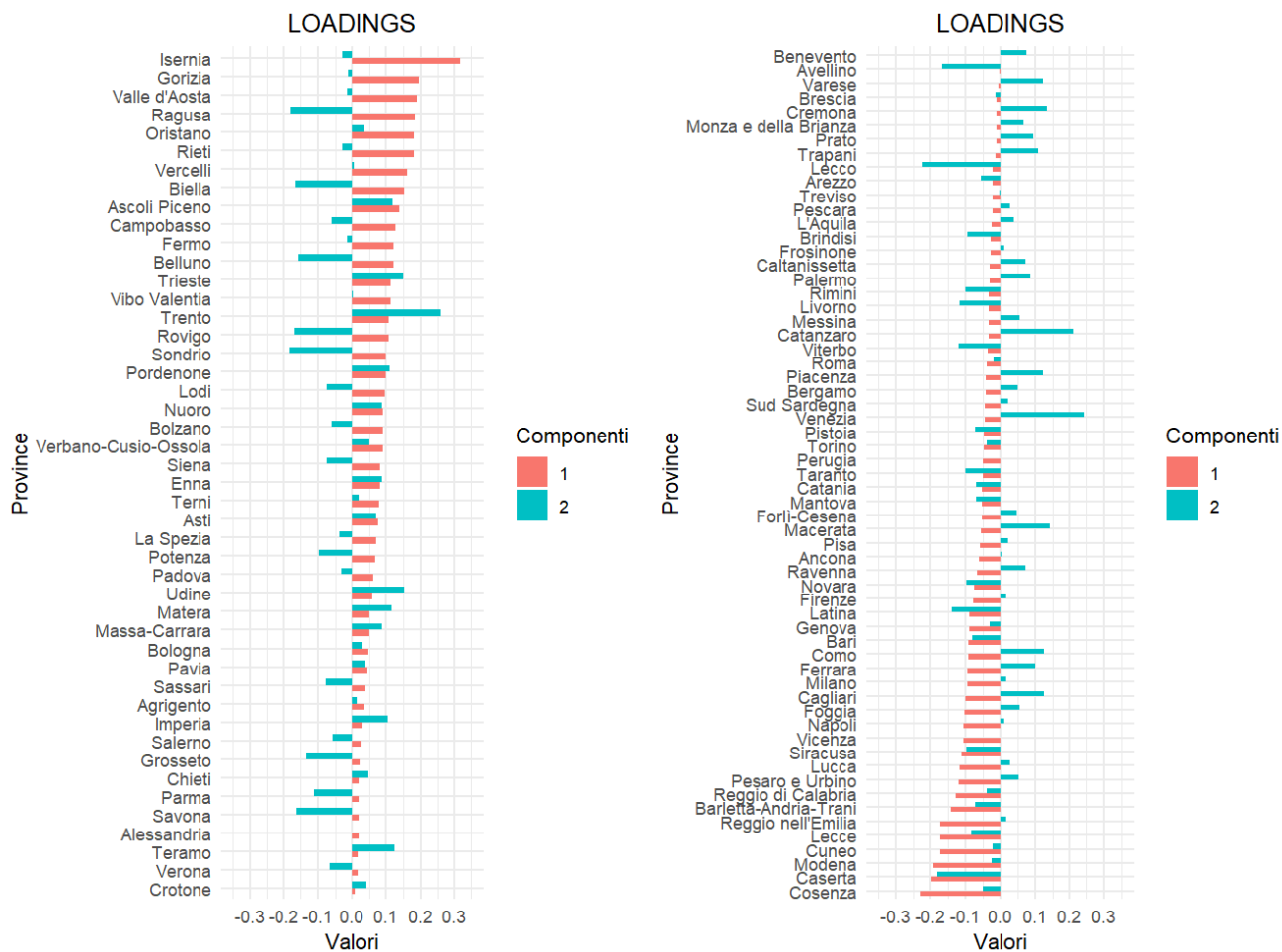


**Figura A.2:** loadings riferiti alla prima e seconda componente principale, maschi, 2021





**Figura A.3:** loadings riferiti alla prima e seconda componente principale, femmine, 2020



**Figura A.4:** loadings riferiti alla prima e seconda componente principale, femmine, 2021