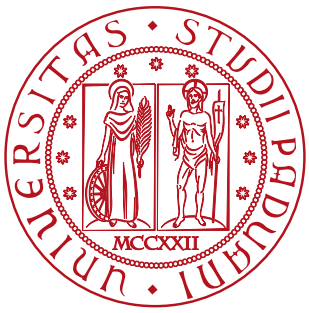
Università degli Studi di Padova

Dipartimento di Scienze Statistiche

Corso di Laurea Magistrale in
Scienze Statistiche



Modellazione di dati multivariati ordinali tramite tensori: un approccio Bayesiano non Parametrico

Relatore: Prof. Emanuele Aliverti
Dipartimento di Scienze Statistiche, Università degli Studi di Padova

Laureando: Riccardo Fassina
Matricola N. 2020314

Anno Accademico 2021/2022

Indice

Elenco delle figure

Elenco delle tabelle

Introduzione

In questi due anni di emergenza sanitaria, dovuta alla diffusione a livello globale del Corona-virus, le vite di moltissime persone sono state stravolte. La capillarità del contagio è stata resa possibile dalle interazioni sociali che caratterizzano le vite di ciascuno di noi. Molti governi del mondo, stimolati anche da alcuni paesi che hanno fatto da precursori a questo tipo di manovre, hanno imposto limitazioni sulle abitudini di vita delle persone (*e.g. lockdown*). Nel presente lavoro si è quindi voluto analizzare in che modo i comportamenti dei cittadini sono cambiati. Più nel dettaglio, sono presenti analisi relative all'ottemperanza di alcune norme sanitarie, sull'opinione in merito all'operato del Governo Italiano e l'obbligatorietà, insieme alla fiducia, dei vaccini. I dati che hanno reso possibile la trattazione di quanto descritto, sono stati collezionati durante lo stato di emergenza dall'Institute of Global Health Innovation (Jones Sarah P. e Plc. 2020) attraverso apposito sondaggio. Le risposte fornite dai soggetti, di natura categoriale ordinale, sono state scelte per catalogare i rispondenti in scale di adempimento o approvazione in merito a degli interrogativi cuciti sugli obiettivi sopra citati. Alcuni indicatori socio-demografici sono altresì disponibili: grazie ad essi è stato possibile capire in che modo l'età, il genere, la regione di provenienza e lo stato occupazionale potessero determinare il comportamento osservato durante l'emergenza.

Il raggruppamento di cui si parla è stato reso possibile grazie ad una Mistura standard di modelli di regressione di esperti, in cui ciascun *kernel* è stato definito con distribuzione Normale troncata: la natura ordinale delle risposte è cosi stata tenuta in considerazione. Il modello ipotizza pertanto che vi siano H profili latenti diversi, ciascuno corrispondente a diversi gradi di conformità alle misure imposte per evitare il diffondersi del virus. Si vedrà in realtà che le

sotto-popolazioni determinate si differenziano in quanto a sentimenti verso l'Esecutivo e su una presa di posizione in relazione ai vaccini. In particolar modo, l'approccio alla modellazione seguito è di tipo Bayesiano non parametrico che, almeno a livello teorico, fa divergere H ad infinito, lasciando che siano i dati a guidarci verso la scelta di questo valore. Tuttavia, attraverso l'utilizzo del processo di Dirichlet e della stick-breaking prior, si è potuto troncare questo numero ad una quantità finita.

Capitolo 1

COVID-19 e conseguenze sul tessuto sociale

Il COVID-19, il cui acronimo sta per COrona VIrus Disease 19, è una malattia infettiva respiratoria causata dal virus denominato SARS-CoV-2, che risulta appartenere alla famiglia dei coronavirus. Poiché il virus è trasmissibile per via aerea, tramite le cosiddette *droplet*, a loro volta emesse attraverso starnuti o tosse, è di facile comprensione la trasmissibilità dell'agente patogeno. A conseguenza di ciò, i governi dei diversi paesi colpiti hanno dovuto adottare, promuovendo, e per taluni periodi imponendo, opportune misure di distanziamento sociale che in qualche modo potessero prevenire il diffondersi del contagio. Citando alcuni numeri, relativi al periodo che intercorre tra il 20 ed il 27 luglio 2022, si può affermare che circa il 75% dei casi ha presentato stato iniziale asintomatico, mentre si sono registrati, dal 11 al 24 luglio, 738 deceduti su 1.105.799 nuovi casi (ISS 2022).
Relativamente al virus, il tratto respiratorio superiore ed inferiore sono i principali target, ma in generale tutti gli organi e apparati potrebbero venire colpiti. Tra i sintomi influenzali presenti si annoverano: febbre, tosse, cefalea (mal di testa), dispnea (respiro corto), artralgie e mialgie (dolore ad articolazioni e ai muscoli), astenia (stanchezza) e disturbi gastrointestinali come la diarrea. Riportando invece i sintomi caratteristici della patologia COVID, si hanno la perdita dell'olfatto (anosmia) e la perdita del gusto (ageusia), le quali risultano essere entrambe transitorie nella maggior parte dei casi.

1.1 Motivazioni delle analisi perseguite

Il COVID-19 è stato, e per certi versi continua ad essere, uno shock economico e sociale (Delanty 2021). Le misure preventive adottate da ciascuno Stato hanno avuto l'effetto di cessare la normale vita per come la si è conosciuta prima del 11 marzo 2020, giorno in cui il WHO (World-Health-Organization) ha dichiarato lo stato di pandemia. Ciò che all'inizio è sembrato essere in tutto e per tutto l'inizio di una dittatura, talvolta definita sanitaria, è stato, giorno dopo giorno, accettato socialmente. A tal proposito, della documentazione che quantifica e qualifica i cambiamenti delle abitudini può essere fornita da Google; questa, tramite l'impostazione *Cronologia delle posizioni* ed attraverso la creazione di report, basati su dati aggregati, anonimizzati e suddivisi per paese e regione, certifica tali cambiamenti delle routine (Google 2022).

Tale emergenza è una sfida sotto punti di vista sociali e politici: si prevede che le conseguenze economiche e sociali si possano perpetrare persino di più della pandemia stessa. Diversamente da lavori che trattano il Corona-virus da un punto di vista epidemiologico, nel presente elaborato lo si esamina sotto una lente sociologica. Non si tratta infatti unicamente di una questione biologica, ma anche sociale considerando che le infezioni virali sono trasmesse attraverso le interazioni umane.

Come ulteriore aggravante della situazione precaria in cui hanno versato i paesi del mondo negli ultimi due anni, vi sono consistenti differenze con taluni disastri passati (J. D. Osofsky, H. J. Osofsky e Mamon 2020). Per fare un esempio, è risaputo come le calamità naturali colpiscono solamente determinate aree di una comunità, stato o paese, permettendo a quelle non colpite di prestare soccorso. Inoltre, la previsione della loro durata e dei processi di ricostruzione è meno soggetta ad errore: mettere la parola "fine" ad uno stato pandemico non è invece semplice. Un ulteriore differenza, strettamente collegata alla prima, riguarda senz'altro il supporto psicologico non ottenibile dalle relazioni inter-personali nel caso pandemico. L'approccio all'emergenza, ha pertanto assunto connotati dissimili da quelli adottati in precedenza.

Ad ulteriore supporto dell'analisi sociologica perseguita, vi è della letteratura che studia come le reazioni psicologiche delle persone giocano un ruolo fondamentale nell'incentivare o meno la diffusione del virus, nella compar-

sa di stress emotivo, ma anche nell'insorgenza di disordini sociali durante e dopo la pandemia (Cullen, Gulati e Kelly 2020). I bisogni psicologici e psichiatrici delle persone non dovrebbero essere trascurati in nessuna delle fasi di gestione della pandemia. I fattori psicologici possono infatti influenzare sia il livello di ottemperanza alle misure di salute pubblica (*e.g.* vaccinazioni), che il rischio di infezione. Inoltre, le persone che già di per sé sono affette da problemi psicologici sono ancora più vulnerabili. Prendere consapevolezza di ciò, ed agire di conseguenza è quindi funzionale ad una quanto più ampia accettazione possibile delle manovre attuate dal Governo (Renault et al. 2022).

Un ulteriore studio (Forte et al. 2020), ha invece dimostrato come, su suolo italiano, le giovani studentesse siano associate ad un maggiore impatto psicologico derivante dall'emergenza. Sembra inoltre che nelle diverse regioni non vi siano differenze così marcate in termini di sintomi psicopatologici e disturbi post-traumatici da stress. Si è quindi potuto evincere che lo stato psicologico non è solamente stato influenzato dalla paura del contagio, ma anche dalle misure restrittive che hanno colpito in modo equo i cittadini delle diverse regioni. Quanto visto per il genere femminile lo si deduce anche da altri studi come, ad esempio, Moccia et al. (2020) ed altri.

Un obiettivo simile è perseguito da Carlucci, D'Ambrosio e Balsamo (2020): in questo lavoro viene infatti studiata l'ottemperanza alle misure di quarantena in base ad alcune caratteristiche socio-demografiche. Ciò che si evince è che le persone di genere femminile, con alti livelli di istruzione, di mezza età e residenti nelle regioni del Sud Italia, sembrano essere più propense a rispettare le misure di quarantena. Si veda anche Wolff et al. (2020).

Le misure di igiene personale, tra cui il disinfettarsi le mani, sono la principale arma che un cittadino ha nel prevenire il contagio. In particolare, in Güner, Hasanoğlu e Aktaş (2020), viene dimostrato come tale pratica sia, oltre a quarantene e distanziamento sociale, uno degli strumenti più efficaci.

Un altro strumento rivelatosi efficace nel dare un freno all'ascesa del numero di contagi è la mascherina. L'accettazione sociale di quest'ultima è stata analizzata in diversi lavori, tra cui Carbon (2021).

Per quanto concerne i vaccini invece, in articoli come Sallam (2021), viene prodotta una rassegna delle percentuali di accettazione, differenziate per paese, dei vaccini. Da quest'ultimo si evince come l'Italia sia uno dei paesi con

percentuale di approvazione del vaccino più bassa, raggiungendo un tasso di appena il 53.7%.

In Fujii, Suzuki e Niimi (2021), attraverso un confronto tra molteplici paesi, viene dimostrato come, le persone che hanno ben chiara l'efficacia dell'indossare la mascherina, del disinfettare le mani, e dell'evitare assembramenti siano più propense a seguire tali normative.

La presente rassegna è ciò che ha dato fondamenta allo studio condotto: confermare o contestare, magari arricchendo, i risultati già evincibili dalla vastissima letteratura, con accezione sociologica, riguardante il COVID-19. In particolare, l'attenzione è stata posta principalmente nel relazionare indicatori socio-demografici con l'accettazione, da parte dei cittadini, di norme sanitarie e vaccini. Un giudizio sulla gestione dell'emergenza da parte del Governo Italiano ha altresì trovato posto nelle analisi effettuate.

1.2 Analisi preliminari

Per questo elaborato si è scelto come paese l'Italia: questo è stato il paese colpito immediatamente dopo la Cina; inoltre, per questo Stato, il Governo ha adottato politiche che hanno cambiato radicalmente le abitudini dei cittadini, tramite l'imposizione di restrizioni, talvolta anche ben più aspre se raffrontate con altri paesi. Diverse sono state le misure di contenimento adottate dal Governo Italiano per i due governi susseguitisi dal 2020 al 2022, tutte tarate sulla base dell'andamento dell'epidemia. Tra gli indici utilizzati nella calibrazione delle misure imposte, si citano (COVIDSTAT 2022):

- I_t come il numero di persone contagiate al tempo t, in genere misurato in numero di giorni. Nei dati ufficiali viene usualmente riportata una stima, necessariamente per difetto;

- R_t come il numero medio di persone che è in grado di contagiare un'altra persona contagiata al tempo t;

- *generation time s* che risulta essere la differenza tra il tempo di infezione di una persona contagiata e il tempo di infezione del suo contagiatore.

Come già detto, ai fini della tipologia di studio perseguito, ci si è concentrati maggiormente sulle manovre con impatto sociale, come le misure di *lock down*, di quarantena obbligatoria ed altre. Per una disamina completa di tutti i DPCM introducenti nuove restrizioni, dall'inizio della pandemia sino alla fine dello stato di emergenza, si faccia riferimento all'appendice A.

1.2.1 Raccolta dei dati

I dati di questo studio provengono da un sondaggio condotto da *YouGov* in collaborazione con l'*Institute of Global Health Innovation* (IGHI) presso l'Imperial College di Londra (Jones Sarah P. e Plc. 2020). In particolare, il sondaggio somministrato ai diversi soggetti ha avuto il fine ultimo di presentare analisi comportamentali relativamente a come le diverse popolazioni mondiali hanno risposto alla pandemia; questi dati potrebbero quindi essere utili ai corpi di sanità pubblica per approvare manovre che siano adeguate nel limitare il più possibile l'impatto del COVID.

Le domande del sondaggio, somministrate dal IGHI, raccolgono dati relativi ai test, ai sintomi e alla volontà di auto-isolamento in risposta ai sintomi. Altre informazioni raccolte riguardano invece i comportamenti quotidiani, come ad esempio uscire di casa, lavorare fuori di casa, avere contatti con persone non del nucleo familiare, e la volontà di seguire misure di prevenzione comune. Ulteriori domande ponevano l'attenzione sull'opinione relativa ai vaccini e in merito alle decisioni prese dal Governo. Pertanto, le risposte dei soggetti, hanno consentito l'esplicitazione di eventuali legami tra di esse ed alcune variabili esplicative. La volontà di catalogare i rispondenti, in merito all'accettazione delle misure preventive, dell'operato dell'Esecutivo e dei vaccini, è ciò che ha orchestrato il tutto.

Le informazioni sopra descritte sono presenti per molti stati del mondo, ma si è deciso di trattare solamente i dati italiani per quanto già detto nella sezione 1.2. L'arco temporale considerato in questo studio è quello che va dal 27 dicembre 2021 fino al 25 marzo 2022. Questo trova motivazione nel fatto che tale periodo è l'unico comune nella rilevazione delle variabili risposta d'interesse. Inoltre, è opportuno far notare come, nel suddetto periodo, si era

già giunti ad uno stadio avanzato dell'epidemia e, con essa, della campagna vaccinale e delle normative per la prevenzione dal contagio.

1.2.2 Variabili risposta

Delle molte variabili disponibili nel dataset, si sono scelte quelle elencate nella tabella 1.1.

Tabella 1.1: Variabili risposta e relativa descrizione.

Codice questionario	**Descrizione**
`i12_health_1`	Hai indossato una mascherina fuori da casa tua.
Modalità:	Sempre, di frequente, qualche volta, raramente, mai.
`i12_health_3`	Hai utilizzato il disinfettante per mani.
Modalità:	Sempre, di frequente, qualche volta, raramente, mai.
`i12_health_4`	Hai coperto il tuo naso e la bocca quando hai tossito o starnutito.
Modalità:	Sempre, di frequente, qualche volta, raramente, mai.
`i12_health_5`	Evitato il contatto con persone che hanno sintomi o che hanno avuto contatti.
Modalità:	Sempre, di frequente, qualche volta, raramente, mai.
`WCRex1`	Il Governo ha gestito bene la questione Covid.
Modalità:	Molto bene, bene, male, molto male, non so.
`r1_1`	Il Coronavirus è molto pericoloso.
Modalità:	1 - Disaccordo,...,7 - Accordo.
`vac7`	Quanto ti fidi dei vaccini per il Covid.
Modalità:	Per nulla, poco, moderatamente, molto.
`r1_8`	Essere vaccinato ti protegge contro il Covid.
Modalità:	1 - Disaccordo,...,7 - Accordo.
`vac_man_99`	Io non penso che le vaccinazioni debbano essere obbligatorie per nessuno.
Modalità:	No, si.

Nella figura 1.1 e 1.2, è possibile prendere visione delle distribuzioni marginali di ciascuna variabile risposta. Si può chiaramente vedere come per le prime 4 variabili, la frequenza relativa alla modalità "Sempre" sia nettamente maggiore, se raffrontata con le altre 4 modalità presenti. Cercando di dare una prima interpretazione descrittiva di questo fatto, si potrebbe dire che, almeno per le prime quattro misure imposte, ci sia una forte prevalenza di soggetti che le rispettano rigorosamente. Per quanto concerne invece la fiducia nella gestione del Covid da parte del governo, la maggior parte degli intervistati, sembra essere contenta, a cui segue una porzione non indifferente di soggetti che invece crede che il Governo abbia gestito male l'emergenza (49% vs 22%). Si può anche notare come, utilizzando una scala *Likert* da 1 a 7 per la variabile relativa alla pericolosità del COVID, la maggior parte dei rispondenti sembra voler prendere una posizione neutra, ossia la modalità 4, a cui segue la sottopopolazione che è molto d'accordo con l'affermazione proposta. Le rimanenti tre variabili considerate cercano invece di suddividere la popolazione in esame in base alla fiducia per i vaccini nel contrastare l'epidemia, e la relativa obbligatorietà imposta da alcuni governi. La maggior parte delle unità statistiche sembra fidarsi moderatamente dei vaccini, a cui segue invece che la frequenza più elevata la si osserva per chi è completamente d'accordo sul fatto che vaccinarsi protegga dal Corona-virus. Infine, circa l'82% dei rispondenti si trova d'accordo con l'obbligatorietà dei vaccini.

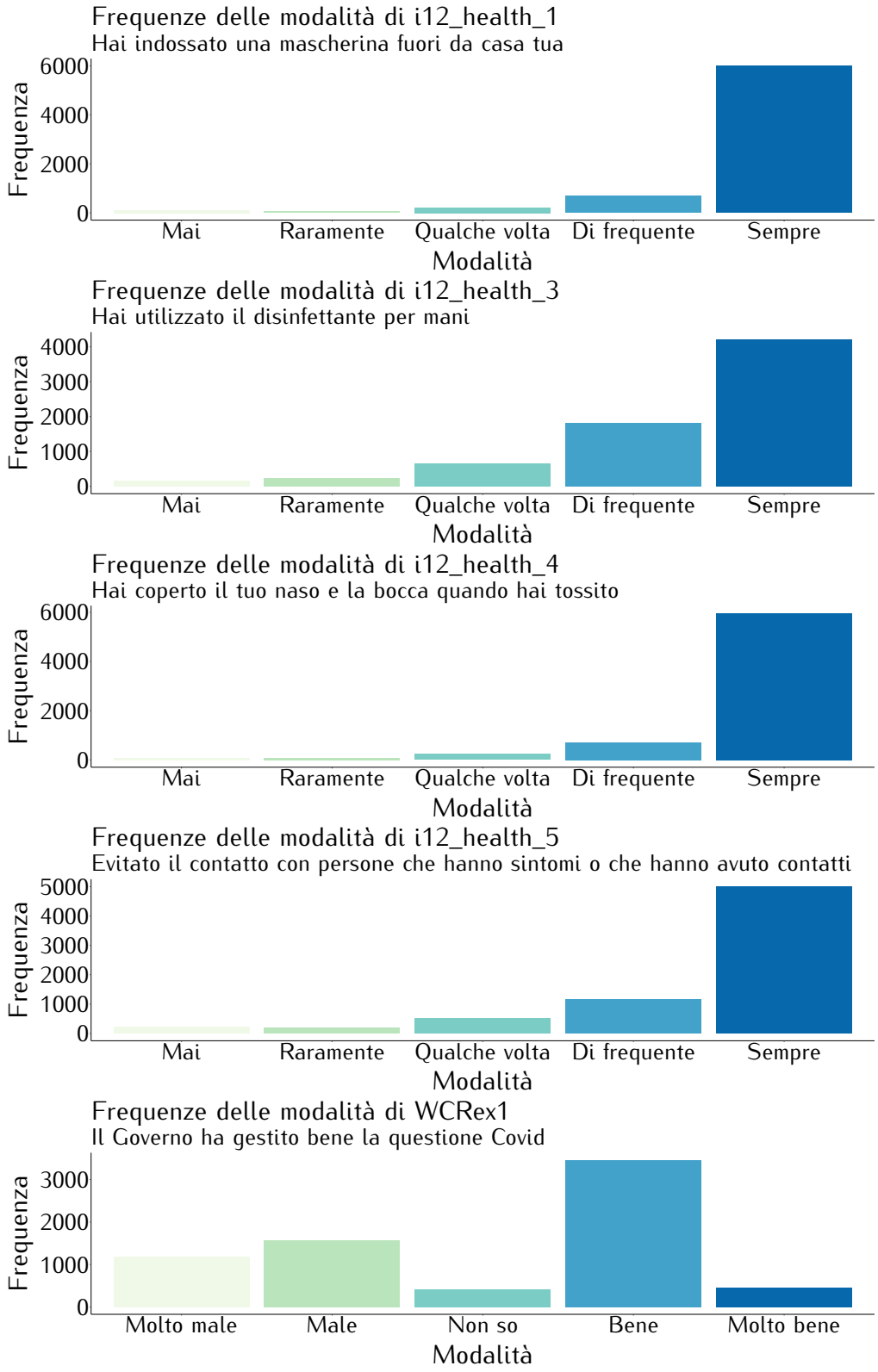


Figura 1.1: Analisi univariate variabili risposta 1 - 5.

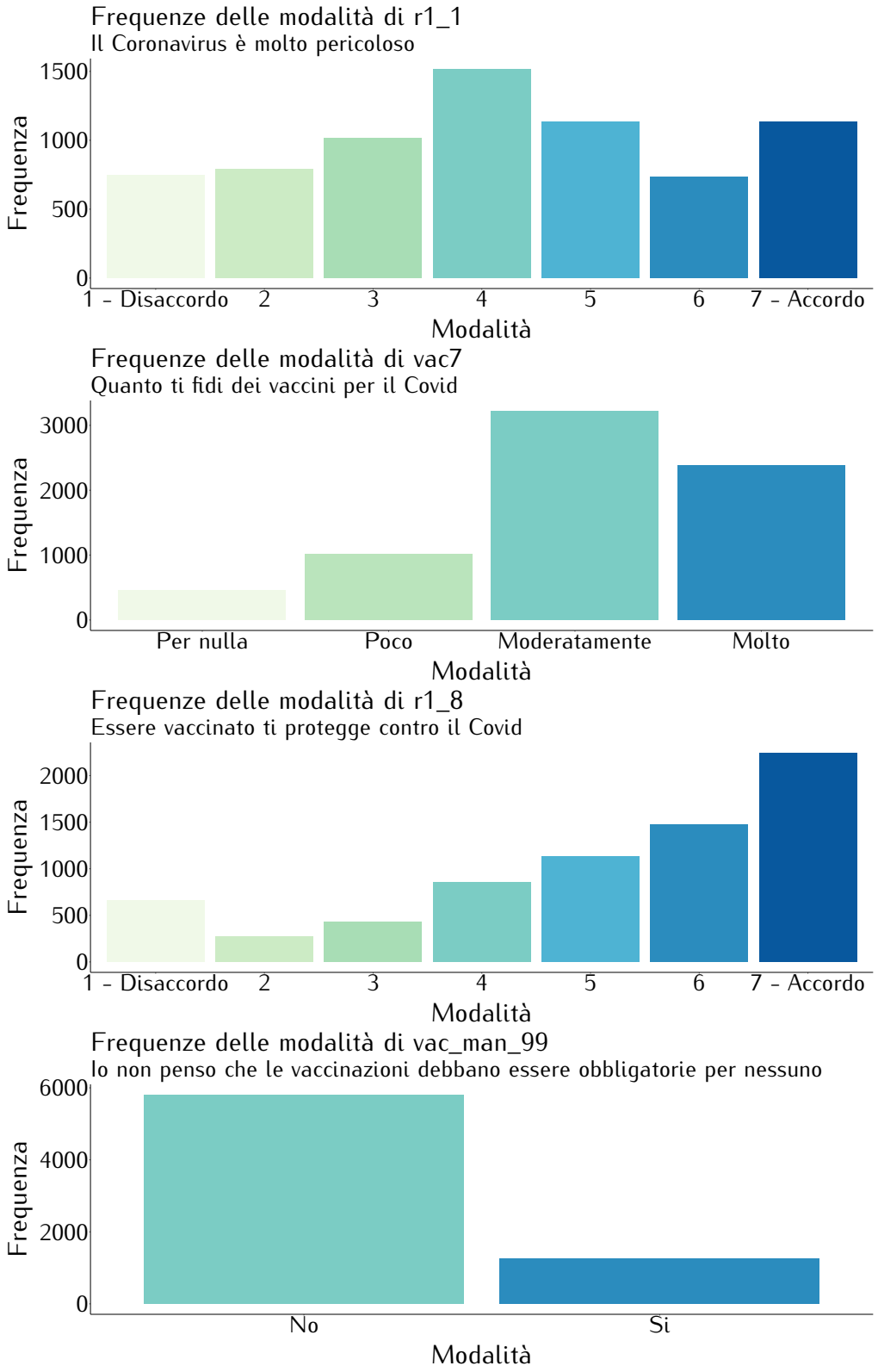


Figura 1.2: Analisi univariate variabili risposta 6 - 9.

1.2.3 Variabili esplicative

Alle variabili risposta si sono in seguito aggiunti degli indicatori socio - demografici, che aiutassero a categorizzare le unità statistiche osservate: di questi si può prendere visione alla tabella 1.2.
Con riferimento alla figura 1.3, si può osservare la distribuzione marginale di ciascuna variabile esplicativa in esame. In queste analisi occorre tener conto che chiunque non abbia accesso alla connessione internet non potrà partecipare al sondaggio (Quinn 2010). Il *violin-plot* relativo alla variabile Età riflette la piramide d'età dell'Italia. Le proporzioni di maschi e femmine sono invece del tutto analoghe. Lo stesso discorso vale per la distribuzione della variabile Regione: tutte e 5 le aree sono presenti con percentuali analoghe. La condizione occupazionale presenta, come modalità a maggior frequenza, il lavoratore a tempo pieno.

Tabella 1.2: Variabili esplicative e relativa descrizione.

Codice questionario	**Descrizione**
`Age`	Età.
`Gender`	Genere.
Modalità:	Maschio, femmina.
`Region`	Area Italiana in cui vivi.
Modalità:	Sud, Nord-Est, Nord-Ovest, Centro, Isole.
`Employment Status`	Condizione occupazionale.
Modalità:	Tempo pieno, Part-time, Studente a tempo pieno, In pensione, Disoccupato, Non lavoro, Altro.

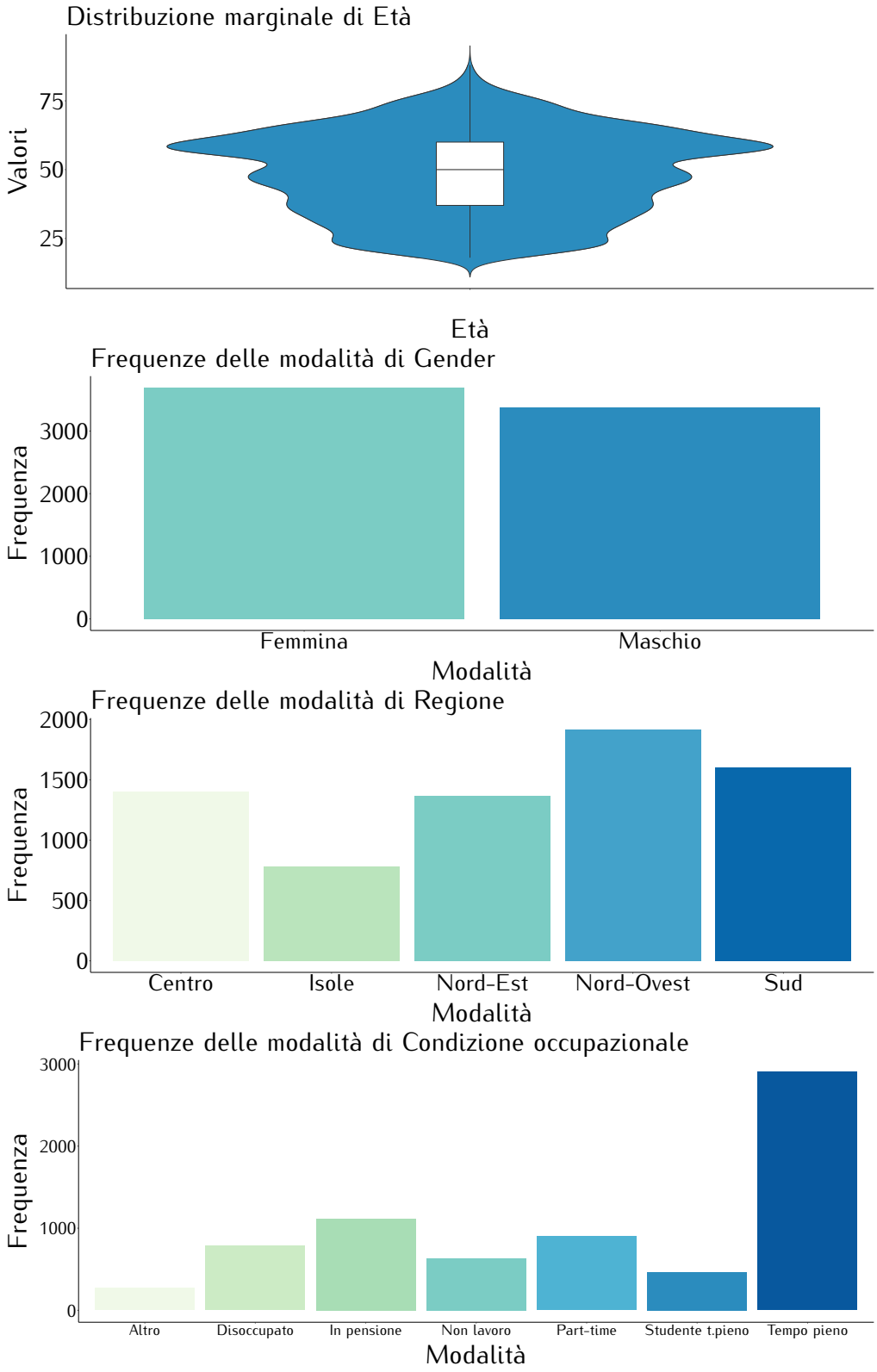


Figura 1.3: Analisi univariate variabili esplicative.

1.2.4 Analisi bivariate

Ulteriori analisi preliminari ritenute d'interesse e pertanto riportate, confrontano ciascuna variabile risposta dapprima con la variabile "Gender", e poi con l'età. I restanti indicatori socio-demografici ed i relativi grafici sono stati omessi in quanto trascurabili.

Genere

Da una disamina dei grafici riportati in figura 1.4 e 1.5 si deduce quanto segue. Nelle prime 4 domande, relative alle abitudini, i soggetti di genere femminile hanno risposto in percentuale maggiore rispetto ai maschi con la modalità "Sempre". Relativamente alla gestione del Covid da parte del Governo, sembrano invece essere i maschi ad esserne i più soddisfatti, con le modalità negative maggiormente assunte dalle femmine. La pericolosità del COVID sembra invece essere maggiormente percepita dalle femmine, suggerendo quindi come il malcontento di questi individui per la gestione del Covid possa voler proporre provvedimenti più restrittivi. Relativamente alla sicurezza e alla protezione offerta dai vaccini, i cittadini maschi sembrano essere i più fiduciosi. Infine, percentuali analoghe per le modalità "No" e "Si" alla domanda di obbligatorietà dei vaccini si evidenziano per i 2 generi.

Età

I *violin-plot* riportati in figura 1.6 e in 1.7 aiutano invece a capire la distribuzione dell'età dei rispondenti alle diverse domande poste, suddivise per modalità. Non sembrano essere presenti differenze marcate tra le mediane dell'età di ciascuna classe. La modalità più alta di ciascuna variabile viene assunta dai soggetti più anziani: la fragilità di questi individui li porta ad osservare più scrupolosamente le norme imposte. Per le restanti modalità e relativi grafici non sono invece presenti differenze così marcate. In ogni caso, molteplici forme di variabilità sono evincibili da ciascun *violin-plot*.

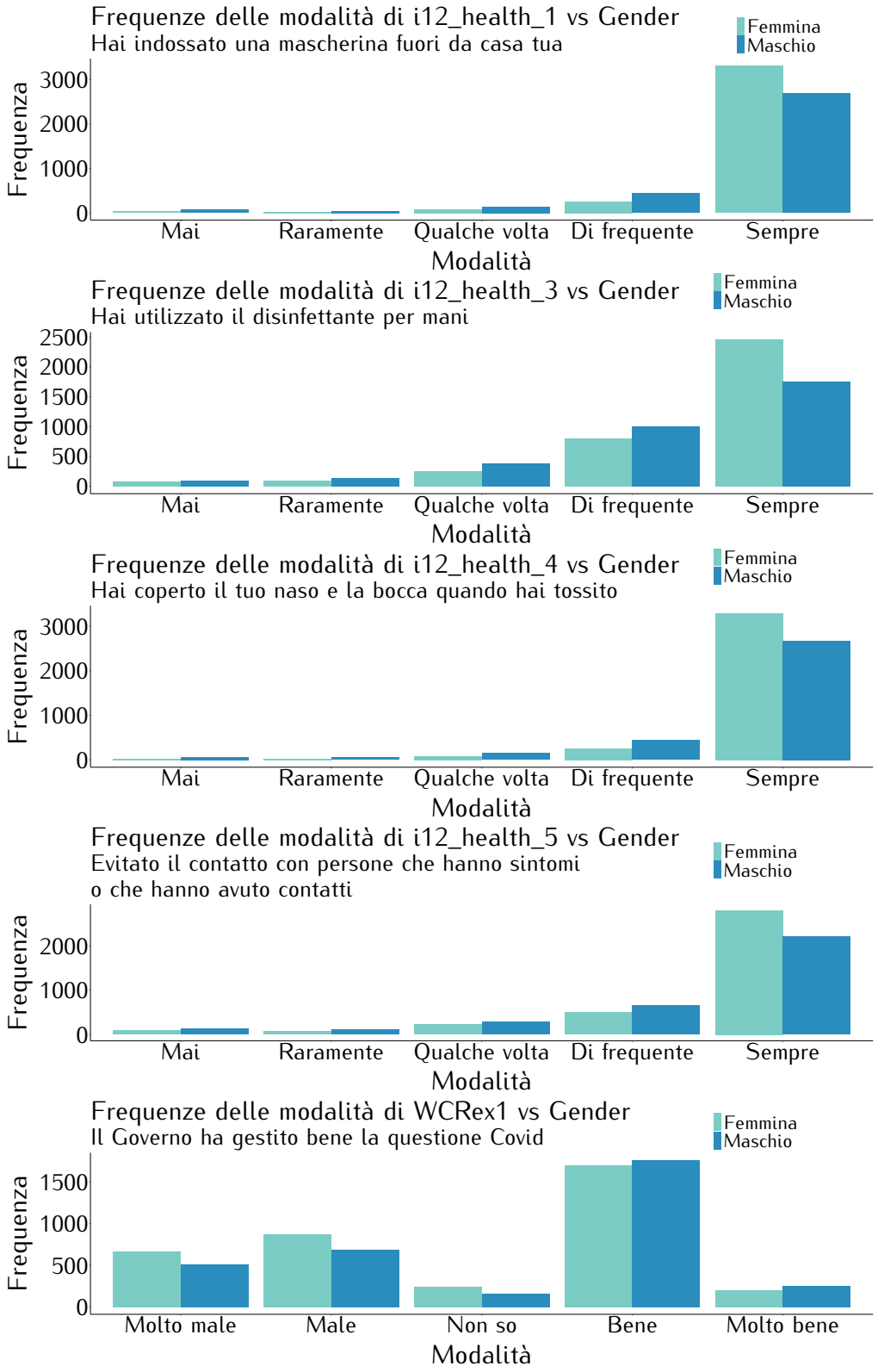


Figura 1.4: Analisi bivariate domande 1 - 5 vs Gender.

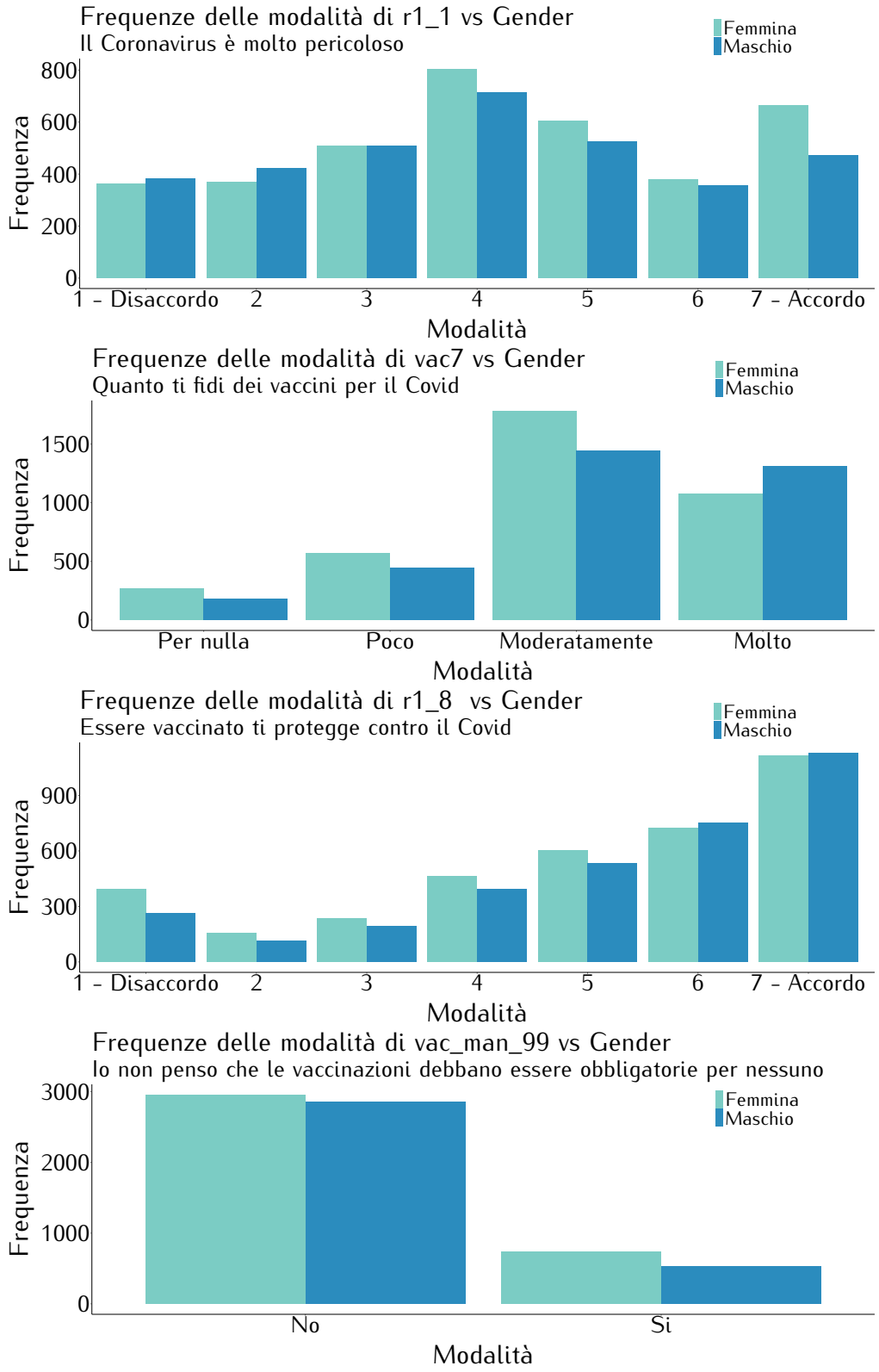


Figura 1.5: Analisi bivariate domande 6 - 9 vs Gender.

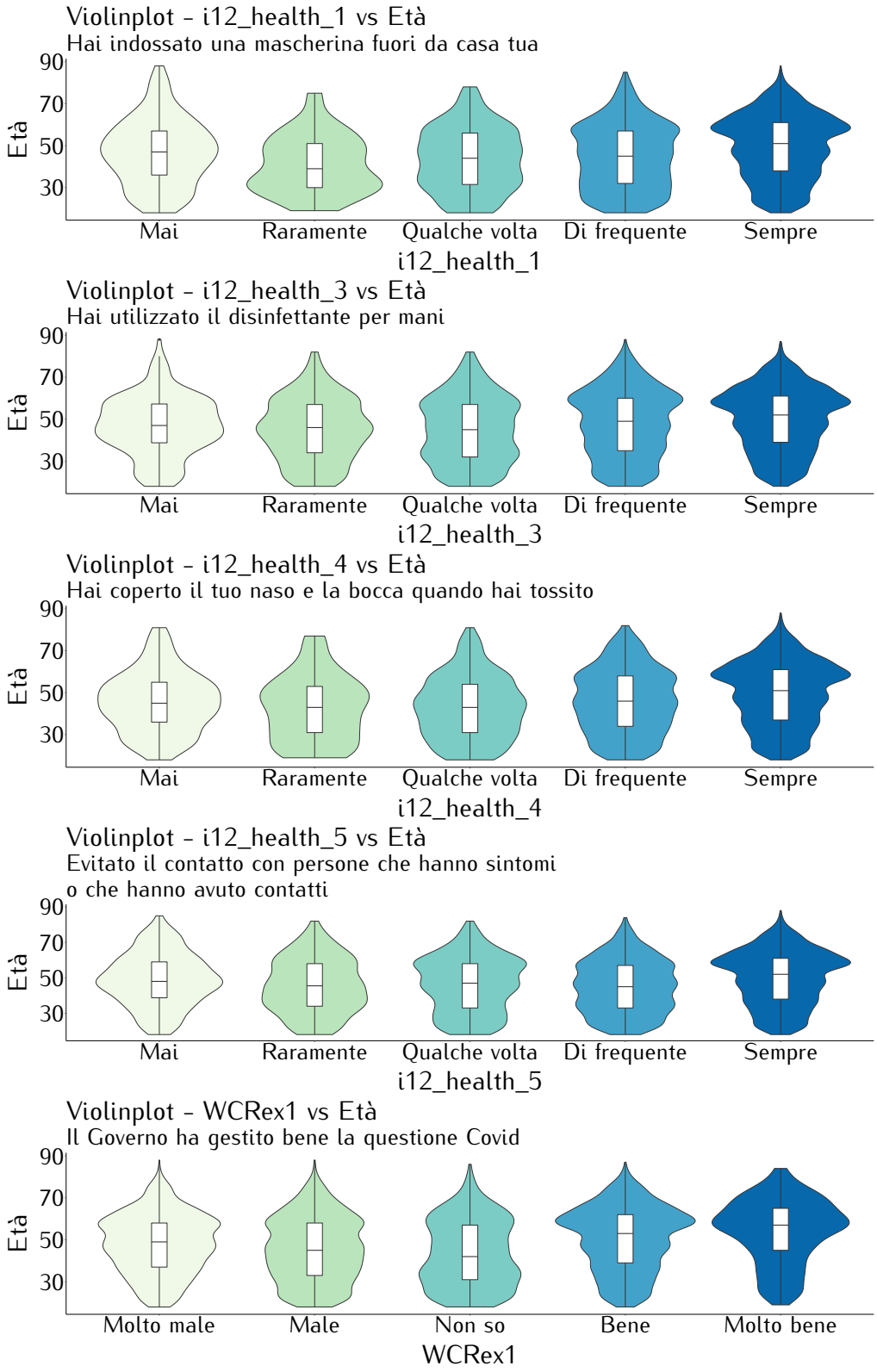


Figura 1.6: *Violin-plot* domande 1 - 5 vs Età.

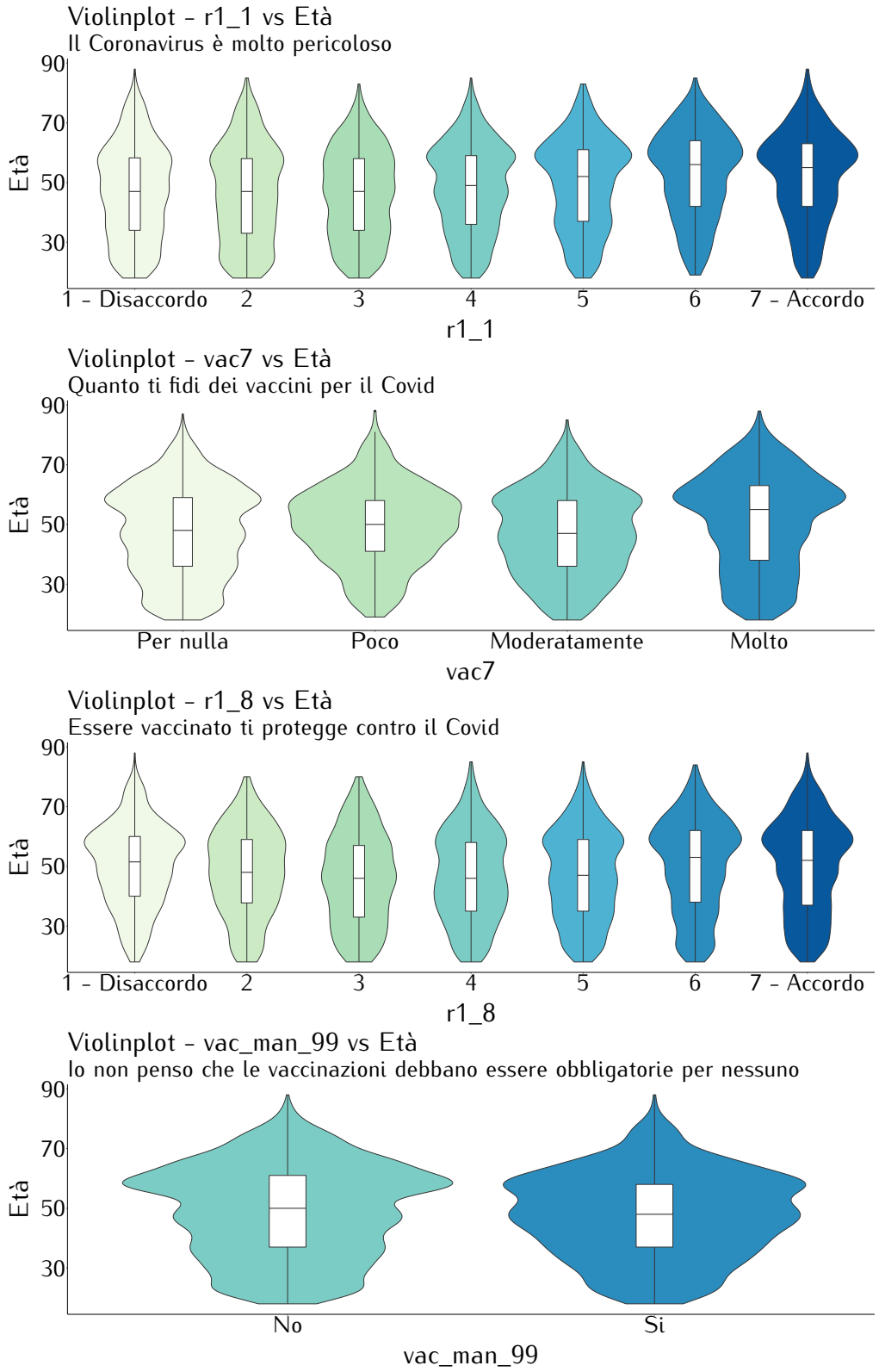


Figura 1.7: *Violin-plot* domande 6 - 9 vs Età.

Capitolo 2

Modelli per dati categoriali ordinali

Prima di poter presentare i modelli utilizzati è necessario contestualizzare e formalizzare il problema che si vuole affrontare. Come si è potuto evincere dalla sezione 1.2.1, tutte le variabili risposta considerate presentano natura categoriale, e dovranno quindi essere trattate di conseguenza.

È perciò possibile indicarle come $y_j,\ j = 1,\dots,9$ ciascuna avente d_j modalità. In particolare, ricordando quanto già visto nella tabella 1.1, si può scrivere $d_j = 5,\ j = 1,\dots,5,\ d_6 = 7,\ d_7 = 4,\ d_8 = 7,\ d_9 = 2$. Più nel dettaglio, con $N = 7072$ unità statistiche, specificando la risposta J-variata per l'i-esimo soggetto, si riporta

$$\boldsymbol{y}_i = (y_1,\dots,y_j,\dots,y_J), \quad J = 9,\ i = 1,\dots,N. \tag{2.1}$$

Per tale soggetto vengono poi osservate le variabili di cui alla tabella 1.2 e, formalizzando, si ha pertanto

$$\boldsymbol{x}_i = (x_1,\dots,x_p,\dots,x_P), \quad p = 1,\dots,P, \tag{2.2}$$

con $P = 4$. Vista la natura del dato osservato per $\boldsymbol{y}_i$, si è dovuto lavorare con una struttura che fosse in grado di ospitare un oggetto multi-dimensionale come quello descritto. In particolare, nel caso triviale con solamente due variabili risposta, si sarebbe utilizzata la classica tabella a doppia entrata; non essendo così in questo problema, si è optato per un tensore. Per immaginare questo oggetto, si pensi ad astrarre una classica matrice ad un iper-cubo J-dimensionale, con ciascuna di queste dimensioni aventi tante celle quan-

te le modalità d_j della variabile risposta j-esima. Il tensore in questione avrà quindi

$$d_1 \times \cdots \times d_j \times \cdots \times d_J = \bigtimes_{j=1}^{J} d_j \tag{2.3}$$

celle. Nello specifico, ciascuna di queste è atta a contenere la probabilità che l'i-esimo soggetto risponda con una certa combinazione di modalità a ciascuna delle risposte considerate. Indicando con $c_j \in \{1, \ldots, d_j\}$ la modalità di risposta alla domanda j-esima osservata per il soggetto i-esimo, la probabilità di cui si parlava poc'anzi viene riscritta come

$$\Pr(y_{i1} = c_{i1}, \ldots, y_{ij} = c_{ij}, \ldots, y_{iJ} = c_{iJ}). \tag{2.4}$$

A scopo illustrativo, nella figura 2.1, viene riportato un tensore tri-dimensionale.

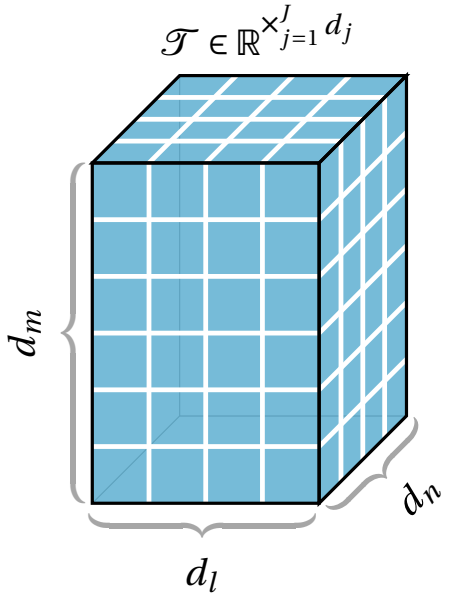


Figura 2.1: Rappresentazione grafica di un tensore tri-dimensionale.

Adoperando le covariate descritte in tabella 1.2, ed indicando con c_{ij} la modalità di risposta del soggetto i-esimo alla domanda j-esima, la modellazione avverrà sulla seguente quantità

$$\Pr(y_{i1} = c_{i1}, \ldots, y_{ij} = c_{ij}, \ldots, y_{iJ} = c_{iJ} \mid x_{i1}, \ldots, x_{ip}, \ldots, x_{iP}). \tag{2.5}$$

Vista la dimensione del tensore in esame, è presto comprensibile come questo presenterà molte celle contenenti frequenze nulle: la maggioranza delle combinazioni determinabili avrà probabilità ≈ 0 di occorrere. Si parla in questi casi di tensori ultra-sparsi (Kunihama e Dunson 2013).

2.1 Paradigma Bayesiano

Prima di poter trattare la strada scelta per modellare il problema presentato, occorre fare un breve *excursus* in merito alla procedura inferenziale di tipo Bayesiano.
Nell'inferenza Bayesiana, la probabilità viene intesa come incertezza sulle quantità osservabili, e su quelle non osservabili, ossia i parametri d'interesse θ. In sostanza, si assume che ciascun ignoto parametro θ sia realizzazione di una variabile casuale, con distribuzione sullo spazio parametrico Θ e con densità a priori $\pi(\theta)$. Quest'ultima ha il compito di riassumere l'informazione preliminare su θ.
Il modello statistico è specificato come $\mathscr{F} = \{p(\boldsymbol{y}_j;\theta),\ \theta \in \Theta,\ \boldsymbol{y}_j \in \mathscr{Y}_j\}$. Dato $\mathscr{F}$ e $\pi(\theta)$, l'approccio Bayesiano alla modellazione è completamente specificato. L'aggiornamento dell'informazione a priori avviene in seguito grazie ai dati osservati, $\boldsymbol{y}_j^{\text{OSS}}$, e alla diretta applicazione del teorema di Bayes, come riportato in 2.6.

$$\underbrace{\pi\left(\theta|\boldsymbol{y}_j^{\text{OSS}}\right)}_{\text{posteriori}} = \frac{\overbrace{p\left(\boldsymbol{y}_j^{\text{OSS}}|\theta\right)}^{\text{verosimiglianza}}\ \overbrace{\pi(\theta)}^{\text{priori}}}{\underbrace{\int_\Theta p\left(\boldsymbol{y}_j^{\text{OSS}}|\theta\right)\ \pi(\theta)\ d\theta}_{\text{costante di normalizzazione}}}. \tag{2.6}$$

2.1.1 Approccio Bayesiano Non Parametrico

I modelli e le tecniche rientranti sotto il nome di *Bayesian nonparametrics* (BNP) fanno parte di un campo di ricerca in Statistica, la cui attenzione ricade nella definizione ed utilizzo di distribuzioni a priori flessibili. In questa accezione quindi, l'etimologia del termine "non parametrico" è da attribuire al fatto che si usa un insieme non finito di parametri. Tra gli articoli che han-

no dato il via a questa branca, è doveroso citare i lavori di Ferguson (1973) e Doksum (1974).
Il processo di Dirichlet (DP) occupa una delle primissime posizioni in quanto a importanza tra queste distribuzioni. Da una primissima, e superficiale, disamina di questo processo, si ha che esso è un processo stocastico che genera misure i cui campioni sono $q.c.$ distribuzioni di probabilità discrete. Le misure casuali che vengono campionate dal DP sono quindi utilizzate per la costruzione di modelli più complicati, tra cui appunto le misure di mistura per un modello a mistura infinita.

In generale quindi, l'interesse dimostrato negli anni verso le tecniche BNP è largamente dovuto all'intrinseca capacità di adattamento a dati complessi, come ad esempio quelli del presente lavoro. In ambito BNP si identificano quindi strumenti utili sia a quantificare l'incertezza a posteriori, sia a garantire un modello altamente flessibile.

2.1.2 Il processo di Dirichlet

Il processo di Dirichlet nasce come un estensione non parametrica della distribuzione di Dirichlet nel simplesso avente la seguente esplicitazione

$$\Delta_H = \left\{ \boldsymbol{x} \in \mathbb{R}^{H^+} : \sum_{h=1}^{H} x_i = 1 \right\}. \tag{2.7}$$

Indicata con $\mathscr{P}$ la collezione di tutte le distribuzioni di probabilità in uno spazio misurabile $(\mathscr{X}, \mathscr{B})$, l'obiettivo per cui tale processo è stato introdotto è quello di specificare una distribuzione in $\mathscr{P}$ tale che le sue realizzazioni siano sufficientemente flessibili in supporto, mantenendo allo stesso tempo la trattabilità a livello analitico. Nel proseguo si dirà quindi che una misura casuale P segue un processo di Dirichlet con misura base P_0 e parametro di concentrazione α, ed indicata con $P \sim \mathscr{D}(\alpha, P_0)$ se, per ogni partizione finita $A_1, \ldots, A_H$ di $\mathscr{X}$, il vettore casuale di probabilità $\big(P(A_1), \ldots, P(A_H)\big)$ ha distribuzione

$$\big(P(A_1), \ldots, P(A_H)\big) \sim \mathscr{D}\big(\alpha P_0(A_1), \ldots, \alpha P_0(A_H)\big). \tag{2.8}$$

Come diretta conseguenza di questa definizione, il processo di Dirichlet è un processo stocastico, le cui realizzazioni sono distribuzioni discrete formate da un numero contabile di atomi (Ferguson 1973), come specificato nel seguito

$$P = \sum_{h=1}^{\infty} \pi_i \delta_{X_i}. \tag{2.9}$$

2.1.3 Rappresentazione Stick-breaking

Per poter campionare dal processo di Dirichlet, occorre un modo per poter definire i pesi. Nel dettaglio, ci si concentra nel campionamento di un insieme di valori indicati con $\{V_h, X_h^*\}$, per poi trovare una semplice regola che colleghi ogni V_h con l'associato peso π_h (Frigyik, Kapila e Gupta 2010). La rappresentazione stick-breaking di un processo di Dirichlet (Sethuraman 1994) viene usata per campionare da un DP come segue

1. cominciando da un bastoncino di dimensione unitaria, una porzione casuale di esso viene rotta. La dimensione V_1 è estratta da una Beta$(1,\alpha)$ con $\alpha > 0$ e viene associata ad un valore X_1^* campionato dalla distribuzione base P_0. Si definisce quindi $\pi_1 = V_1$ come la prima porzione rotta del segmento;

2. sia ora $1-V_1$ la dimensione del bastoncino rimanente. Un'altra porzione casuale, indicata con V_2 viene rotta da esso, in modo tale che la lunghezza rimanente sia ora uguale a $(1-V_1)(1-V_2)$, con il nuovo peso dato dalla nuova lunghezza di rottura $\pi_2 = V_2(1-V_1)$. Come in precedenza, V_2 viene associato ad un altro valore X_2^* campionato dalla distribuzione di base P_0.

Riassumendo quanto detto è possibile scrivere

$$\begin{aligned} &V_h \overset{iid}{\sim} \text{Beta}(1,\alpha), \quad X_h^* \overset{iid}{\sim} P_0, \\ &\pi_h = V_h \prod_{i=1}^{h-1}(1-V_j), \quad P = \sum_{h=1}^{\infty} \pi_k \delta_{X_k^*}. \end{aligned} \tag{2.10}$$

La figura 2.2 fornisce una rappresentazione grafica di quanto detto.

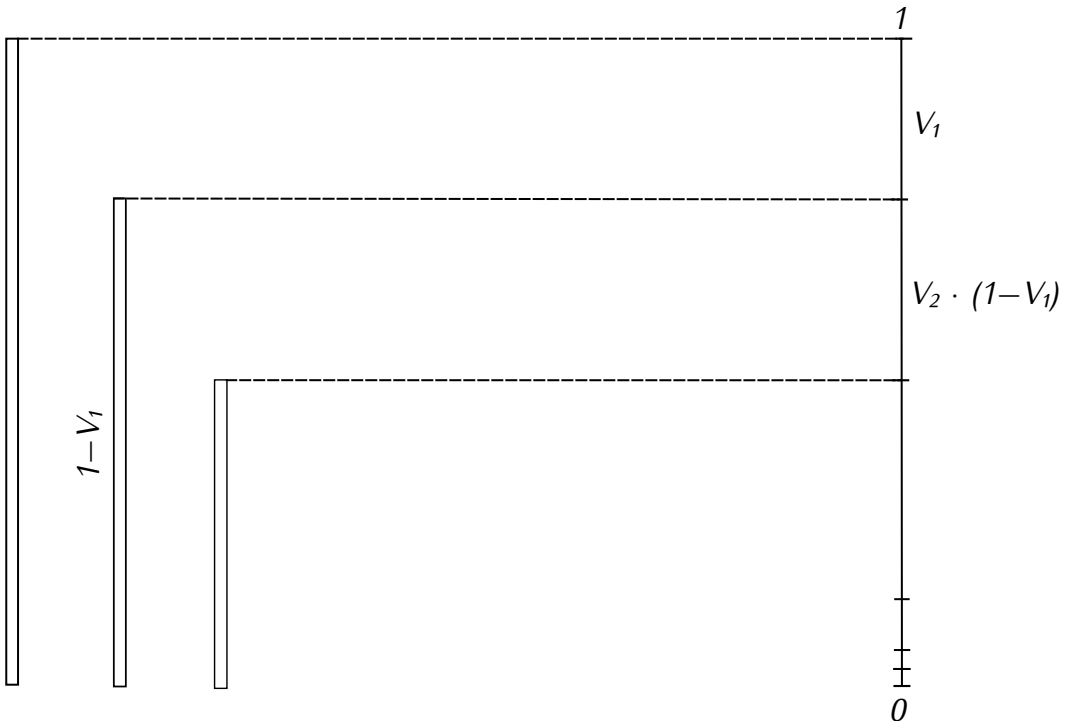


Figura 2.2: Rappresentazione grafica stick-breaking.

2.2 Modelli mistura

In antecedenza alla specificazione dei modelli mistura utilizzati, occorre delinerare l'obiettivo che si sta perseguendo, così da avere ben chiara la scelta di questa tipologia di modellazione.
Il fine ultimo del presente lavoro è quello di caratterizzare i comportamenti tenuti dalla popolazione italiana durante la pandemia. Per conseguire questo obiettivo, si è deciso di adottare un modello a bassa dimensionalità per dati multivariati categoriali in grado di caratterizzare la funzione di massa di probabilità per $\boldsymbol{y}_i$ in termini di un insieme di classi latenti (Aliverti e Russo 2022). Si assume quindi che la popolazione sia divisa in H comportamenti ideali, anche chiamati profili, che sono proprio le classi di cui si parlava poc'anzi. Questi H identikit possono esplicitarsi in diversi gradi di ottemperanza alle normative anti-Covid; *e.g.* le persone che osservano tutte le regole potrebbero caratterizzare un profilo, oppure le persone che si comportano esattamente come prima dell'inizio della pandemia potrebbero invece caratterizzare una secon-

da classe di soggetti. I modelli mistura quindi, essendo basati sull'assunzione che si cerca di modellare una popolazione composta da H sotto-popolazioni, sono lo strumento più naturale da utilizzare.

Il modello mistura adoperato fa utilizzo di informazioni derivanti dai regressori per costruire sia i *kernel* che i pesi di mistura. Esso risulta specificato come segue

$$p(y,z|x) = p(y|x,z)\,p(z|x) = \sum_{h=1}^{H} \nu_h(x)\,k_h(\cdot|x), \tag{2.11}$$

dove, $\nu_h(x)$ sono i pesi della mistura e, i $k_h(\cdot|x)$ sono invece i *kernel*. Vista la dipendenza sia da h che da x, si sta appunto indicando la dipendenza dei pesi e dei *kernel* dalla popolazione h-esima e dalle covariate x. A partire dalle suddette relazioni, tali modelli, nella letteratura, prendono il nome di "Mistura standard di modelli di regressione di esperti". Per ulteriori dettagli sui modelli mistura di esperti si veda, ad esempio, Fruhwirth-Schnatter, Celeux e Robert (2019).

2.2.1 Misture Bayesiane Non Parametriche

Con il fine ultimo di garantire maggior flessibilità nella modellazione dei dati, si è proceduto senza fissare un numero H di *kernel* di mistura, ma bensì lasciando che $H \to \infty$ con i pesi via via più piccoli e tendenti a 0, grazie alla rappresentazione stick-breaking.
Si può quindi scrivere, per ciascuna variabile risposta j-esima quanto segue

$$p(y_{ij}|\boldsymbol{\theta},\boldsymbol{\pi}) = \sum_{h=1}^{\infty} \pi_{hi}\,k(y_j;\boldsymbol{\theta}_h), \tag{2.12}$$

assegnando a $\boldsymbol{\pi}$ un'opportuna distribuzione a priori.
Si è poi adoperata una variabile latente categoriale discreta, denominata z_i, per l'assegnazione ad una delle H classi latenti di ciascun soggetto i, $i = 1,\dots,N$. Essa è specificata come segue

$$z_i \in \{1,\dots,h,\dots,H\} \text{ con probabilità } \pi_h(\boldsymbol{x}_i) = \Pr(z_i = h|\boldsymbol{x}_i). \tag{2.13}$$

Utilizzando poi tali quantità per la definizione dei $\nu_h(\boldsymbol{x}_i)$ partendo dal primo peso, ossia $\pi_1(\boldsymbol{x}_i) = \nu_1(\boldsymbol{x}_i)$. La collezione dei $\pi_h(\boldsymbol{x}_i)$, con $h = 1, \ldots, H$ definisce quindi tutto l'insieme dei pesi della mistura. La determinazione dei pesi successivi al primo, avviene come esplicitato in 2.10, con il generico valore così definito

$$\pi_h(\boldsymbol{x}_i) = \nu_h(\boldsymbol{x}_i) \prod_{j=1}^{h-1} \big(1 - \nu_j(\boldsymbol{x}_i)\big), \quad h \in \mathbb{N}. \tag{2.14}$$

Per la specificazione del generico $\nu_h(\boldsymbol{x}_i)$ si è scelto di adoperare la trasformazione *inverse-logit* utilizzata in Rigon e Durante (2021), ossia

$$\nu_h(\boldsymbol{x}_i) = \frac{\exp\big(\gamma_{0h} + \boldsymbol{x}_i \boldsymbol{\gamma}_h\big)}{1 + \exp\big(\gamma_{0h} + \boldsymbol{x}_i \boldsymbol{\gamma}_h\big)}, \tag{2.15}$$

questa ha il compito di mappare in $[0,1]$ il predittore lineare.
La dipendenza dei pesi di mistura, $\pi_h(\boldsymbol{x}_i)$, dalle covariate, definisce il cosiddetto covariate-dependent Dirichlet Process. Per una revisione di quest'ultimo si veda, ad esempio, F. A. Quintana et al. (2022).

2.2.2 Distribuzione Normale Troncata

Finora, nulla è stato detto sulle distribuzioni definenti i diversi *kernel*. In primo luogo si è considerata una distribuzione Normale per ciascuna delle J variabili risposta. La scelta di questa distribuzione è dovuta al fatto che, grazie ad essa, è possibile modellare una vasta gamma di dati. Comunque, da un'analisi superficiale del problema e delle relative domande, ci si può aspettare, viste anche le scale utilizzate, che i rispondenti tendano a distribuirsi al centro, ossia nelle modalità intermedie. Le variazioni da questo comportamento medio vengono quindi catturate dalla varianza della distribuzione Normale.
In ogni caso, la distribuzione Normale ha supporto in $(-\infty, \infty)$, e come tale assegna massa di probabilità positiva a ciascun valore dell'asse reale. Considerando che le variabili d'interesse con scala ordinale discretizzate presentano al più modalità con supporto discreto da 1 a 7, si è pensato di troncare la Normale in base alle modalità della variabile risposta y_{ij} che si sta modellando.

Più nel dettaglio quindi, i $J = 9$ supporti prefissati sono

$$s_j = [1,5],\ j = 1,\ldots,5,\ s_6 = s_8 = [1,7],\ s_7 = [1,4],\ s_9 = [1,2]. \tag{2.16}$$

Occorre ora fare una piccola digressione in merito ai valori di frontiera utilizzati, anche chiamati *cutoff*, per l'assegnazione di una modalità di risposta ad una domanda j per il soggetto i, ossia y_{ij}. Quando si modellano dati ordinali, non si hanno a disposizione le distanze tra una modalità e l'altra della variabile considerata; scegliere dei valori di *cutoff* prefissati non è pertanto propriamente corretto in quanto si starebbe implicitamente assumendo la conoscenza dell'esatta ubicazione di questi, o l'equidistanza tra le categorie. In approcci di tipo parametrico questi valori devono essere considerati come quantità casuali e, in quanto tali, devono essere opportunamente stimati; il processo di stima non è affatto semplice, vista anche la moltitudine di vincoli che si è tenuti a rispettare, tra gli altri, l'ordinalità che queste stime devono avere. Per ulteriori informazioni in merito si veda, ad esempio, Canale e Dunson (2011). In ogni caso, grazie ai risultati ottenuti in Kottas, Müller e F. Quintana (2005), si può aggirare il problema di stima considerando che, grazie all'approccio Bayesiano non parametrico che si sta utilizzando, il numero di componenti della mistura è casuale e, pertanto, non si ha perdita di generalità nel considerare come fissati i valori di *cutoff* definenti gli intervalli.
Ad esemplificazione di quanto detto, in figura 2.3 viene riportata la distribuzione Normale suddivisa in 7 intervalli tramite diversi valori di *cutoff*. I colori di questa corrispondono alle $d_j = 7$ modalità di y_{ij} con $j = 6,8$.

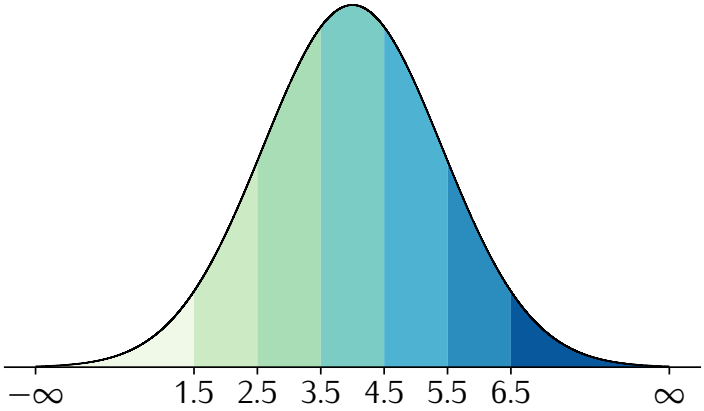


Figura 2.3: Grafico esemplificativo della suddivisione in 7 intervalli, tramite valori di *cutoff* equi-spaziati, di una distribuzione Normale.

Tornando al problema che si sta trattando, una volta arrotondato al valore successivo il valore ottenuto dalla j-esima Normale troncata con supporto s_j, si è pertanto ottenuta la modalità di risposta alla j-esima domanda del soggetto i-esimo. In 2.17 e 2.18 vengono riportate la funzione di densità di probabilità e la funzione di ripartizione della distribuzione Normale troncata (Burkardt 2014).

$$f_{Y_j=y_j}(Y_j = y_j; \mu, \sigma, a_j, b_j) = \begin{cases} 0 & \text{se } y_j \leq a_j, \\ \frac{\phi(y_j;\mu,\sigma^2)}{\Phi(b_j;\mu,\sigma^2)-\Phi(a_j;\mu,\sigma^2)} & \text{se } a_j < y_j < b_j, \\ 0 & \text{se } b_j \leq y_j. \end{cases} \quad (2.17)$$

$$F_{Y_j=y_j}(Y_j = y_j; \mu, \sigma, a_j, b_j) = \begin{cases} 0 & \text{se } y_j \leq a_j, \\ \frac{\Phi(y_j;\mu,\sigma^2)-\Phi(a_j;\mu,\sigma^2)}{\Phi(b_j;\mu,\sigma^2)-\Phi(a_j;\mu,\sigma^2)} & \text{se } a_j < y_j < b_j, \\ 1 & \text{se } b_j \leq y_j. \end{cases} \quad (2.18)$$

Si ha che $\phi(\cdot;\mu,\sigma^2)$ e $\Phi(\cdot;\mu,\sigma^2)$ sono, rispettivamente, la funzione di densità di probabilità e di ripartizione della Normale con media μ e varianza σ^2, $\mathcal{N}(\mu,\sigma^2)$. Invece, con riferimento ad a_j e b_j, questi sono gli estremi del gene-

rico supporto s_j.

2.2.3 Modellazione

Come visto, si è scelto di consentire la dipendenza dalle covariate nei pesi di mistura e nei *kernel*, adoperando così una Mistura standard di modelli di regressione di esperti.
Si ottiene pertanto

$$p(y_{ij}|x_{i1},\dots,x_{iP}) = \sum_{h=1}^{\infty} \pi_h(\boldsymbol{x}_i)\mathcal{NT}\left(\mu_h^{(j)}(\boldsymbol{x}_i), \sigma_h^2, a_j, b_j\right), \tag{2.19}$$

dove si indica con $\mathcal{NT}$ la distribuzione Normale Troncata. Relativamente alla specificazione per $\pi_h(\boldsymbol{x}_i)$, si faccia riferimento a quanto visto nella 2.14. Ponendo invece l'attenzione su $\mu_h^{(j)}(\boldsymbol{x}_i)$, ossia la media di ciascuna Normale troncata, si è scelto di adottare il seguente predittore lineare

$$\mu_h^{(j)}(\boldsymbol{x}_i) = \xi_{0h}^{(j)} + \boldsymbol{x}_i\boldsymbol{\xi}_h^{(j)}, \quad j = 1,\dots,J, \tag{2.20}$$

con $\xi_{0h}^{(j)}$ scalare, $\boldsymbol{x}_i \in \mathbb{R}^{13}$ vettore riga, ed infine $\boldsymbol{\xi}_h^{(j)} \in \mathbb{R}^{13}$ vettore colonna contenente i coefficienti relativi alle variabili osservate.
La suddetta modellazione della j-esima variabile risposta ha quindi visto l'utilizzo della distribuzione Normale troncata con media $\mu_h^{(j)}(\boldsymbol{x}_i)$, varianza σ_h^2 e supporto s_j definito dagli estremi a_j e b_j.
Raccogliendo tutte le quantità ignote, si può scrivere

$$\boldsymbol{\theta} = \left(z_i, \gamma_{0h}, \boldsymbol{\gamma}_h, \xi_{0h}^{(j)}, \boldsymbol{\xi}_h^{(j)}, \sigma_h^2\right), \tag{2.21}$$

con $\boldsymbol{\gamma}_h$ e $\boldsymbol{\xi}_h^{(j)}$ vettori $\in \mathbb{R}^{13}$ e con indici $h = 1,\dots,H,\ j = 1,\dots,J,\ i = 1,\dots,N$.

Relativamente all'elicitazione delle distribuzioni a priori si può tener conto o meno di informazioni pregresse sul problema. Poiché in questo caso non si ha a disposizione questo genere di conoscenza, si è optato per utilizzare distribuzioni non informative, con la sola limitazione di rispettare vincoli sul supporto, qualora i parametri siano definiti in supporti limitati.

- La distribuzione a priori per $\gamma_{0h} \in \mathbb{R}$, con $h = 1,\dots,H$, è una $\mathcal{N}(\mu = 0, \sigma =$

100); questa, essendo centrata in 0, non assume un valore non nullo come più plausibile, ed inoltre, grazie alla varianza elevata, è sparsa.

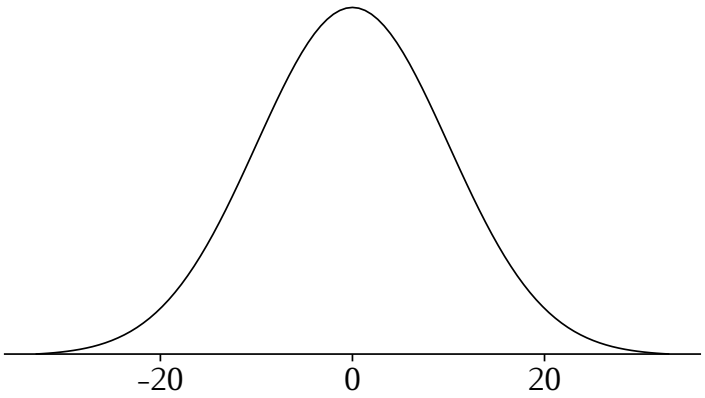


Figura 2.4: Grafico distribuzione a priori per γ_{0h} e $\boldsymbol{\gamma}_h$.

- Per quanto riguarda invece $\boldsymbol{\gamma}_h \in \mathbb{R}^{13}$, con $h = 1,\dots,H$, si è proceduto per ciascuna delle $H \cdot 13$ distribuzioni da specificare, esattamente come già visto per γ_{0h}.

- Relativamente alla distribuzione a priori per $\xi_{0h}^{(j)}$ e per $\boldsymbol{\xi}_h^{(j)} \in \mathbb{R}^{13}$ (prendendo un elemento alla volta), con $h = 1,\dots,H$ e $j = 1,\dots,J$, si è scelta una Normale standard. In questo caso si è preferito procedere con una distribuzione informativa, non tanto perché si avevano conoscenze pregresse sulla distribuzione dei suddetti parametri, quanto piuttosto per ovviare ad alcuni problemi di convergenza che si sono riscontrati nelle catene simulate a posteriori. I dati a disposizione sono comunque in quantità considerevole e, di conseguenza, come sarà possibile vedere nel prossimo capitolo, la distribuzione a posteriori è stata comunque trainata dalla verosimiglianza verso altri valori dei parametri.

- Le ultime priori da specificare sono quelle riguardanti i σ_h, ossia le deviazioni standard degli H *kernel*. In questo caso, si è scelta una distribuzione che avesse supporto positivo, ossia la Inverse-Gamma con parametro di forma $\alpha = 13$ e scala $\beta = 9$, $IG(\alpha = 13, \beta = 9)$. Con la priori soggettiva scelta, si è proceduto assegnando massa di probabilità non

nulla ad elementi piccoli della varianza. In particolare, data la suddetta parametrizzazione si osserva per la distribuzione un valore medio pari a 0.75, una moda di 0.64 e, i valori $q_{0.05} = 0.46$, $q_{0.95} = 1.17$. La sopracitata specificazione è stata quella che ha garantito convergenza più rapida delle catene dei valori simulati a posteriori.

Per avere ben chiaro il modello implementato, in figura 2.5, se ne riporta la rappresentazione grafica.

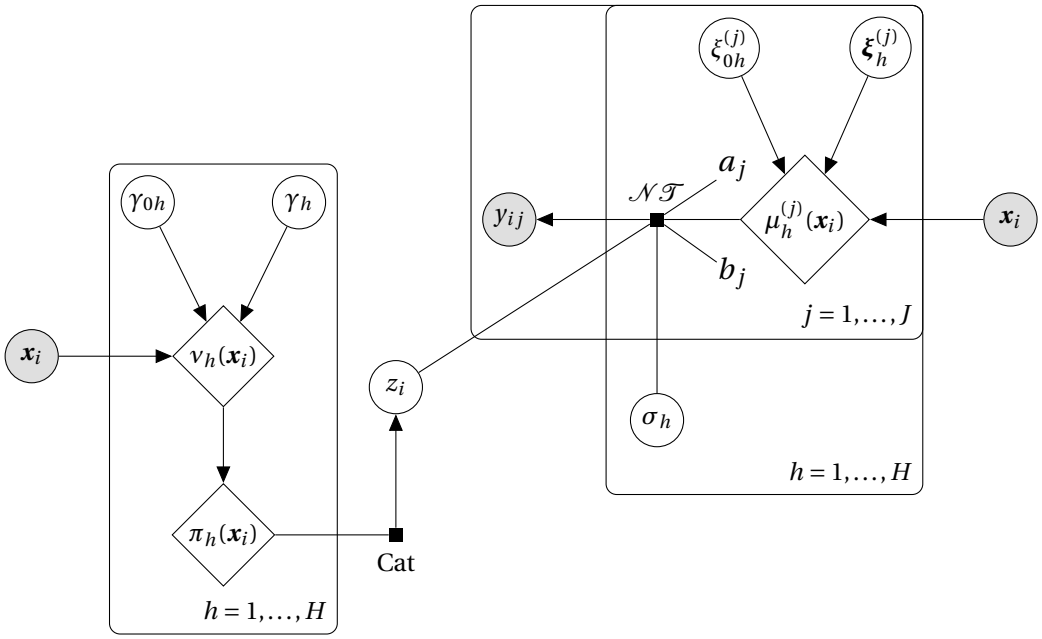


Figura 2.5: Rappresentazione grafica del meccanismo che genera i dati y_{ij} dal modello utilizzato.

2.2.4 Modellazione in NIMBLE

NIMBLE (de Valpine, Turek et al. 2017) è un sistema per costruire e condividere metodi di analisi per modelli statistici, specialmente per modelli gerarchici e metodi computazionalmente intensivi. NIMBLE è costruito in R, ma compila i modelli e gli algoritmi usando C++, garantendo così una maggior velocità. Esso include tre componenti:

1. Un sistema per usare modelli scritti nel linguaggio dei modelli BUGS, come oggetti programmabili in R;

2. Una libreria iniziale di algoritmi per modelli scritti in BUGS, includendo MCMC basico, che può essere usato direttamente o modificato a piacere con R prima di essere compilato e avviato;

3. Un linguaggio incorporato in R per programmare algoritmi per i modelli, entrambi compilati in codice C++e caricati in R.

NIMBLE può anche essere utilizzato come un modo per compilare del codice simile ad R in C++; questo sarà poi compilato e caricato in R con funzioni d'interfaccia od oggetti.

Una volta specificate tutte le distribuzioni a priori, si è altresì implementata la verosimiglianza. Poiché la distribuzione della Normale troncata, descritta in 2.17, non è presente nelle distribuzioni già esistenti in NIMBLE, si è dovuto procedere implementandola *ex-novo*.

Capitolo 3

Risultati della modellazione

Nel proseguo, vengono riportati e commentati i grafici relativi alle distribuzioni a posteriori per i parametri ritenuti più informativi. La simulazione, avvenuta in un PC fisso provvisto di una CPU hexa-core AMD Ryzen 5 1600 e 2 banchi da 8 GB di RAM DDR4 da 3200 MHz, ha impiegato circa 6 giorni ad essere completata. Nella parte finale del presente capitolo sono riportati i risultati dello studio, coadiuvati anche da opportuni grafici.

3.1 Simulazione a posteriori

Nella presente sezione viene presentata una disamina delle catene simulate a posteriori per i diversi parametri del modello. Innanzitutto occorre notare come, quando si utilizzano misture non parametriche, si è soliti porre un numero di cluster H molto elevato. Ciò a cui si auspica infatti, attraverso un approccio non parametrico, è il lasciare che siano i dati a "parlare", senza che venga imposto a priori un determinato numero H di cluster. Saranno quindi le unità statistiche, con le relative risposte ed indicatori socio-demografici, a guidare il modello verso un'opportuna scelta del numero di cluster.

Alcuni studi, seppur diversi dal presente lavoro e che prendono in esame dati analoghi ai presenti, hanno sottolineato come vi siano poche sottopopolazioni nel campione considerato (Aliverti e Russo 2022). Si è pertanto deciso di porre $H = 6$, benché solitamente si tenda ad impostarlo più elevato.

Con la scelta effettuata è stato pertanto possibile troncare il processo di Dirichlet: questo infatti, a livello teorico, è definito per $H \to \infty$, e pertanto non implementabile su un calcolatore.
A certificare la validità della scelta fatta, si sono osservate le percentuali di riempimento a posteriori di ciascun gruppo trovato, e ciò che ne è risultato è che la maggioranza di unità statistiche viene allocata nei primi 3 gruppi, mostrando invece percentuali esigue di popolamento dal quarto cluster in poi.

In relazione alla lunghezza delle catene, si è scelto di produrle da 20000 iterazioni; da queste si è proceduto con l'eliminazione di 500 valori (*burn-in*), per poi cercare di ridurre l'autocorrelazione dei valori generati attraverso un *thinning* di 10. Le catene finali ottenute per ciascun parametro hanno quindi lunghezza pari a 1950: come si vedrà da alcuni grafici sembra che sia stata raggiunta la convergenza delle distribuzioni a posteriori per i parametri.

3.1.1 Campionatori e catene

Una volta scelte le distribuzioni a priori per i parametri del modello, ed implementata la verosimiglianza, NIMBLE si è occupato della costruzione del modello.
Si sono utilizzati 2 diversi campionatori

1. *Categorical sampler*: specifico per le variabili a supporto discreto, in questo caso solamente z_i, ossia l'indicatore del cluster h per il soggetto i-esimo.

2. *Random Walk sampler*: la simulazione è avvenuta tramite passeggiata casuale per tutti i restanti parametri contenuti in $\boldsymbol{\theta}$ (2.21);

Prima di passare all'interpretazione dei risultati ottenuti, occorre procedere con una disamina delle catene ottenute.

Catene e diagnostiche di convergenza

La prima diagnostica di convergenza utilizzata è di tipo grafico; questa, andando a studiare i grafici delle traiettorie per ogni componente, deve verificare

che i moti oscillino molto velocemente senza mostrare particolari andamenti ricorrenti. In questi casi si parla di buon *mixing*. Un ulteriore tecnica è quella che va a disegnare nei grafici la media cumulata. Quest'ultimo approccio aiuta a capire se la catena è arrivata o meno a convergenza. Infine, cercando di capire quante osservazioni di quelle generate possano effettivamente considerarsi provenienti da un campione $i.i.d.$, è stata calcolata l'*Effective-Sample-Size* di ciascuna catena.
Per snellire la presente trattazione, tutti i grafici delle catene esplorate sono stati riportati in appendice B. In ogni caso, basti sapere che si è ottenuto un buon *mixing* per tutte le catene scandagliate, inoltre, la media cumulata dei parametri ad ogni iterazione non pone dubbi sulla convergenza delle catene ottenute. Per quanto riguarda l'*Effective-Sample-Size* di ciascuna catena, occorre far notare come questa raggiungesse almeno un valore di 1800 su 1950 osservazioni facenti parte di ciascuna traiettoria, e pertanto ci si è ritenuti soddisfatti delle effettive dimensioni dei campioni $i.i.d.$ ottenute.

3.2 Risultati a posteriori

3.2.1 Indicatore del cluster

Vista l'importanza di allocare correttamente un soggetto i-esimo in uno degli H cluster, e considerando il supporto discreto delle z_i, la stima puntuale, a posteriori, del cluster di appartenenza di ciascun soggetto non è un problema banale. A tal proposito si è pertanto deciso di utilizzare, come metrica atta a misurare la distanza tra 2 raggruppamenti, la Variazione di Informazione. Si considerano 2 partizioni X e Y di un insieme A in sottoinsiemi disgiunti $X = \{X_1, \dots, X_k\}$, $Y = \{Y_1, \dots, Y_l\}$, e siano inoltre $n = \sum_i |X_i| = \sum_j |Y_j| = |A|$, $p_i = |X_i|/n$, $q_j = |Y_j|/n$ ed $r_{ij} = |X_i \cap Y_j|/n$ allora, la Variazione di Informazione, è così esplicitata

$$VI(X;Y) = -\sum_{i,j} r_{ij}\left(\log\frac{r_{ij}}{p_i} + \log\frac{r_{ij}}{q_j}\right). \tag{3.1}$$

Poiché il calcolo della VI attesa a posteriori è computazionalmente intensivo si è proceduto con una sua versione modificata, ossia quella che minimizza la VI invertendo il logaritmo ed il valore atteso. Il procedimento è risultato co-

munque molto oneroso.
Per maggiori informazioni sulla teoria sottostante e relativamente al pacchetto in R utilizzato, `mcclust.ext`, si veda Wade e Ghahramani (2018) e, per un ulteriore approfondimento sul tema del confronto di clustering, Meilă (2003).

Una volta ottenuto l'indicatore del cluster per ciascun soggetto, si è quindi proceduto con la caratterizzazione di ciascuno dei raggruppamenti trovati.
Si riportano ora dei gruppi di tabelle, ciascuno dei quali concernente uno dei tre cluster presi in considerazione. La scelta di interpretare solamente i primi tre raggruppamenti si esplicita in una caratterizzazione più agevole delle sotto-popolazioni; inoltre, in relazione a quanto già discusso sull'occupazione di ciascun gruppo, la stick-breaking-prior ha allocato sempre meno soggetti nei cluster all'aumentare di H.

Una generica cella di queste tabelle contiene la probabilità, espressa in percentuale, che, data l'appartenenza del soggetto i-esimo al cluster h-esimo, esso abbia risposto con modalità c_j tra le d_j modalità possibili (indicate sul fondo) della domanda j-esima (indicata sulla sinistra), con $j = 1, \dots, J = 9$.

Primo cluster

Prendendo visione della figura 3.1 e 3.2 è possibile interpretare il primo cluster. In particolare, dalle prime 4 domande, ossia quelle relative alle normative di salute, non risulta chiara una classe predominante: i soggetti tendono a distribuirsi in modo piuttosto uniforme nelle 5 modalità possibili. Un discorso diverso lo si ha invece per le ultime domande, ossia quelle relative alla gestione dell'emergenza da parte del Governo, sulla pericolosità del Covid ed infine sul fidarsi, anche relativamente alla protezione offerta, oltre che all'obbligatorietà o meno, del vaccino. Si ha infatti che il 79.3% dei soggetti pensa che il Governo abbia gestito male l'emergenza, il 61.2% non pensa affatto che il Covid sia pericoloso, e in larga parte non ripone fiducia nel vaccino, né crede che possa proteggerlo dal contagio; infine, un'alta percentuale di soggetti, pari al 90.9%, non si trova d'accordo con l'obbligatorietà dei vaccini.
Cercando quindi di raffigurare questo profilo, lo si potrebbe individuare come soggetti molto insofferenti per le normative imposte dal governo, non impauriti dal Covid, e pertanto neanche interessati al vaccino, né alla sua ob-

bligatorietà, ma che in larga parte continuano a rispettare le normative di salute imposte. In sostanza, questa sotto-popolazione può prendere il nome di "Soggetti insofferenti non impauriti dal Covid e No-vax".

Secondo cluster

Considerando la figura 3.3 e 3.4, si può identificare il secondo cluster. Ciò che si nota immediatamente riguarda le prime 4 domande: la maggioranza dei soggetti ha risposto con la modalità "Sempre", ad indicare come questo raggruppamento comprenda i soggetti che rispettano scrupolosamente le norme sanitarie imposte. In ogni caso, la maggior parte di questi individui è molto insoddisfatta dell'operato del governo, con ben il 57.7% che ha infatti risposto con "Molto male" alla domanda relativa all'approvazione dell'operato del Governo. Segue poi una grossa fetta di popolazione che non ritiene che il virus sia pericoloso, e tende a posizionarsi nella modalità neutra 4. La maggioranza di questi individui comunque crede che il vaccino possa offrire protezione dal Covid, fidandosi in gran parte del vaccino (il 79.9% ha risposto con la modalità "Moderatamente"). Infine, per quanto riguarda l'obbligatorietà, l'88.6% ha risposto che è giusto che il vaccino sia obbligatorio.

Riassumendo quindi quanto detto, ed interrogandoci sul perché in questo cluster la maggior parte dei soggetti non sia contenta della gestione da parte del Governo, si potrebbe pensare che, benché tutte le norme sanitarie vengano rispettate rigorosamente, e che ci siano molti più soggetti che sono a favore del vaccino anziché contro, pare che questi o ritengano che le norme imposte non siano sufficientemente aspre, e che quindi auspichino in manovre più severe, oppure che, dopo aver rispettato alla lettera quanto imposto dal Governo, sia sorto in essi del malumore. Cercando dunque di affibbiare un nome a questo gruppo, lo si potrebbe definire come i "Pro-vax scontenti del Governo".

Terzo cluster

Per l'interpretazione del terzo cluster si possono osservare le tabelle in figura 3.5 e 3.6. Prendendo atto delle percentuali riportate per la modalità "Sempre" nelle prime quattro domande, come nel caso precedente, la maggioranza dei rispondenti mostra un livello di ottemperanza molto elevato alle normati-

ve anti-Covid. I soggetti sono inoltre contenti dell'operato del Governo (modalità "Bene" assunta per il 42.9% del totale). Per quanto riguarda la pericolosità percepita del Covid, non vi è una chiara modalità sovrastante le altre, piuttosto sembrano essere presenti opinioni molto discordanti vista anche la distribuzione piuttosto uniforme delle modalità. Ciò che invece non pone dubbi è il pensiero che il vaccino non protegga affatto dal Covid: le modalità negative sono assunte in percentuale schiacciante se raffrontate con quelle positive. Tutto ciò trova poi conferma anche nella fiducia riposta nei vaccini: la modalità "Poco" risulta la più frequente, con una percentuale osservata pari al 41.3%. Infine, relativamente all'obbligatorietà del vaccino, le modalità no e sì sono assunte pressoché in egual misura: non è stato pertanto possibile discernere un pensiero di fondo maggiormente condiviso.
Cercando quindi di delineare tratti comuni dei rispondenti appartenenti alla terza sotto-popolazione, si può evidenziare come questi siano molto scrupolosi nel seguire le normative, siano contenti dell'operato del Governo, ma d'altro canto non ripongono fiducia nei vaccini, né sulla loro efficacia. Questo gruppo viene quindi definito come i "No-vax contenti del Governo".

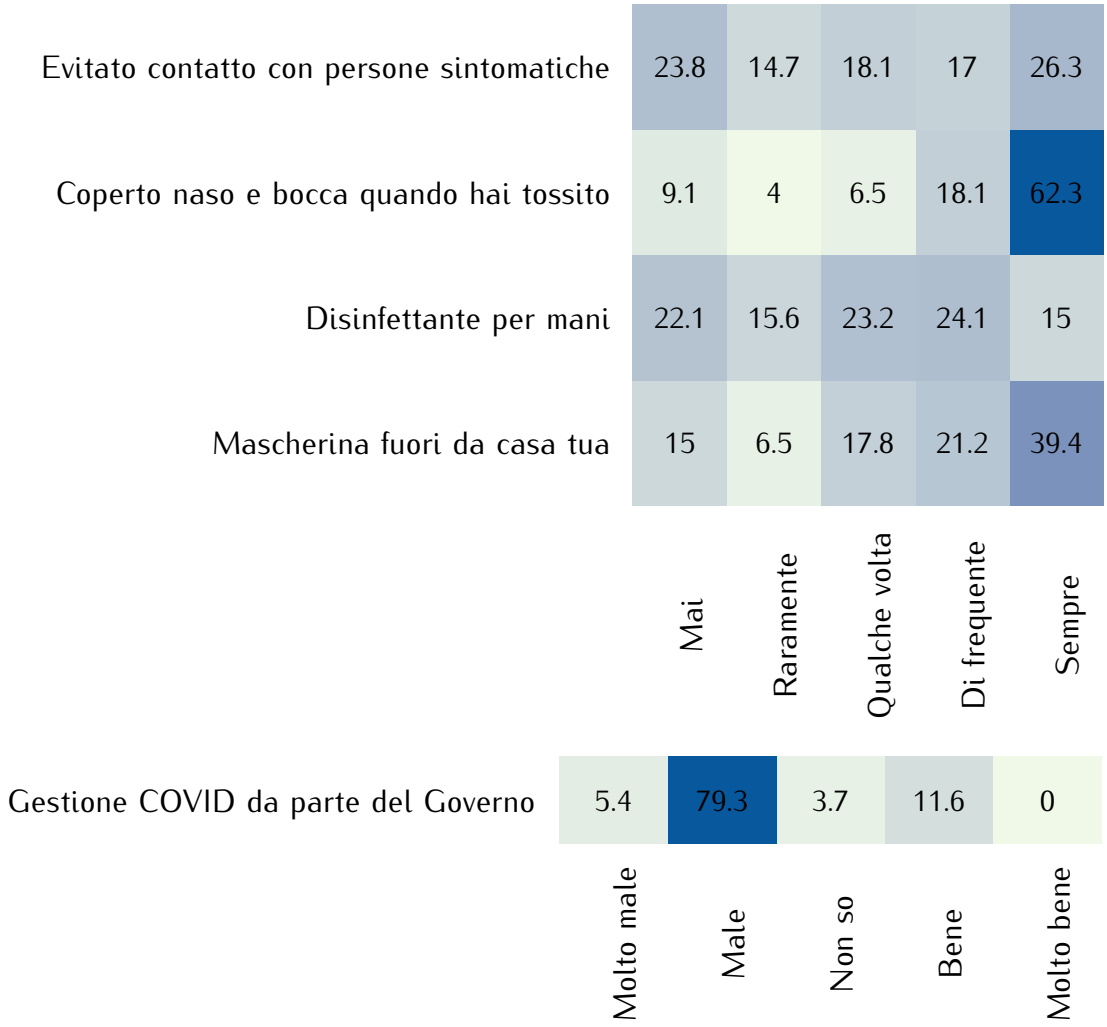


Figura 3.1: Percentuali di risposta alle modalità delle prime 5 domande, cluster 1.

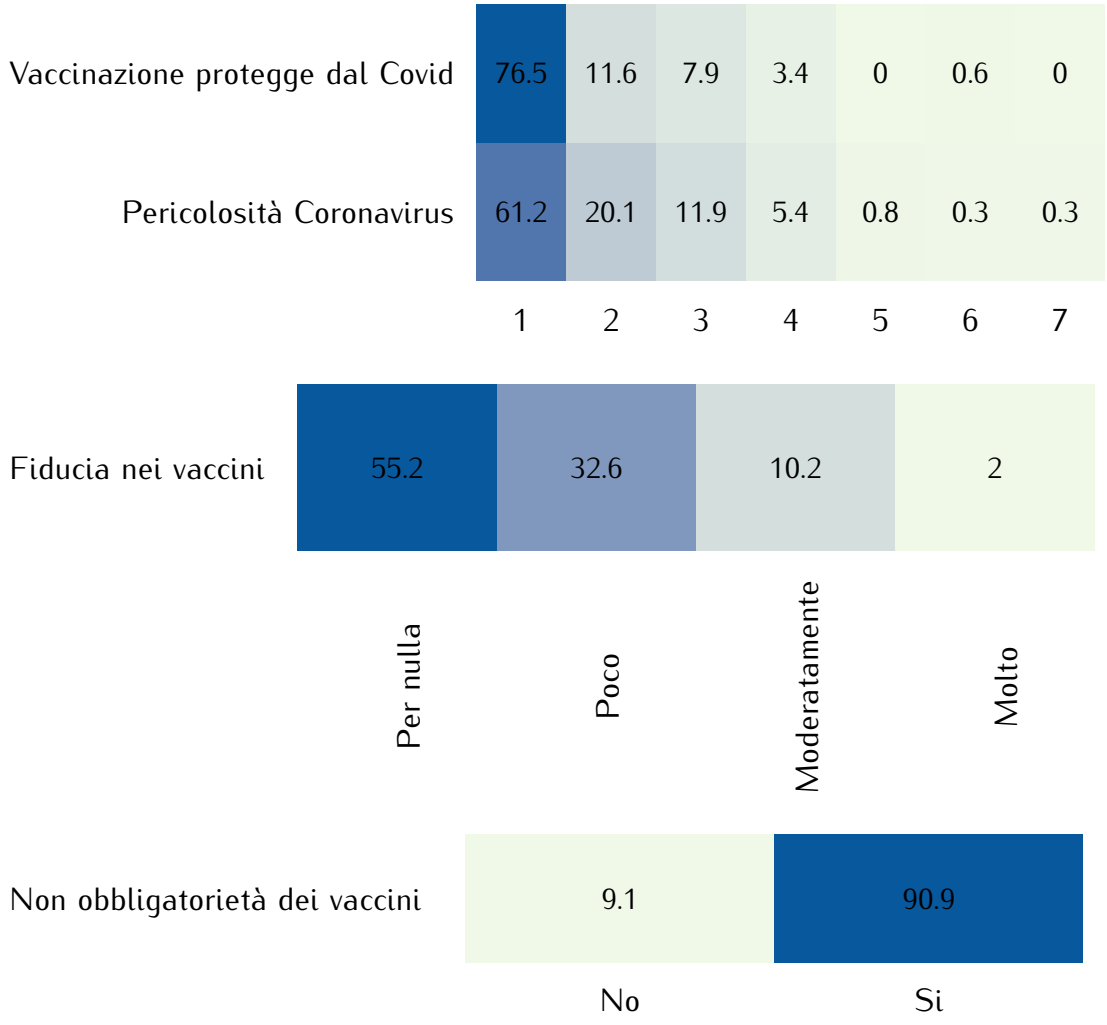


Figura 3.2: Percentuali di risposta alle modalità delle ultime 4 domande, cluster 1.

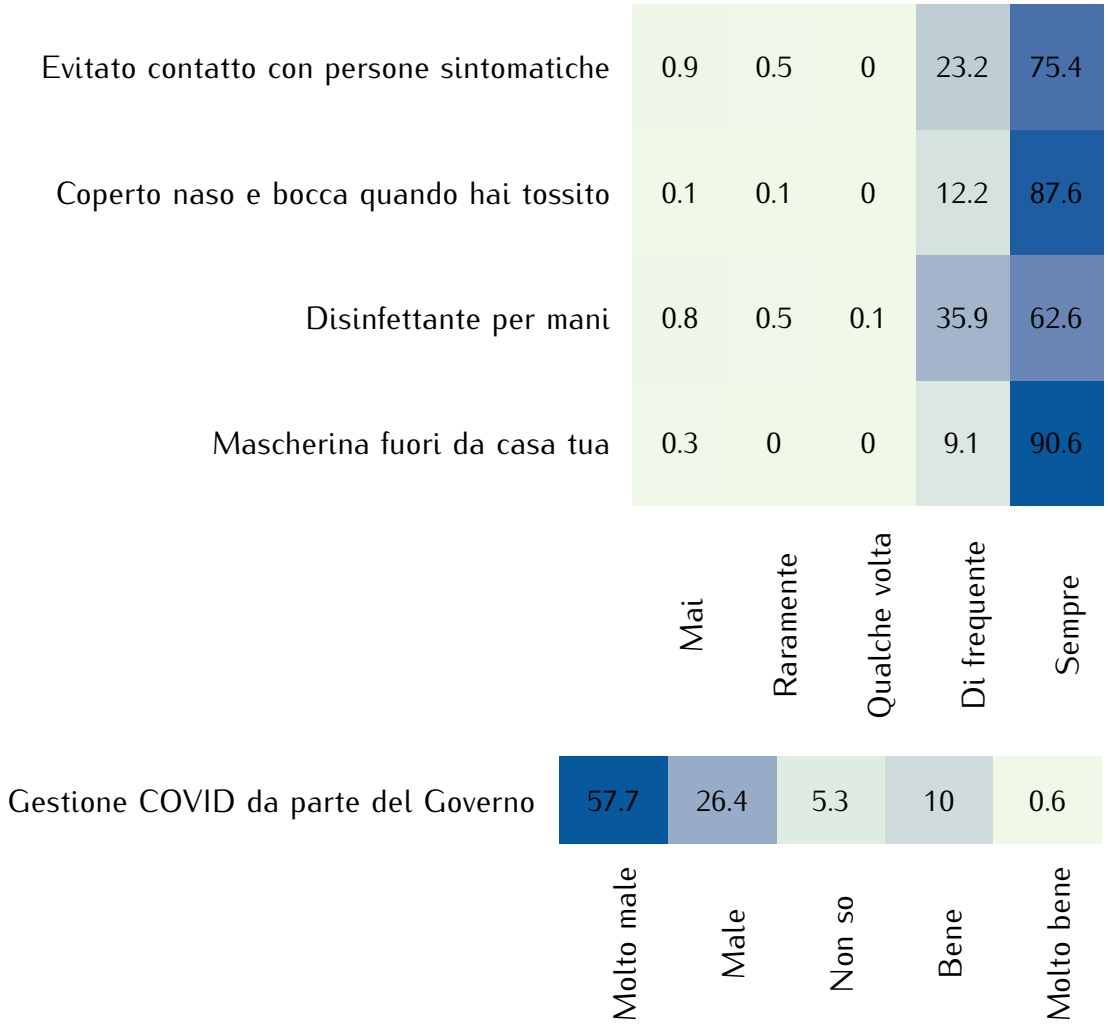


Figura 3.3: Percentuali di risposta alle modalità delle prime 5 domande, cluster 2.

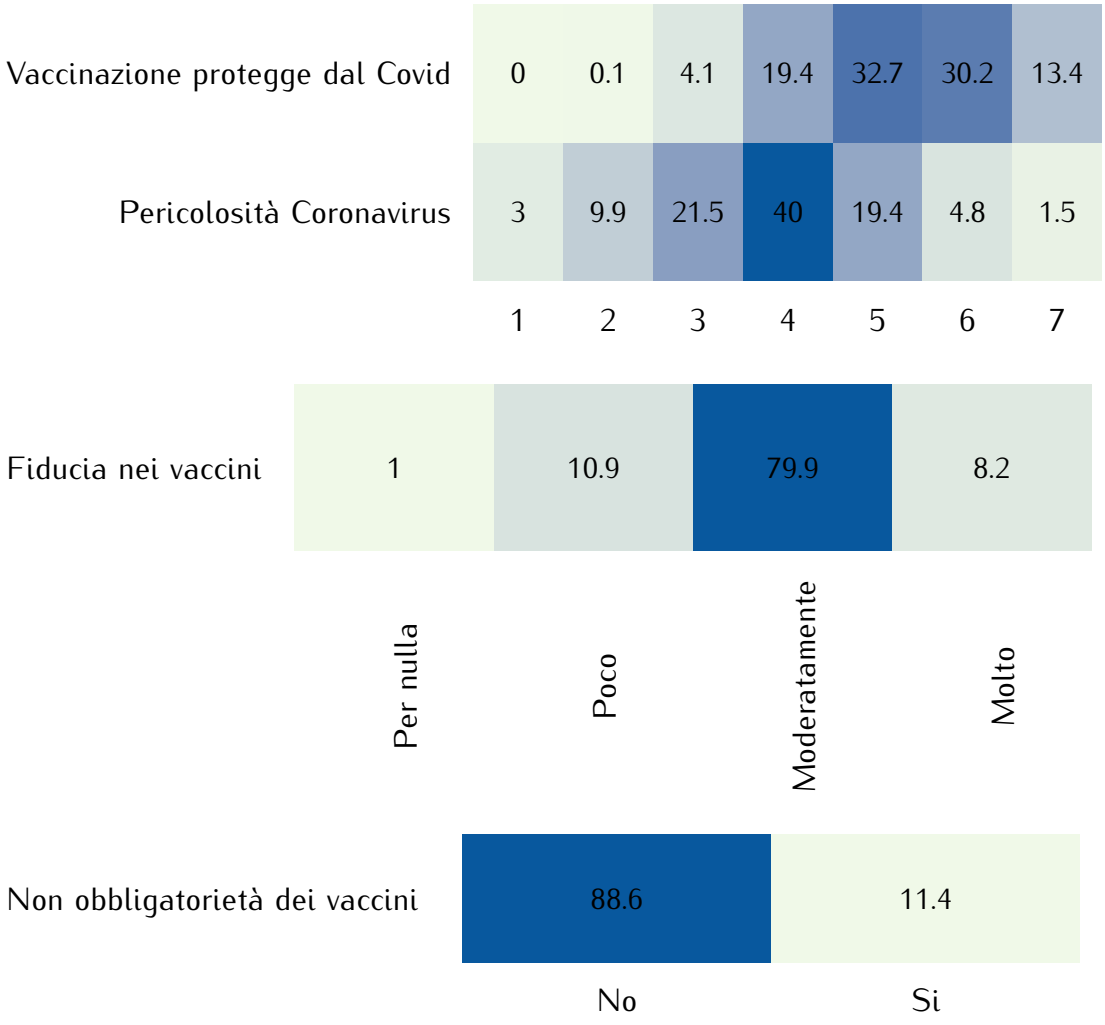


Figura 3.4: Percentuali di risposta alle modalità delle ultime 4 domande, cluster 2.

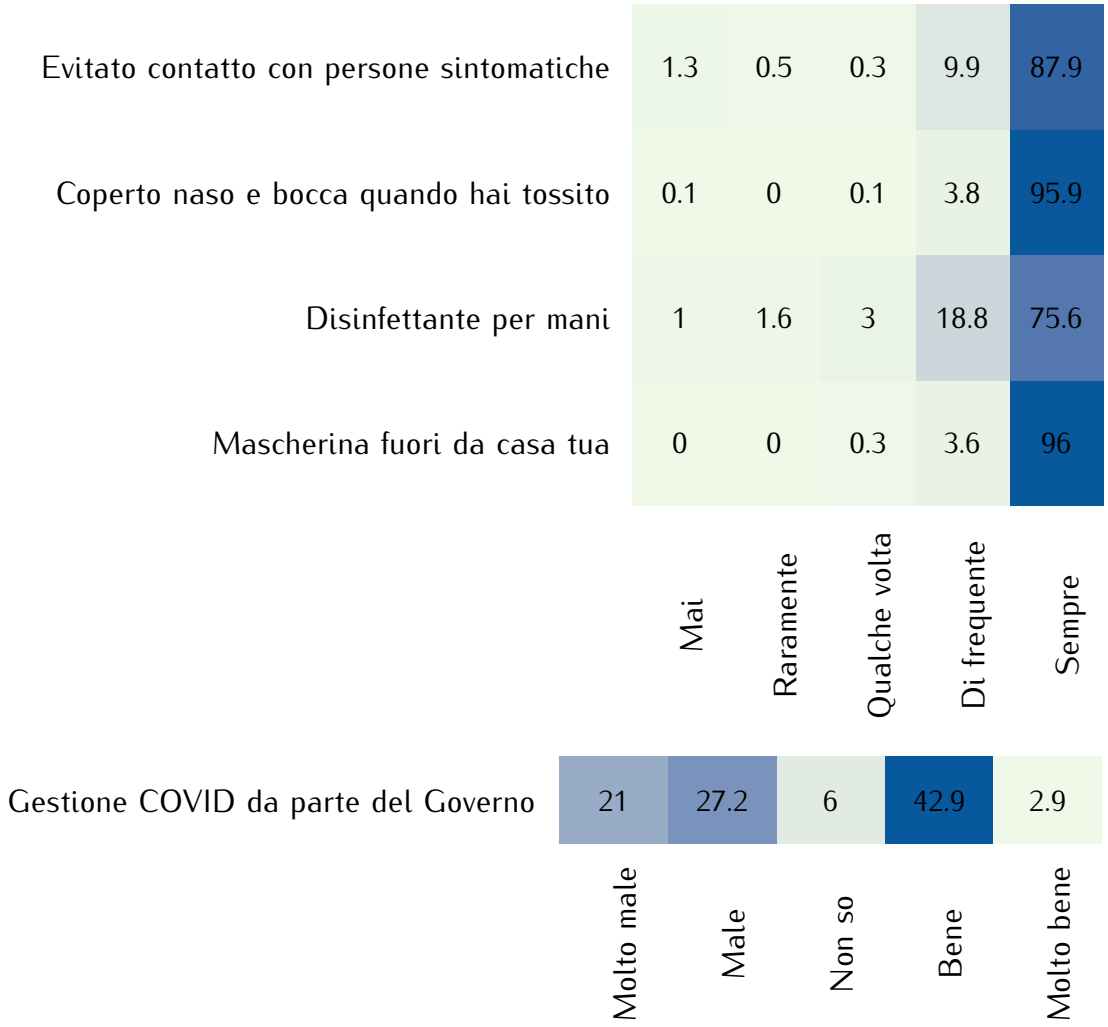


Figura 3.5: Percentuali di risposta alle modalità delle prime 5 domande, cluster 3.

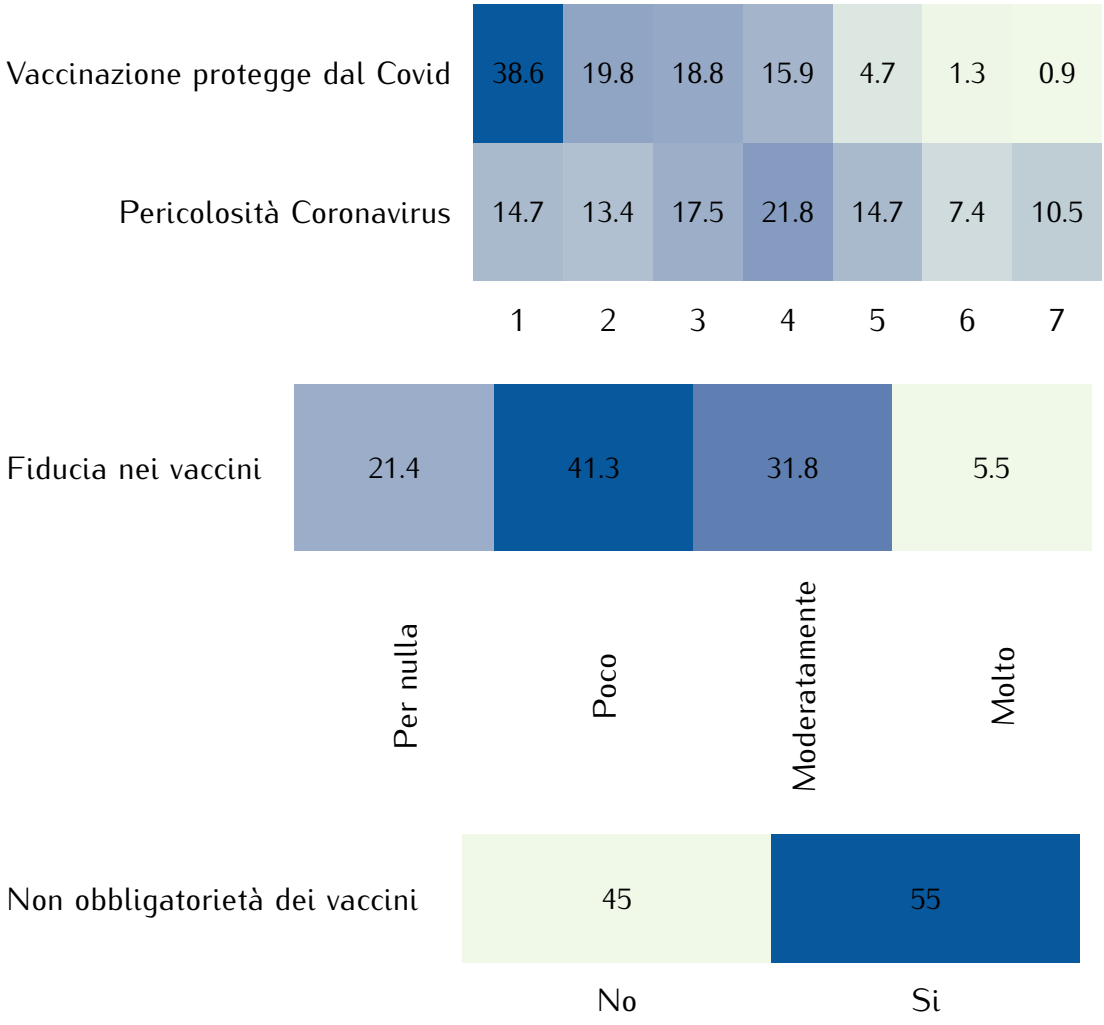


Figura 3.6: Percentuali di risposta alle modalità delle ultime 4 domande, cluster 3.

3.2.2 Caratterizzazione socio-demografica di ciascun cluster

Successivamente alla caratterizzazione di ciascun cluster, basata su tutte le percentuali poco fa descritte, si è cercato di capire quali indicatori socio-demografici contribuissero a determinare l'appartenenza di un soggetto ad uno specifico cluster. Per poter dare un interpretazione a ciascuno dei coefficienti, occorre prima rimaneggiare la definizione della logit-stick-breaking-prior vista in 2.15. In particolare, ricordando

$$\Pr(z_i = h|\boldsymbol{x}_i) = \nu_h(\boldsymbol{x}_i) \prod_{l=1}^{h-1} (1 - \nu_l(\boldsymbol{x}_i)), \tag{3.2}$$

come la probabilità che l'i-esimo soggetto appartenga al cluster h-esimo, si può notare che

$$\nu_h(\boldsymbol{x}_i) = \frac{\pi_h(\boldsymbol{x}_i)}{1 - \sum_{l=1}^{h-1} \pi_l(\boldsymbol{x}_i)} = \frac{\Pr(z_i = h|\boldsymbol{x}_i)}{\Pr(z_i > h-1|\boldsymbol{x}_i)}, \quad h \in \mathbb{N}, \tag{3.3}$$

garantendo quindi un'interpretazione avveduta. Più nel dettaglio, facendo riferimento a quanto descritto in Rigon e Durante (2021), si interpretano i coefficienti γ_{0h} e γ_{hp} definenti il generico $\nu_h(\boldsymbol{x}_i)$, con $h = 1,\dots,H$ e $p = 1,\dots,P$ a seconda della loro positività o negatività, rispettivamente come aumento o diminuzione della probabilità di essere assegnati al gruppo h-esimo condizionatamente all'essere "sopravvissuti" all'allocazione negli $h-1$ gruppi precedenti. Relativamente ai coefficienti γ_{1p}, con $p = 1,\dots,P$ questi possono invece essere interpretati in base al loro segno, come contributi all'aumento o alla diminuzione della probabilità di venire allocati al primo cluster. La probabilità complementare a quella appena descritta definisce invece quanto verosimile sia l'allocazione ai successivi $H-1$ cluster rimanenti.
La figura 3.7, 3.8 e 3.9 aiutano a visualizzare il segno di ciascun coefficiente e la loro variabilità. I commenti fatti, per i parametri delle covariate categoriali, sono da considerarsi relativamente alla modalità base di ciascuna variabile qualitativa cui fanno riferimento, nell'ordine: genere femminile, Centro e lavoratore a tempo pieno.

Primo cluster

Con riferimento al primo raggruppamento trovato, ossia quello in cui vi sono allocati più soggetti, l'interpretazione dei coefficienti è più agevole se raffrontata ai successivi cluster. In particolare, vista la positività dei coefficienti relativi all'età, al genere maschile, all'abitare nelle Isole o nel Sud-Italia, e l'essere studente a tempo pieno o in pensione, avere tali tratti socio-demografici contribuirà a rendere più probabile l'appartenenza al cluster dei "Soggetti insofferenti non impauriti dal Covid e No-vax". Tuttavia, si può notare come le distribuzioni a posteriori per i coefficienti relativi alle modalità "Isole", "Nord-Ovest" e "Disoccupato", comprendano lo 0, indicando quindi una possibile non significatività di questi.

Secondo cluster

Vista la positività del coefficiente Età, all'aumentare di questa aumenterà la probabilità di appartenere al secondo cluster, condizionatamente al fatto di non essere stato allocato al primo. Un analogo discorso lo si ha per i coefficienti relativi al genere femminile ed alle modalità "Isole", "Nord-Ovest" e "Sud" rispetto alla modalità base "Centro" della variabile `Region` e con le modalità "Non lavora" e "Part-time" rispetto alla baseline "Tempo pieno" della variabile `Employment_status`. Tra i coefficienti esaminati si ipotizza comunque la non significatività di "Part-time", considerata la presenza del valore 0 nella scatola del relativo boxplot. D'altro canto, avere età inferiore, genere maschile, provenire dal Nord-Est d'Italia, essere studente a tempo pieno, in pensione, o disoccupato contribuisce invece a diminuire la probabilità di appartenere al gruppo dei "Pro-vax scontenti del Governo", relativamente al non essere stato allocato nel primo gruppo.

Terzo cluster

Guardando alla variabile Età si vede che, anche per questo cluster, in corrispondenza di un suo aumento, la probabilità di appartenere al terzo cluster, condizionatamente al non appartenere ai primi due raggruppamenti, aumenta. Essere studente a tempo pieno, di genere maschile, abitare nel Nord-Est o

nel Nord-Ovest Italia renderà meno verosimile l'assegnazione a questo gruppo, sempre con riferimento al non essere stato collocato nei primi 2 gruppi trovati. I commenti fatti devono però tenere in considerazione che le modalità "Nord-Ovest" e "Studente a tempo pieno" non sembrano essere significative. Il gruppo dei "No-vax contenti del Governo" sarà perciò più frequentemente composto da femmine abitanti nelle Isole oppure nel Sud-Italia, non lavoratrici o con contratto part-time, in pensione oppure disoccupate.

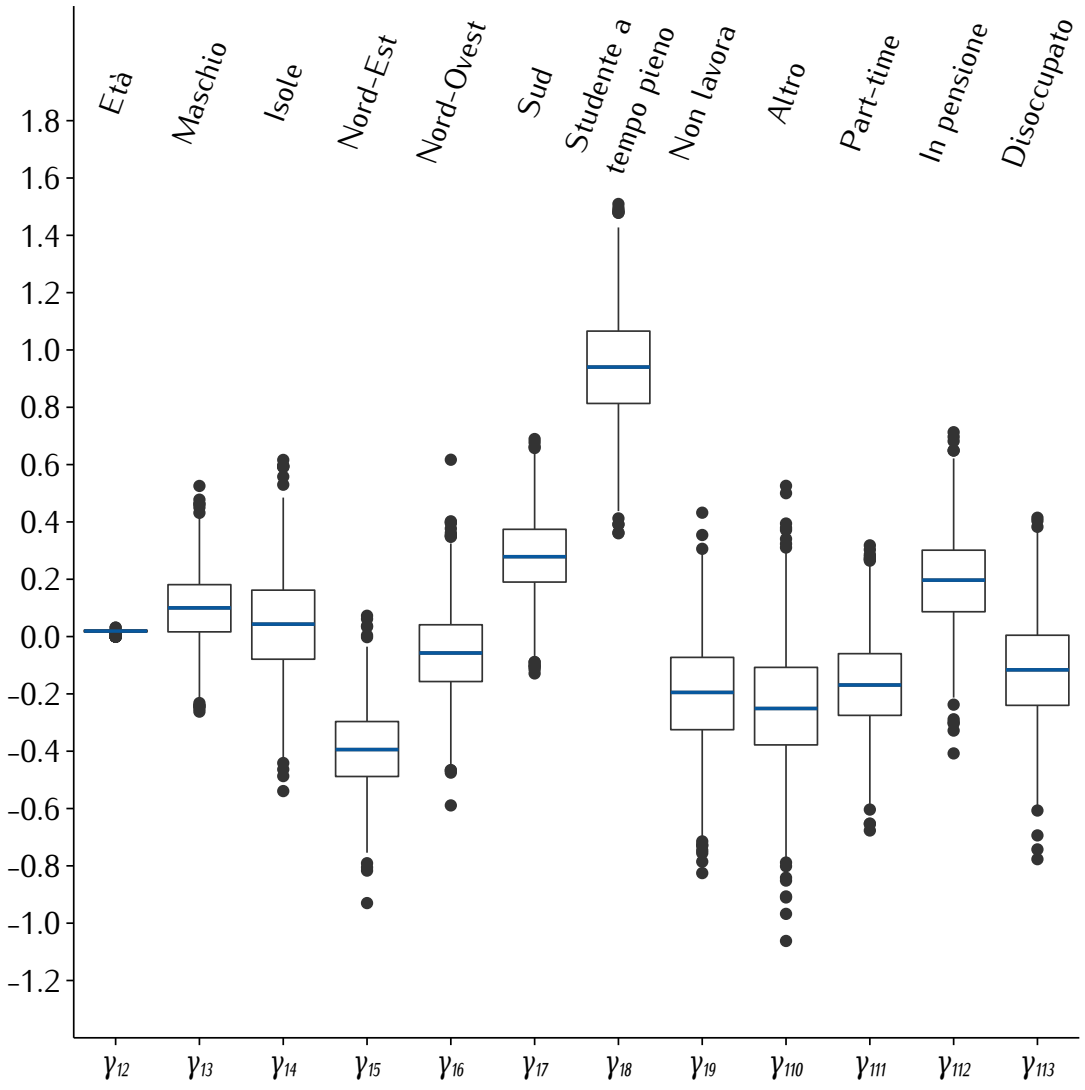


Figura 3.7: Boxplot coefficienti a posteriori relativi al primo cluster.

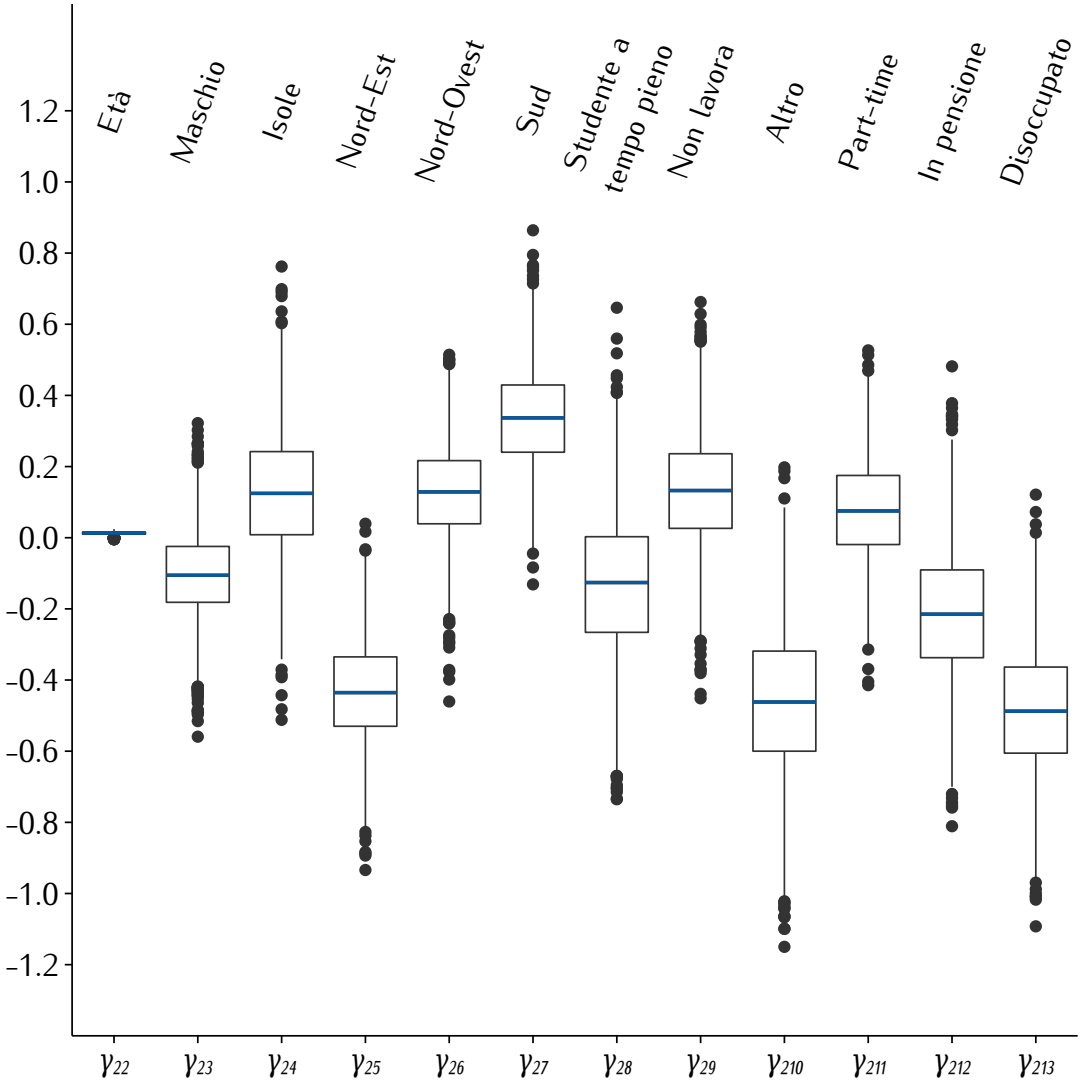


Figura 3.8: Boxplot coefficienti a posteriori relativi al secondo cluster.

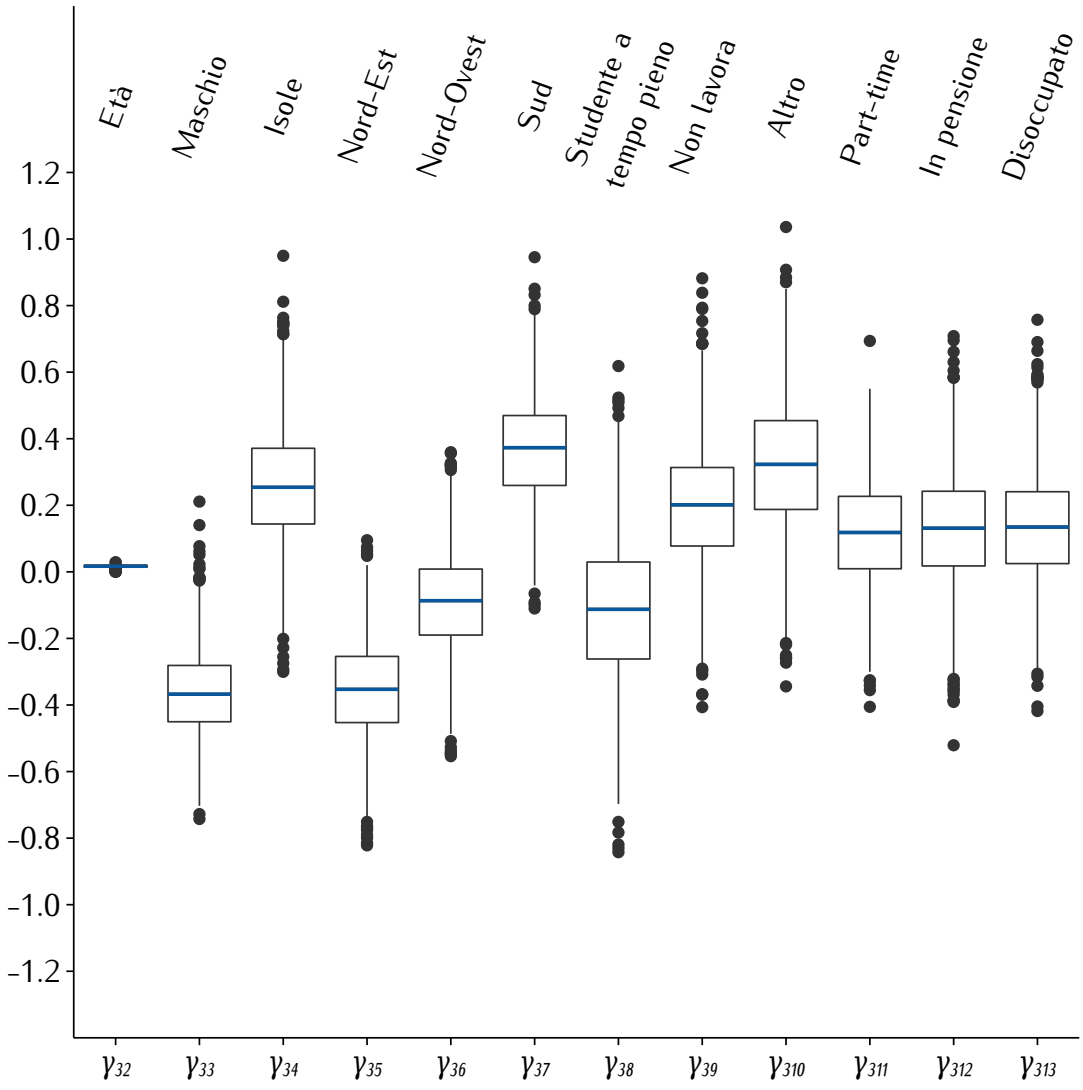


Figura 3.9: Boxplot coefficienti a posteriori relativi al terzo cluster.

3.2.3 Caratterizzazione della risposta media di un individuo in base al cluster di appartenenza e ad alcuni regressori

In quest'ultima parte si vuole analizzare in che modo, data l'appartenenza di un soggetto i ad un cluster h, gli indicatori socio demografici contribuiscano ad aumentare od abbassare la risposta media ad una determinata domanda. Per caratterizzare quanto descritto, si utilizzeranno i parametri $\boldsymbol{\xi}_h^{(j)}$. In ogni caso, vista l'elevata numerosità di questi ultimi, si è proceduto considerando solamente $h = 1,2,3$, ossia i cluster più popolati, con particolare attenzione verso le domande contenute in y_1, y_5, y_6 ed y_8, che si ricordano essere le seguenti

- y_1 corrisponde a "Hai indossato la mascherina fuori da casa tua";
- y_5 corrisponde a "Il Governo ha gestito bene la questione COVID";
- y_6 corrisponde a "Il Corona-virus è molto pericoloso";
- y_8 corrisponde a "Essere vaccinato ti protegge contro il COVID".

La scelta è ricaduta pertanto su quelle domande che si è pensato fossero più interessanti tra le $J = 9$ analizzate. Un ulteriore semplificazione adottata è stata quella di considerare solamente l'età, il genere e l'area italiana di provenienza dei soggetti, omettendo quindi la variabile categoriale più numerosa in merito a modalità: lo stato occupazionale. Tutti i risultati ottenuti per le variabili di natura categoriale sono da considerarsi rispetto alle relative modalità base: genere femminile e Centro Italia. Come si potrà notare, commenti relativi alla variabile Età non sono stati riportati: i boxplot dei coefficienti mostrano come lo 0 sia il valore assunto in mediana, con la scatola che evidenzia una bassissima variabilità di stima, rendendo pertanto difficile qualsiasi sbilanciamento in merito alla positività o negatività di tali parametri.

Primo cluster

Osservando la figura 3.10, è possibile dare un'interpretazione ai vettori di coefficienti $\boldsymbol{\xi}_1^{(j)}$, con $j = 1,5,6,8$. Si sta facendo riferimento al cluster dei "Soggetti insofferenti non impauriti dal Covid e No-vax".

- Relativamente alla domanda y_1, essere di genere maschile, abitare nelle Isole o del Nord-Est Italia contribuisce ad aumentare la risposta media alla domanda, e pertanto la mascherina verrà indossata più di frequente;

- Relativamente alla domanda y_5, essere di genere maschile o vivere nel Nord-Ovest Italia contribuisce ad aumentare la risposta media alla domanda, e pertanto l'opinione sul Governo tenderà ad essere più positiva;

- Relativamente alla domanda y_6, essere isolano contribuisce ad aumentare la risposta media alla domanda, e pertanto la pericolosità del Corona-virus sarà percepita più alta;

- Relativamente alla domanda y_8, non si ha invece una risposta netta: sembra che tutti i coefficienti considerati tendano a fare aumentare la risposta media alla domanda relativa alla protezione offerta dai vaccini.

Secondo cluster

Osservando la figura 3.11, è possibile dare un'interpretazione ai vettori di coefficienti $\boldsymbol{\xi}_2^{(j)}$, con $j = 1,5,6,8$. Si sta facendo riferimento al cluster dei "Provax scontenti del Governo".

- Relativamente alla domanda y_1, essere di genere maschile, o essere del Nord-Ovest Italia contribuisce a diminuire la risposta media alla domanda, e pertanto la mascherina verrà indossata meno di frequente;

- Relativamente alla domanda y_5, essere di genere maschile, abitare nel Nord-Est o Nord-Ovest Italia contribuisce ad aumentare la risposta media alla domanda, e pertanto l'opinione sul Governo tenderà ad essere più positiva;

- Relativamente alla domanda y_6, essere isolano, del Nord-Ovest o del Sud Italia contribuisce ad aumentare la risposta media alla domanda, e pertanto la pericolosità del Corona-virus sarà percepita più alta. Si noti anche come spicca il coefficiente relativo al Sud;

- Relativamente alla domanda y_8, sembra che, fatta eccezione per abitare nel Nord-Est Italia, tutti gli altri coefficienti tendano a diminuire la risposta media, e di conseguenza si avrà una minor fiducia nell'efficacia dei vaccini.

Terzo cluster

Osservando la figura 3.12, è possibile dare un'interpretazione ai vettori di coefficienti $\boldsymbol{\xi}_3^{(j)}$, con $j = 1,5,6,8$. Si sta facendo riferimento al cluster dei "No-vax contenti del Governo".

- Relativamente alla domanda y_1, essere isolano è ciò che contribuisce maggiormente ad incrementare la media della risposta, pertanto la mascherina verrà indossata più frequentemente. A questo coefficiente segue quello relativo al genere maschile, seppur questo contributo sia davvero molto piccolo;

- Relativamente alla domanda y_5, il contributo maggiore all'incremento della media della risposta proviene in primo luogo dal genere maschile e, in quantità minore, dall'abitare nel Nord-Ovest Italia: con queste due modalità quindi si avranno più verosimilmente giudizi positivi in merito all'operato del governo;

- Relativamente alla domanda y_6, ossia quella relativa alla pericolosità percepita del COVID, le modalità "Isole", "Nord-Ovest" e "Sud", con i loro coefficienti positivi, decretano come più plausibile che il Corona-virus crei timore;

- Relativamente alla domanda y_8, abitare nel Sud-Italia o nelle Isole contribuisce ad innalzare la fiducia riposta verso la protezione offerta dai vaccini.

Si noti comunque che, per diversi dei parametri esaminati nei raggruppamenti descritti, la distribuzione a posteriori contiene lo 0, facendone quindi presagire la non significatività.

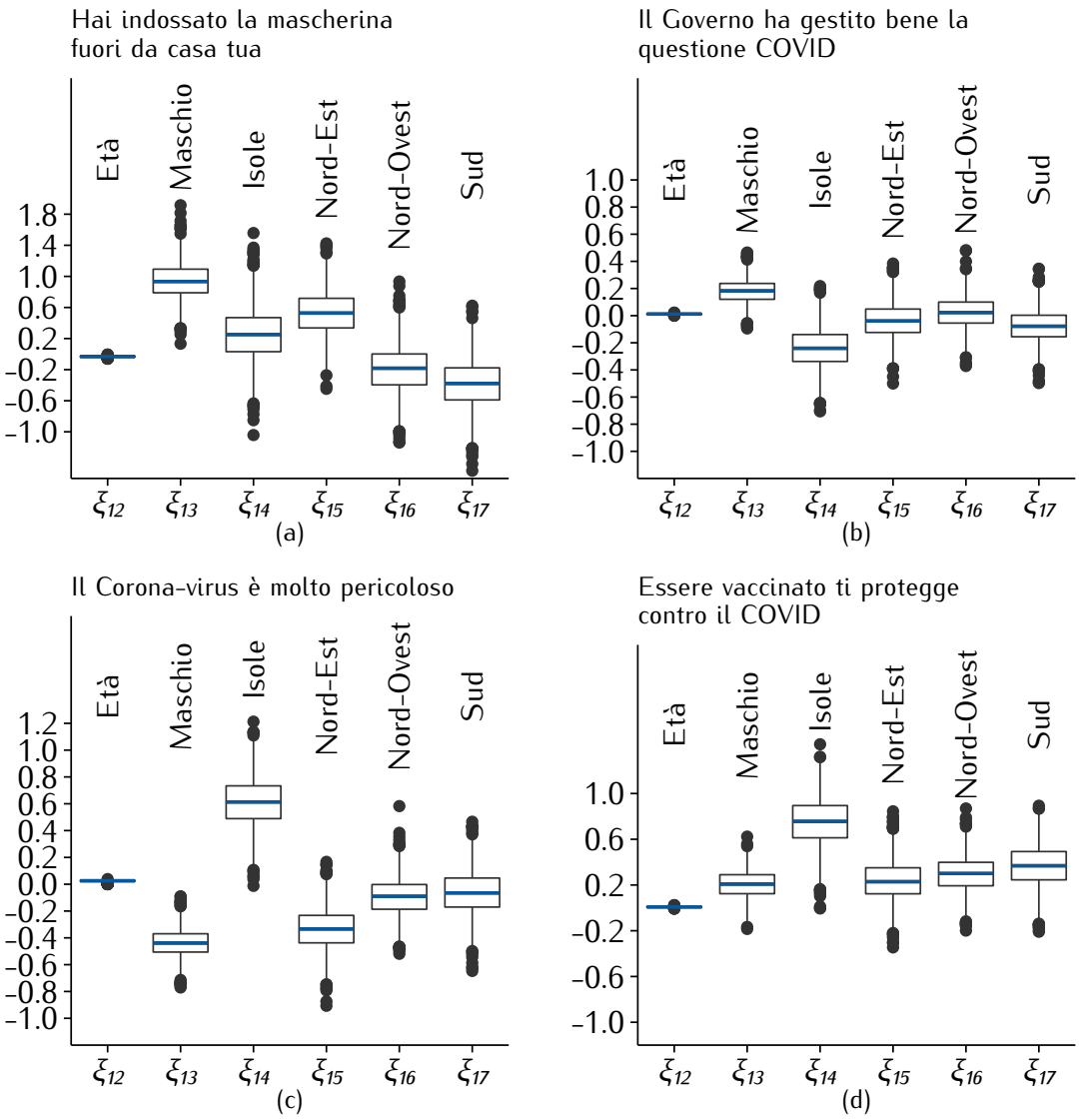


Figura 3.10: Boxplot $\boldsymbol{\xi}_1^{(j)}$ a posteriori: (a) $\boldsymbol{\xi}_1^{(1)}$, (b) $\boldsymbol{\xi}_1^{(5)}$, (c) $\boldsymbol{\xi}_1^{(6)}$, (d) $\boldsymbol{\xi}_1^{(8)}$.

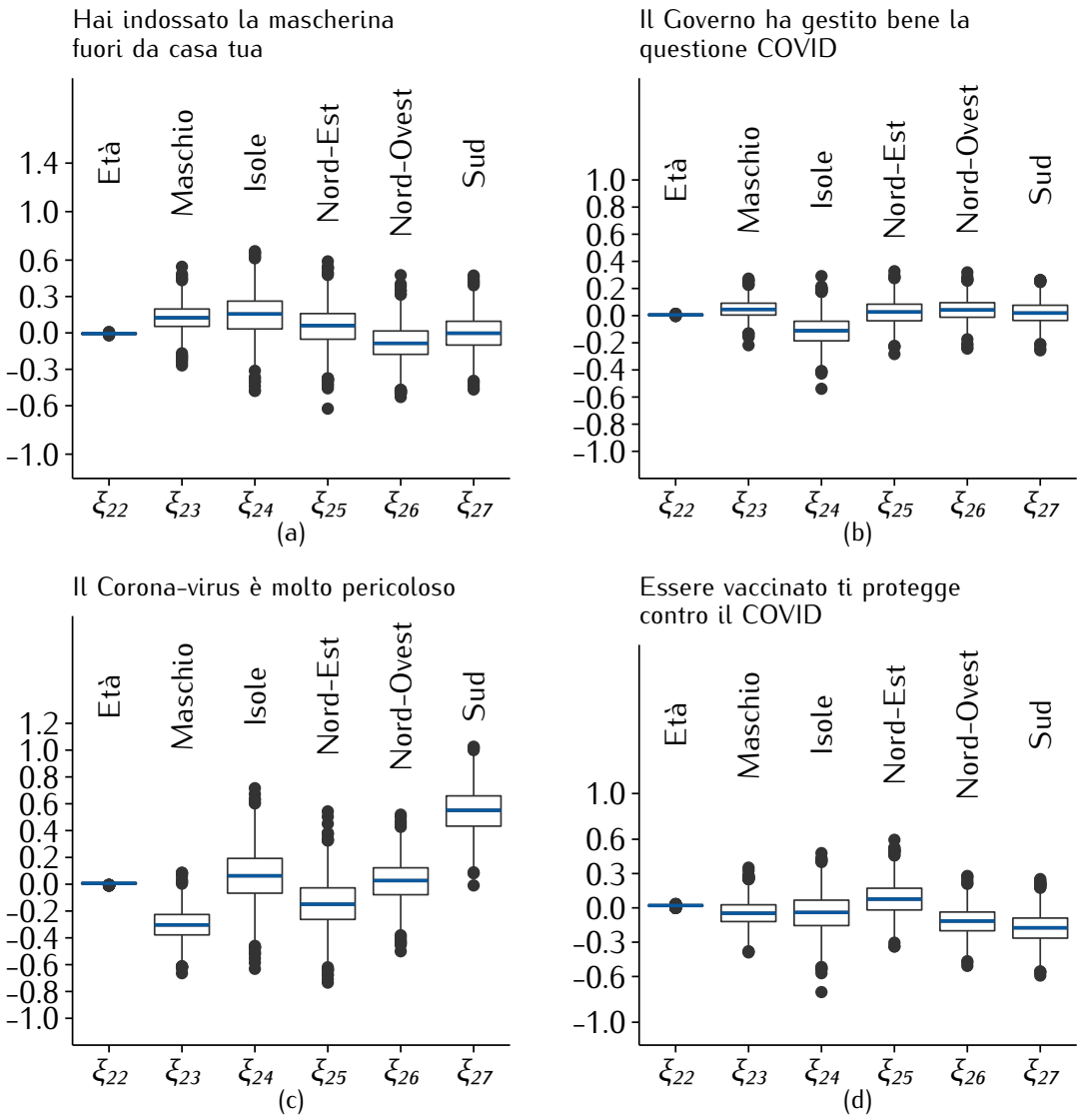


Figura 3.11: Boxplot $\boldsymbol{\xi}_2^{(j)}$ a posteriori: (a) $\boldsymbol{\xi}_2^{(1)}$, (b) $\boldsymbol{\xi}_2^{(5)}$, (c) $\boldsymbol{\xi}_2^{(6)}$, (d) $\boldsymbol{\xi}_2^{(8)}$.

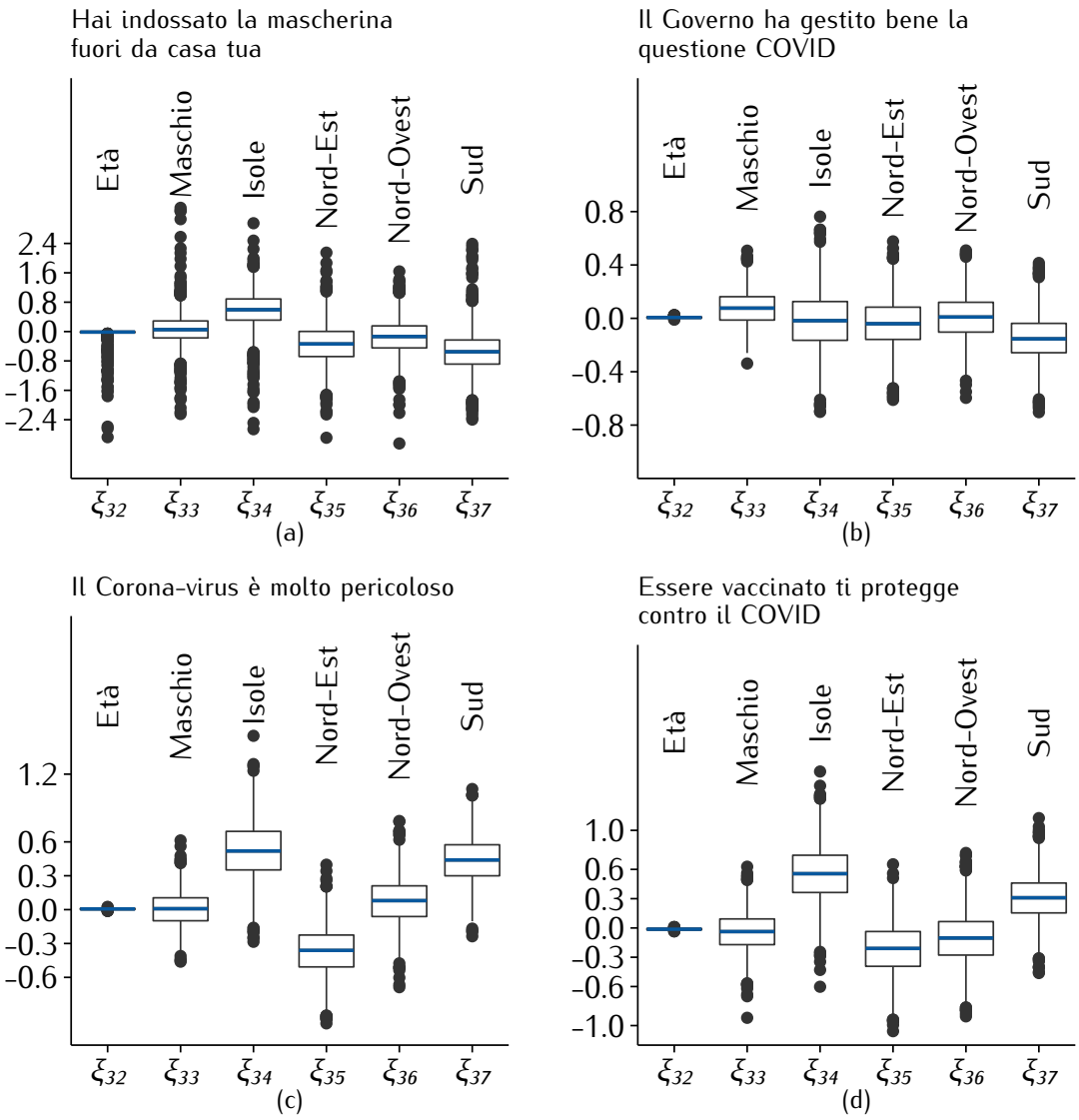


Figura 3.12: Boxplot $\boldsymbol{\xi}_3^{(j)}$ a posteriori: (a) $\boldsymbol{\xi}_3^{(1)}$, (b) $\boldsymbol{\xi}_3^{(5)}$, (c) $\boldsymbol{\xi}_3^{(6)}$, (d) $\boldsymbol{\xi}_3^{(8)}$.

Considerazioni conclusive

In quest'ultima parte del presente elaborato si cerca di dare un resoconto dei diversi risultati ottenuti. A distanza di più di due anni dall'inizio della pandemia, e grazie ai dati messi a disposizione da YouGov, raccolti attraverso un sondaggio (Jones Sarah P. e Plc. 2020), si è cercato di categorizzare in diversi comportamenti "ideali" i rispondenti, estendendo poi i risultati alla popolazione italiana. Le unità statistiche campionate, hanno risposto a delle domande a risposta categoriale, che riguardano l'ottemperanza o meno ad alcune misure sanitarie per la prevenzione del contagio, oltre che ad alcune domande che invece cercano di riassumere il pensiero dei cittadini verso le decisioni del Governo ed i vaccini. Vengono poi richiesti alcuni dati personali dei suddetti soggetti, che aiutino in qualche modo a classificarli da un punto di vista socio-demografico.

In particolare, grazie al modello Bayesiano non parametrico implementato, di cui si può trovare la completa specificazione passo-passo nel capitolo 2, si è potuto procedere con abbastanza flessibilità da poter trattare il presente problema. La flessibilità richiesta infatti non è stata poca, considerata anche la dimensionalità e la complessità del tensore adoperato per incasellare le risposte alle diverse domande di ciascun individuo.
Inoltre, l'utilizzo di una Mistura standard di modelli di regressione di esperti, e l'aver quindi consentito la dipendenza dalle covariate sia per i pesi che per i *kernel* componenti la mistura, ha complicato ulteriormente la trattazione, garantendo però allo stesso tempo una maggior malleabilità. Complice anche il diverso periodo considerato, che si ricorda essere dal 27 dicembre 2021 al 25 marzo 2022, i profili dei raggruppamenti ottenuti non si esplicitano in livelli di ottemperanza alle misure anti-Covid, come invece visto in Aliverti e Russo

(2022) ed ipotizzato nella sezione 2.2; questi piuttosto si identificano in livelli di approvazione delle misure imposte dal Governo, e in termini di efficacia sia nel prevenire il contagio da parte dei vaccini, sia nella fiducia in essi riposta. Infatti, per tutti i raggruppamenti trovati, i cittadini seguivano più o meno scrupolosamente le misure di prevenzione; sotto questo punto di vista, non si è pertanto trovata una differenziazione netta tra i diversi cluster.

Il primo cluster, denominato come "Soggetti insofferenti non impauriti dal Covid e No-vax", riguarda appunto tutti quegli individui che non si ritengono soddisfatti dalle manovre perseguite con i diversi DPCM, né si fidano dei vaccini: per tali soggetti il Covid non è pericoloso.

Nel raggruppamento "Pro-vax scontenti del Governo", si identificano invece tutti coloro che non approvano quanto deciso dall'Esecutivo, ma che si dichiarano fortemente a favore dei vaccini, invocandone anche l'obbligatorietà.

Nel terzo ed ultimo cluster analizzato, "No-vax contenti del Governo", si possono trovare tutti coloro che assecondano le mosse dei ministri, ma che non approvano i vaccini, non riponendovi di fatto molta fiducia; non vi è però un pensiero comune maggioritario in merito alla loro obbligatorietà.

Si è poi proceduto con la caratterizzazione socio-demografica di ciascuna sotto-popolazione.
Il primo cluster trovato comprende più probabilmente al suo interno soggetti di genere maschile, che abitano nelle Isole o nel Sud-Italia, e che siano studenti a tempo pieno o in pensione.
Il secondo raggruppamento comprende più plausibilmente, condizionatamente al non essere stati assegnati al gruppo precedente, persone di genere femminile che abitano nelle Isole, nel Nord-Ovest oppure nel Sud Italia. Relativamente allo stato occupazionale, si osserva la positività per i coefficienti dei parametri "non lavora" e "part-time".
Infine, all'interno del terzo cluster, *conditio sine qua non* di non partecipare ai primi 2 gruppi, si troveranno maggiormente individui di genere femminile, abitanti nelle Isole o nel Sud-Italia che siano non lavoratrici, in pensione, disoccupate o con impiego part-time. Tutto ciò è ovviamente da considerarsi in relazione alle modalità base di riferimento.

In ultima istanza, la caratterizzazione della risposta media ad alcune do-

mande, in base al cluster di appartenenza e a dei regressori, è stata effettuata. Nel primo cluster, in relazione alla domanda "Hai indossato la mascherina fuori da casa tua", i maggiori contribuenti alle categorie più scrupolose di risposta sono il genere maschile, e l'abitare nel Nord-Est Italia o nelle Isole. Considerando invece la gestione del COVID da parte del governo, questa assumerà più probabilmente modalità positive se il genere è maschile, o si abita nel Nord-Ovest. La pericolosità del COVID viene maggiormente percepita dagli isolani, d'altro canto, non è stato possibile discernere nettamente chi ripone più fiducia nei vaccini.
In relazione al secondo raggruppamento, contrariamente a quanto visto nel primo, il genere maschile fa tendere la risposta media, della domanda concernente l'utilizzo delle mascherine, verso le modalità che indicano una minor frequenza di indossaggio. Giudizi positivi sull'operato dell'Esecutivo vengono maggiormente dati dai maschi o dagli abitanti nel Nord-Est o Nord-Ovest. Una percezione del pericolo più alta è invece percepita prevalentemente nel Nord-Ovest, nelle Isole o nel Sud Italia. La maggior fiducia nei vaccini viene prevalentemente riposta dagli abitanti del Nord-Est.
Per il terzo ed ultimo gruppo esaminato, essere isolano e di genere maschile contribuirà ad un utilizzo più frequente delle mascherine. Gli abitanti a Nord-Ovest ed i maschi tenderanno a dare giudizi maggiormente positivi in merito alla gestione dell'emergenza. Come visto per il secondo cluster, sarà sempre chi abita nelle Isole, nel Nord-Ovest o nel Sud a percepire di più la pericolosità del COVID. Infine, una maggiore aspettativa sull'efficacia dei vaccini verrà riposta da chi vive al Sud o nelle Isole.

I risultati ottenuti possono essere utilizzati per molteplici scopi. Primo fra tutti si annovera sicuramente la possibilità, da parte del Governo, di agire per cercare di tutelare tutte quelle persone che non sono contente dei provvedimenti emanati. Queste pratiche aiuterebbero molto anche nel rallentare la diffusione del virus, come già discusso nella sezione 1.1. Un altro impiego concerne invece una campagna di sensibilizzazione vaccinale che sia più mirata verso tutte quelle persone che ad oggi risultano ancora impermeabili a quanto già é stato detto a riguardo.

Possibili sviluppi e miglioramenti

Primo fra tutti, un miglioramento fin da subito implementabile è quello di utilizzare un numero H di cluster più elevato nella scrittura del modello a mistura infinita in NIMBLE. In ogni caso, avendo a disposizione un computer più performante ed un intervallo di tempo sensibilmente maggiore a quello utilizzato, si può ovviare a questo limite. Il secondo limite, invece, è conseguenza di questo tipo di sondaggi: non tutta la popolazione viene raggiunta allo stesso modo e, come già detto nella sezione 1.2.3, si avranno certe fasce che sono sotto-rappresentate.

Uno sviluppo invece perseguibile in un contesto di modellazione dinamica del fenomeno, sarebbe quello di consentire la dipendenza dal tempo t nella matrice di covariate, e magari procedere ad un confronto per diversi intervalli temporali, osservando come variano i profili latenti nel tempo. In tal senso, una possibile specificazione è riportata in Aliverti e Russo (2022), dove però viene utilizzata una mistura finita di componenti.

Appendice A

Linea temporale eventi

Tabella A.1: Linea temporale eventi

13/02/20	**Governo Conte II**.
21/02/20	Misure di isolamento quarantenario obbligatorio per i contatti stretti con un caso risultato positivo, e viene disposta la sorveglianza attiva con permanenza domiciliare fiduciaria per chi è stato nelle aree a rischio negli ultimi 14 giorni, con obbligo di segnalazione da parte del soggetto interessato alle autorità sanitarie locali.
23/02/20	Registrati focolai in Lombardia e Veneto, firma dell'odierno DPCM di attuazione delle disposizioni del DL 6/2020 per i comuni delle Regioni Lombardia e Veneto interessati dalla misure di contenimento del contagio da Coronavirus.
09/03/20	Il PdC Conte firma il DPCM 9 marzo 2020 recante nuove misure per il contenimento e il contrasto del diffondersi del virus Covid-19 sull'intero territorio nazionale. È inoltre vietata ogni forma di assembramento di persone in luoghi pubblici o aperti al pubblico.

20/03/20	Il ministro della salute ha firmato l'ordinanza che vieta: l'accesso del pubblico ai parchi, alle ville, alle aree gioco e ai giardini pubblici, di svolgere attività ludica o ricreativa all'aperto. È vietato ogni spostamento verso abitazioni diverse da quella principale, comprese le seconde case utilizzate per le vacanze.
22/03/20	Nuova ordinanza che vieta a tutte le persone fisiche di trasferirsi o spostarsi con mezzi di trasporto pubblici o privati in comune diverso da quello in cui si trovano, salvo che per comprovate esigenze lavorative, di assoluta urgenza ovvero per motivi di salute. Rimangono aperti solamente alimentari, farmacie, negozi di generi di prima necessità e i servizi essenziali.
10/04/20	Firma del nuovo DPCM che proroga fino al 3 maggio le misure restrittive sin qui adottate per il contenimento dell'emergenza epidemiologica.
15/05/20	Approvazione di un nuovo decreto-legge che delinea il quadro normativo nazionale all'interno del quale, dal 18 maggio al 31 luglio 2020, con appositi decreti od ordinanze, statali, regionali o comunali, potranno essere disciplinati gli spostamenti delle persone fisiche e le modalità di svolgimento delle attività economiche, produttive e sociali.
11/06/20	Il PdC firma un nuovo DPCM che autorizza la ripresa di ulteriori attività a partire dal 15 giugno tra cui: centri estivi per i bambini, sale giochi, così come le attività di centri benessere, centri culturali. Riprendono, inoltre, gli spettacoli aperti al pubblico ma con alcune cautele/precauzioni.
30/07/20	Il CdM ha approvato un nuovo decreto-legge: il testo proroga, dal 31 luglio al 15 ottobre 2020, la possibilità di adottare specifiche misure di contenimento dell'epidemia.
07/08/20	Approvazione di un nuovo decreto-legge che proroga fino al 7 settembre 2020 le misure precauzionali minime per contrastare e contenere il diffondersi del COVID-19.
07/09/20	Proroga fino al 7 ottobre delle misure contenute nel DPCM del 7 agosto 2020.

07/10/20	Il CdM, vista la nota del Ministro della salute e il parere del Comitato tecnico scientifico, ha deliberato la proroga, fino al 31 gennaio 2021, dello stato d'emergenza dichiarato in conseguenza della dichiarazione di "emergenza di sanità pubblica di rilevanza internazionale" da parte della Organizzazione mondiale della sanità (OMS).
18/10/20	Il PdC firma un nuovo DPCM che prevede ulteriori misure di natura restrittiva, al fine di contenere quanto più possibile il contagio, in presenza di una recrudescenza del virus, ormai in atto da alcune settimane.
03/11/20	Un nuovo corpus di misure restrittive viene approvato. Il nuovo DPCM individua tre aree - gialla, arancione e rossa - rispettive ai differenti livelli di criticità nelle Regioni del Paese e per le quali sono previste misure specifiche.
02/12/20	Approvazione di un nuovo decreto legge che estende il limite massimo di vigenza dei decreti del Presidente del Consiglio dei Ministri attuativi delle norme emergenziali, portandolo dagli attuali trenta a cinquanta giorni. Il provvedimento, inoltre, prevede ulteriori misure restrittive per il periodo 21 dicembre 2020 - 6 gennaio 2021.
04/01/21	Il CdM ha approvato un decreto-legge che introduce ulteriori misure restrittive in merito agli spostamenti per il periodo 7-15 gennaio 2021.
13/01/20	Proroga dello stato di emergenza fino al 30 aprile 2021. Istituzione di una nuova area bianca nella quale si collocano le Regioni con un livello di rischio basso.
12/02/21	Il CdM approva nuovo decreto-legge che proroga fino al 25 febbraio 2021 il divieto di spostarsi tra regioni o province autonome diverse, fatta eccezione per comprovate esigenze lavorative o motivi di salute.

13/02/21	**Governo Draghi**.
22/02/21	Proroga di quanto visto in precedenza fino al 27 marzo 2021. Nelle zone rosse non sono consentiti gli spostamenti verso abitazioni private abitate diverse dalla propria, salvo motivazioni di salute o lavorative.
12/03/21	In considerazione della maggiore diffusività del virus e delle sue varianti e in vista delle festività pasquali, il provvedimento stabilisce misure di maggiore intensità rispetto a quelle già in vigore. Dal 15 marzo al 2 aprile 2021 e nella giornata del 6 aprile 2021, in tutte le zone gialle si applicano le disposizioni previste per le zone arancioni e nei giorni 3, 4 e 5 aprile 2021, su tutto il territorio nazionale (tranne che nelle zone bianche), si applicheranno le restrizioni previste per le zone rosse.
26/03/21	Il PdC Mario Draghi e il Ministro della Salute Roberto Speranza hanno tenuto una conferenza stampa, si è parlato di conclusioni del Consiglio Europeo, della campagna vaccinale e delle riunioni per l'emergenza Covid.
31/03/21	Approvazione Decreto-legge che conferma quanto già visto ed inoltre dispone che dal 7 al 30 aprile 2021 sia assicurato, sull'intero territorio nazionale, lo svolgimento in presenza dei servizi educativi per l'infanzia e della scuola dell'infanzia, nonché dell'attività didattica del primo ciclo di istruzione e del primo anno della scuola secondaria di primo grado.
31/03/21	Per i successivi gradi di istruzione è confermato lo svolgimento delle attività in presenza dal 50% al 75% della popolazione studentesca in zona arancione mentre in zona rossa le relative attività si svolgono a distanza, garantendo comunque la possibilità di svolgere attività in presenza per gli alunni con disabilità e con bisogni educativi speciali.
21/04/21	Approvazione decreto Riaperture, il cui testo delinea il cronoprogramma relativo alla progressiva eliminazione delle restrizioni rese necessarie per limitare il contagio da virus SARS-CoV-2, alla luce dei dati scientifici sull'epidemia e dell'andamento della campagna di vaccinazione. Proroga fino al 31 luglio 2021 dello stato di emergenza.

17/05/21	Tramite Decreto-legge vengono modificati i parametri di ingresso nelle "zone colorate": assumono principale rilievo l'incidenza dei contagi rispetto alla popolazione complessiva nonché il tasso di occupazione dei posti letto in area medica e in terapia intensiva. Ulteriori modifiche al "calendario delle riaperture" per la ripresa delle attività economiche e sociali nelle "zone gialle".
22/07/21	Nuova proroga dello stato di emergenza fino al 31 dicembre 2021 e si sono decise le modalità di utilizzo del Green Pass e nuovi criteri per la "colorazione" delle regioni.
10/09/21	Estensione dell'obbligatorietà del Green Pass in ambito scolastico, della formazione superiore e socio sanitario-assistenziale.
16/09/21	Nuovo decreto-legge che introduce misure urgenti per assicurare lo svolgimento in sicurezza del lavoro pubblico e privato mediante l'estensione dell'ambito applicativo della certificazione verde COVID-19 e il rafforzamento del sistema di screening.
12/10/21	Approvazione DPCM che sancisce le linee guida relative all'obbligo di possesso e di esibizione della certificazione verde COVID-19 da parte del personale delle PA, a partire dal prossimo 15 ottobre; altro DPCM contenente le modalità di verifica del possesso delle certificazioni verdi COVID-19 in ambito lavorativo.
24/11/21	Approvazione decreto-legge il cui testo prevede una serie di misure di contenimento della "quarta ondata" della pandemia Sars-Cov2 in quattro ambiti: obbligo vaccinale e terza dose; estensione dell'obbligo vaccinale a nuove categorie; istituzione del Green Pass rafforzato; rafforzamento dei controlli e campagne promozionali sulla vaccinazione.
15/12/21	Proroga dello stato di emergenza nazionale fino al 31 marzo 2022. Restano in vigore le norme relative all'impiego del Green pass normale e rafforzato.

23/12/21	Riduzione della durata del green pass vaccinale da 9 a 6 mesi a partire dal 01/02/22, obbligo di indossare le mascherine anche all'aperto e anche in zona bianca; obbligo di indossare le mascherine di tipo FFP2 su tutti i mezzi di trasporto e in occasione di spettacoli all'aperto e al chiuso e per gli eventi e competizioni sportive.
29/12/21	Decreto-legge che prevede nuove misure in merito all'estensione del Green Pass rafforzato (che si può ottenere con il completamento del ciclo vaccinale e la guarigione) e le quarantene per i vaccinati.
05/01/22	Decreto-legge che introduce Misure urgenti per fronteggiare l'emergenza COVID-19, in particolare nei luoghi di lavoro e nelle scuole.
31/03/22	Termine stato di emergenza.

Appendice B

Catene e diagnostiche di convergenza

Nella figura B.1 si nota un buon *mixing* per ciascuna delle varianze σ_h. La media cumulata dei parametri ad ogni iterazione, che viene riportata in blu, non pone dubbi sulla convergenza delle catene ottenute.

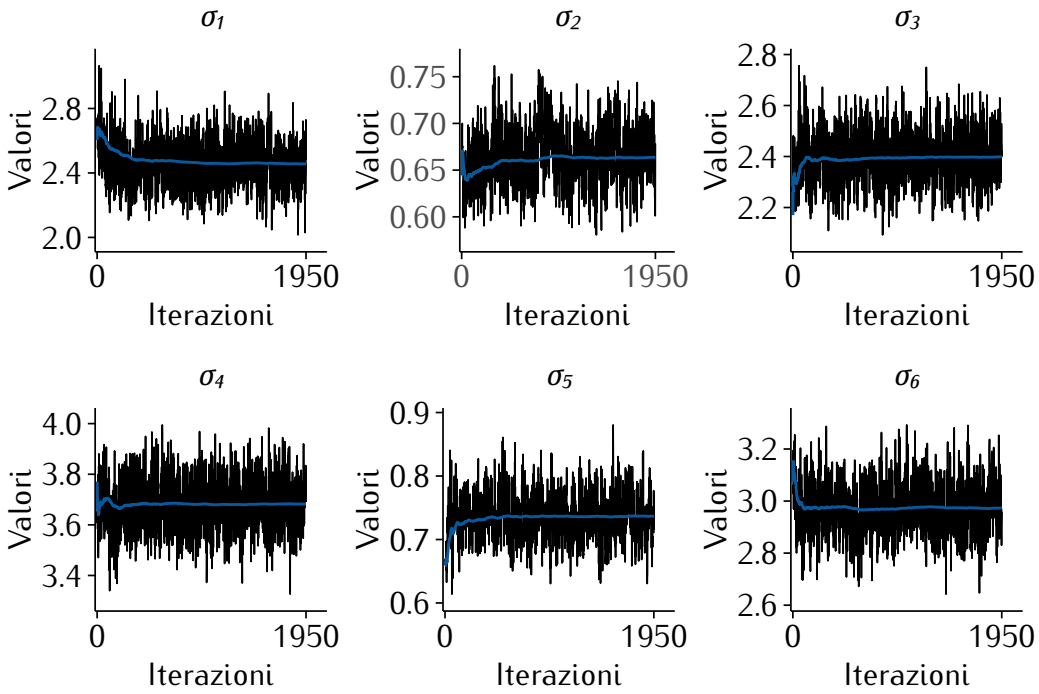


Figura B.1: Catena simulazione a posteriori per σ_h, $h = 1, \ldots, 6$.

Per quanto riguarda i parametri γ_{hp}, con $h = 1,\ldots,6$ e $p = 1,\ldots,13$ si è deciso di riportare, nella figura B.2, solo le catene relative a γ_{2p}. La scelta è ricaduta su quelle relative al secondo cluster in quanto tutti gli altri grafici mostravano andamenti del tutto analoghi in termini di *mixing*. Si può chiaramente notare, grazie alla media cumulata e all'andamento della catene, un buon *mixing* ed il raggiungimento della convergenza.

Infine, relativamente a $\boldsymbol{\xi}_h^{(j)}$, con $j = 1,\ldots,9$, si può prendere visione alla figura B.3 dei risultati ottenuti. Si è deciso di riportare solamente le catene ottenute per la variabile risposta y_2, ossia con $j = 2$. In ogni caso, il *mixing* è risultato essere ottimo anche per tutte le altre catene relative alle altre risposte, con la media cumulata che effettivamente arrivava a convergenza.

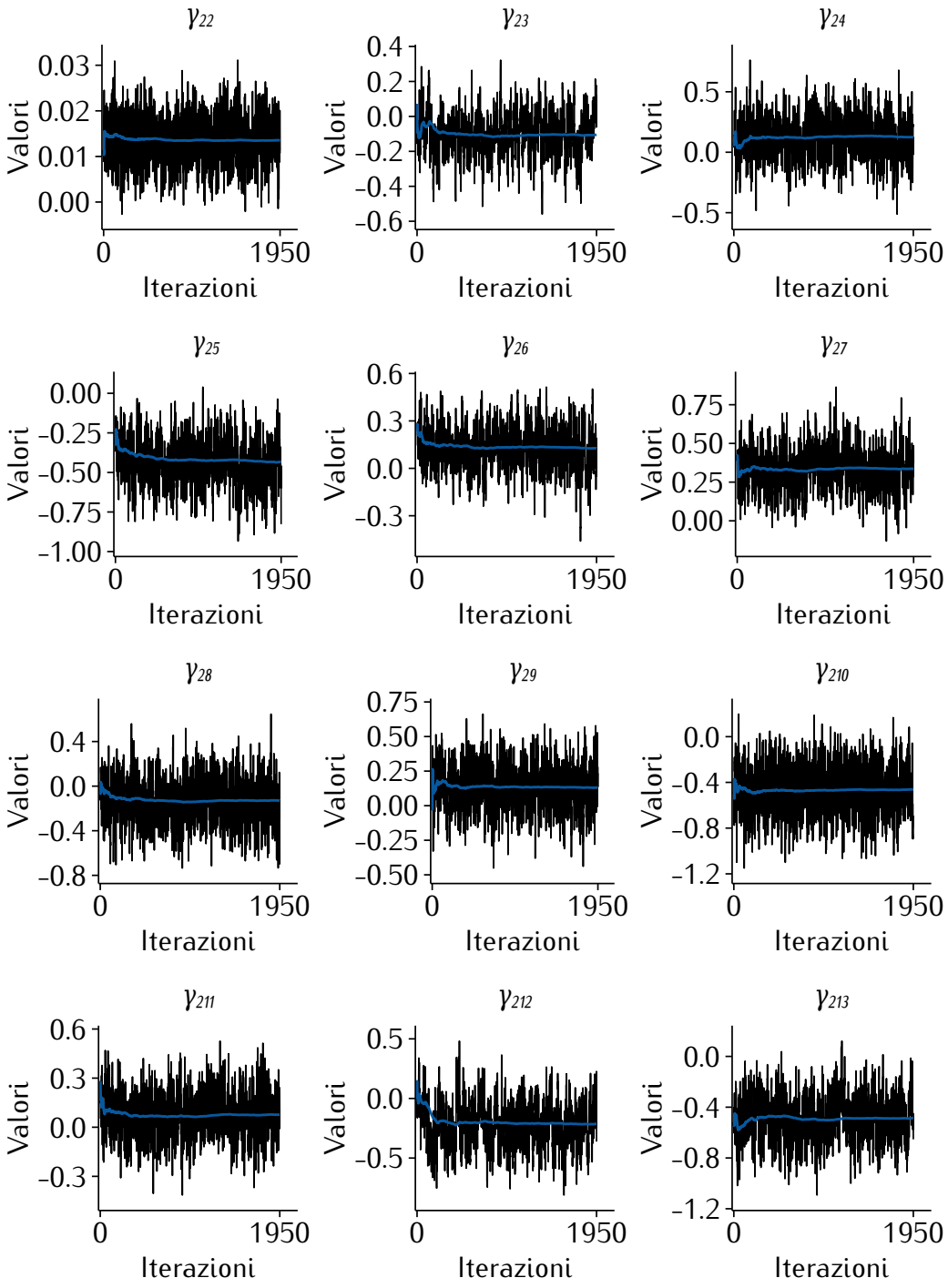


Figura B.2: Catena simulazioni a posteriori per γ_{2p}, $p = 2, \ldots, 13$.

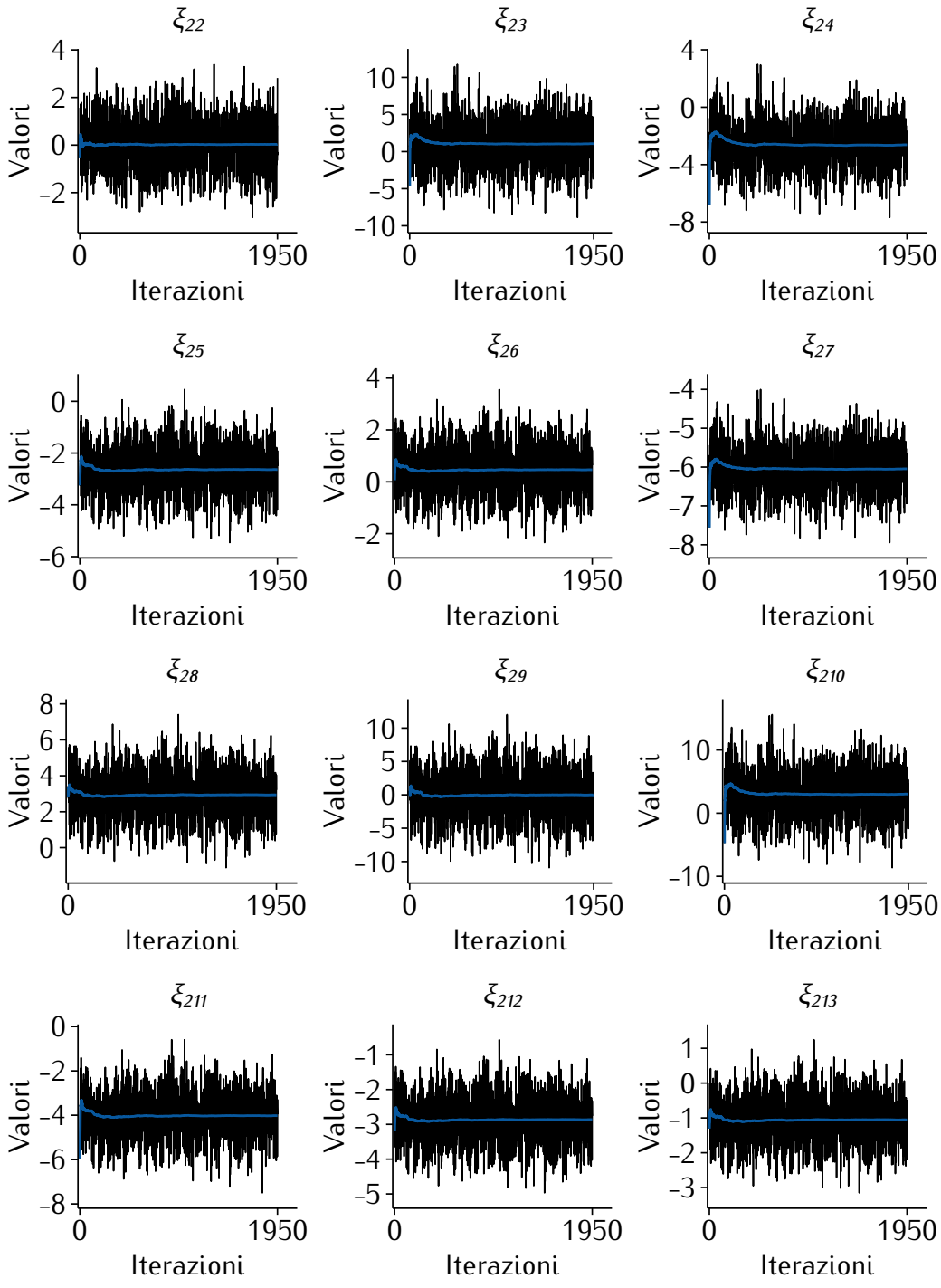


Figura B.3: Catena simulazioni a posteriori per $\boldsymbol{\xi}_2^{(2)}$.

Bibliografia

[1] Emanuele Aliverti e Massimilano Russo. «Dynamic modeling of the Italians' attitude towards Covid-19». In: *Statistics in Medicine* (in press) (2022).

[2] John Burkardt. *The Truncated Normal Distribution*. Florida State University: Department of Scientific Computing, 2014. URL: http://people.sc.fsu.edu/%E2%88%BCjburkardt/presentations/truncated%20normal.pdf.

[3] Antonio Canale e David B. Dunson. «Bayesian Kernel Mixtures for Counts». In: *Journal of the American Statistical Association* 106.496 (2011), pp. 1528–1539. DOI: 10.1198/jasa.2011.tm10552.

[4] Claus-Christian Carbon. «About the acceptance of wearing face masks in times of a pandemic». In: *i-Perception* 12.3 (2021).

[5] Leonardo Carlucci, Ines D'Ambrosio e Michela Balsamo. «Demographic and Attitudinal Factors of Adherence to Quarantine Guidelines During COVID-19: The Italian Model». In: *Frontiers in Psychology* 11 (2020). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2020.559288.

[6] COVIDSTAT. *Il tasso di contagio Rt*. CovidStat INFN. 2022. URL: https://covid19.infn.it/sommario/rt-info.html.

[7] W Cullen, G Gulati e B D Kelly. «Mental health in the COVID-19 pandemic». In: *QJM: An International Journal of Medicine* 113.5 (mar. 2020), pp. 311–312. ISSN: 1460-2725. DOI: 10.1093/qjmed/hcaa110.

[8] Perry de Valpine, Christopher Paciorek et al. *NIMBLE User Manual*. R package manual Version 0.12.2. 2022. DOI: 10.5281/zenodo.1211190.

[9] Perry de Valpine, Christopher Paciorek et al. *NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling*. R package Version 0.12.2. 2022. DOI: 10.5281/zenodo.1211190.

[10] Perry de Valpine, Daniel Turek et al. «Programming with models: writing statistical algorithms for general model structures with NIMBLE». In: *Journal of Computational and Graphical Statistics* 26 (2 2017), pp. 403–413. DOI: 10.1080/10618600.2016.1172487.

[11] Gerard Delanty, cur. *Critical Perspectives on the Covid-19 Crisis*. Berlin, Boston: De Gruyter, 2021. ISBN: 9783110713350. DOI: doi:10.1515/9783110713350.

[12] Kjell Doksum. «Tailfree and Neutral Random Probabilities and Their Posterior Distributions». In: *The Annals of Probability* 2.2 (1974), pp. 183–201. DOI: 10.1214/aop/1176996703.

[13] Thomas S. Ferguson. «A Bayesian Analysis of Some Nonparametric Problems». In: *The Annals of Statistics* 1.2 (1973), pp. 209–230. DOI: 10.1214/aos/1176342360.

[14] Giuseppe Forte et al. «The Enemy Which Sealed the World: Effects of COVID-19 Diffusion on the Psychological State of the Italian Population». In: *Journal of Clinical Medicine* 9.6 (2020). ISSN: 2077-0383. DOI: 10.3390/jcm9061802.

[15] Andrew Béla Frigyik, Amol Kapila e Maya R. Gupta. «Introduction to the Dirichlet Distribution and Related Processes». In: 2010.

[16] S. Fruhwirth-Schnatter, G. Celeux e C.P. Robert. *Handbook of Mixture Analysis*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2019. ISBN: 9780429508240. DOI: https://doi.org/10.1201/9780429055911.

[17] Ryosuke Fujii, Kensuke Suzuki e Junichiro Niimi. «Public perceptions, individual characteristics, and preventive behaviors for COVID-19 in six countries: A cross-sectional study». In: *Environmental Health and Preventive Medicine* 26.1 (2021), pp. 1–12.

[18] Google. *COVID-19 - Report sugli spostamenti della comunità*. Visitato il 28/07/2022. 2022. URL: https://www.google.com/covid19/mobility/.

[19] Jinyang Gu, Bing Han e Jian Wang. «COVID-19: Gastrointestinal Manifestations and Potential Fecal–Oral Transmission». In: *Gastroenterology* 158.6 (2020), pp. 1518–1519. ISSN: 0016-5085. DOI: https://doi.org/10.1053/j.gastro.2020.02.054.

[20] Hatice Rahmet Güner, İmran Hasanoğlu e Firdevs Aktaş. «COVID-19: Prevention and control measures in community». In: *Turkish Journal of medical sciences* 50.9 (2020), pp. 571–577.

[21] ISS. *Aggiornamento nazionale*. Report Esteso ISS. Task force COVID-19 del Dipartimento Malattie Infettive e Servizio di Informatica. Roma: Istituto Superiore di Sanità (ISS), 2022.

[22] Imperial College London Big Data Analytical Unit Jones Sarah P. e YouGov Plc. *Imperial College London YouGov Covid Data Hub*. Aprile 2020. YouGov Plc. 2020. URL: https://github.com/YouGov-Data/covid-19-tracker.

[23] Athanasios Kottas, Peter Müller e Fernando Quintana. «Nonparametric Bayesian Modeling for Multivariate Ordinal Data». In: *Journal of Computational and Graphical Statistics* 14.3 (2005), pp. 610–625. DOI: 10.1198/106186005X63185.

[24] Tsuyoshi Kunihama e David B. Dunson. «Bayesian Modeling of Temporal Dependence in Large Sparse Contingency Tables». In: *Journal of the American Statistical Association* 108.504 (2013), pp. 1324–1338. DOI: 10.1080/01621459.2013.823866.

[25] Marina Meilă. «Comparing Clusterings by the Variation of Information». In: *Learning Theory and Kernel Machines*. A cura di Bernhard Schölkopf e Manfred K. Warmuth. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 173–187. ISBN: 978-3-540-45167-9.

[26] Lorenzo Moccia et al. «Affective temperament, attachment style, and the psychological impact of the COVID-19 outbreak: an early report on the Italian general population». In: *Brain, Behavior, and Immunity* 87 (2020), pp. 75–79. ISSN: 0889-1591. DOI: https://doi.org/10.1016/j.bbi.2020.04.048.

[27] Joy D Osofsky, Howard J Osofsky e Lakisha Y Mamon. «Psychological and social impact of COVID-19.» In: *Psychological Trauma: Theory, Research, Practice, and Policy* 12.5 (2020), p. 468.

[28] Stack overflow. 2022. URL: https://stackoverflow.com/.

[29] Kelly Quinn. «Methodological considerations in surveys of older adults: technology matters». In: *International Journal of Emerging Technologies and Society* 8 (gen. 2010), pp. 114–133. DOI: 10.13140/2.1.3897.9209.

[30] Fernando A. Quintana et al. «The Dependent Dirichlet Process and Related Models». In: *Statistical Science* 37.1 (2022), pp. 24–41. DOI: 10.1214/20-STS819.

[31] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2022. URL: https://www.R-project.org/.

[32] Véronique Renault et al. «Main determinants of the acceptance of COVID-19 control measures by the population: A first pilot survey at the University of Liege, Belgium». In: *Transboundary and Emerging Diseases* 69.4 (2022), e1065–e1078. DOI: https://doi.org/10.1111/tbed.14410.

[33] Tommaso Rigon e Daniele Durante. «Tractable Bayesian density regression via logit stick-breaking priors». In: *Journal of Statistical Planning and Inference* 211 (2021), pp. 131–142. ISSN: 0378-3758. DOI: https://doi.org/10.1016/j.jspi.2020.05.009.

[34] Malik Sallam. «COVID-19 Vaccine Hesitancy Worldwide: A Concise Systematic Review of Vaccine Acceptance Rates». In: *Vaccines* 9.2 (2021). ISSN: 2076-393X. DOI: 10.3390/vaccines9020160.

[35] Jayaram Sethuraman. «A constructive definition of Dirichlet Priors». In: *Statistica Sinica* 4.2 (1994), pp. 639–650. ISSN: 10170405, 19968507. URL: http://www.jstor.org/stable/24305538 (visitato il 09/08/2022).

[36] Sara Wade e Zoubin Ghahramani. «Bayesian Cluster Analysis: Point Estimation and Credible Balls (with Discussion)». In: *Bayesian Analysis* 13.2 (2018), pp. 559–626. DOI: 10.1214/17-BA1073.

[37] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

[38] Wanja Wolff et al. «High Boredom Proneness and Low Trait Self-Control Impair Adherence to Social Distancing Guidelines during the COVID-19 Pandemic». In: *International Journal of Environmental Research and Public Health* 17.15 (2020). ISSN: 1660-4601. DOI: 10.3390/ijerph17155420.