

**UNIVERSITÀ DEGLI STUDI MILANO-BICOCCA**

Scuola di Economia e Statistica

Corso di laurea in

**STATISTICA E GESTIONE DELL'INFORMAZIONE**



**APPLICAZIONE MODELLI DI REGRESSIONE  
SELF-SELECTION PER L'ANALISI DEI SINTOMI  
E DETERMINANTI DEL LONG COVID**

Relatore: Prof. Pietro Giorgio Lovaglio

Tesi di Laurea di:  
Luca De Simone  
Matr. N. 854294

Anno Accademico 2021/202

## Indice

|   |           |
|---|-----------|
| <b>Introduzione</b> .....   | <b>1</b>  |
| <b>1. Descrizione dei dati</b> .....  | <b>3</b>  |
| 1.1. Descrizione variabili esplicative.....   | 3         |
| 1.2. Descrizione variabili dipendenti.....  | 6         |
| <b>2. Pre-Processing variabili dipendenti, esplicative e Componenti Principali</b> .....                    | <b>8</b>  |
| 2.1. Pre-Processing variabili dipendenti .....  | 8         |
| 2.2. Costruzione componenti principali.....   | 9         |
| 2.3. Pre-Processing variabili esplicative.....  | 14        |
| <b>3. Analisi descrittive</b> .....   | <b>16</b> |
| 3.1. Analisi delle sintomatologie per ondate .....  | 16        |
| 3.2. Analisi profili dei soggetti per ondate .....  | 17        |
| <b>4. Modelli di Self-Selection</b> .....   | <b>21</b> |
| 4.1. Censored Sample e Correzione di Heckman .....  | 21        |
| 4.2. Modello Heckit.....  | 22        |
| <b>5. Analisi Preliminari e Model Selection</b> .....   | <b>24</b> |
| 5.1. Analisi Preliminari: studio della Collinearità, Separation/Quasi-Separation e Near-Zero-Variance ..... | 24        |
| 5.2. Model Selection.....   | 25        |
| 5.3. Studio delle Interazioni.....  | 27        |
| <b>6. Analisi dei risultati</b> .....   | <b>29</b> |
| 6.1. Interpretazione dei punteggi dei Macro-Sintomi.....  | 32        |
| 6.2. Analisi Macro-Sintomo Psicologico.....   | 33        |
| 6.3. Analisi Macro-Sintomo Fisico.....  | 35        |
| <b>7. Conclusioni</b> .....   | <b>38</b> |
| <b>Bibliografia e Sitografia</b> .....  | <b>39</b> |

## Introduzione

L'Istituto Superiore di Sanità, nel Rapporto ISS CoViD-19 n. 15/2021, definisce Long Covid *“la condizione di persistenza di segni e sintomi che continuano o si sviluppano dopo un'infezione acuta da SARS-CoV-2. Se i sintomi continuano a manifestarsi oltre quattro settimane dall'infezione fino a 12 settimane, si parla di malattia CoViD-19 sintomatica persistente; se i sintomi si prolungano per più di 12 settimane e non possono essere spiegati da nessun'altra condizione, si parla di Sindrome post-CoViD. Il Long-CoViD include entrambe queste condizioni.”*

Essendo un fenomeno recente e in continuo sviluppo, la definizione sopra citata è in costante aggiornamento e, in seguito ad un'attenta revisione bibliografica, le formulazioni risultano essere molteplici. Per citarne alcune:

- *“La condizione post COVID-19 si verifica in individui con una storia di infezione da SARS CoV-2 probabile o confermata, di solito 3 mesi dall'inizio del COVID-19 con sintomi e che durano per almeno 2 mesi e non possono essere spiegati da una diagnosi alternativa”* (World Health Organization)
- *“Long-COVID è il termine utilizzato al fine di denotare una persistenza di sintomi in quelle persone che sono state ricoverate per il virus SARS-CoV-2”* (Raveendran et al., 2021).
- *“Segni e sintomi che si sviluppano durante o dopo un'infezione compatibile con COVID-19, continuano per più di 12 settimane e non sono spiegati da una diagnosi alternativa. Di solito si presenta con grappoli di sintomi, spesso sovrapposti, che possono fluttuare e cambiare nel tempo, oltre ad essere in grado di interessare qualsiasi sistema ed apparato dell'organismo.”* (National Institute for Health and Clinical Excellence)

Il seguente studio si pone come scopo l'analisi dei determinanti delle sintomatologie Long-Covid, fenomeno che, in questa trattazione, è definito generalmente come la persistenza e lo sviluppo dei sintomi oltre il decorso clinico della malattia

I dati, per affrontare la ricerca, sono stati forniti da un grande ospedale metropolitano specializzato nella cura del Covid-19 e del Long-Covid. Il campione, costituito da 1391 osservazioni, è composto da soggetti che, dopo aver contratto l'infezione da SARSCOV2 ed esserne guariti, hanno aderito autonomamente all'indagine promossa dalla struttura

ospedaliera. Gli individui in analisi hanno contratto l'infezione in due ondate pandemiche differenti, riguardanti rispettivamente il periodo dal 23.02.2020 al 01.10.2020 e quello dal 01.10.2020 al 26.07.2021. Poiché i soggetti hanno contratto il virus in periodi diversi, i tempi di follow-up risultano differenti e le osservazioni si sono concluse in data 14.10.2021 con l'acquisizione dell'ultimo questionario.

Per perseguire l'obiettivo in studio sono state costituite delle Macro-Sintomatologie, espresse in termini di durate, mediante la tecnica della Componenti Principali. Successivamente lo studio dei determinanti è stato affrontato tramite l'applicazione di modelli di regressione lineari di Self-Selection.

La trattazione sarà composta da due parti e si aprirà con la descrizione delle variabili che costituiscono il campione in studio. Queste sono suddivise in sintomatologie post-Covid, variabili dipendenti, e caratteristiche clinico-anagrafiche dei pazienti, variabili esplicative.

Inseguito ad una fase di pulizia dei dati, il secondo capitolo si concluderà con l'aggregazione dei sintomi in macro-gruppi tramite la tecnica delle componenti principali. Poiché uno degli scopi dell'analisi consiste nell'identificare i determinanti Long-Covid, nel terzo capitolo, verranno eseguite delle analisi esplorative in grado di evidenziare ciò che ha contraddistinto ogni ondata infettiva.

La seconda parte della trattazione si concentrerà sulle fasi di analisi. Il quarto capitolo, infatti, è dedicato allo studio dei modelli di Self Selection, in particolare verrà descritto il metodo di stima, le sue proprietà, i vantaggi e i limiti d'applicazione.

Lo studio degli effetti delle covariate sui macro-sintomi suddivisi per ondate, riportato nel sesto capitolo, è stato preceduto da una fase preliminare di selezione delle variabili d'interesse. Questo processo ha permesso di esplicitare al meglio i nessi esistenti tra le caratteristiche dei pazienti e le sintomatologie post-Covid. Infine, lo studio si concluderà con l'esposizione dei risultati ottenuti, eventuali limiti e alcuni spunti di riflessione per eventuali analisi future.

# Capitolo 1

## Descrizione dei dati

Le informazioni relative ai soggetti in studio, rilevate dall'istituto ospedaliero, sono di diversa natura e ripartite nelle seguenti macroaree:

- Anagrafiche dei pazienti e informazioni personali.
- Farmaci e terapie assunte prima dell'infezione acuta da SARSCOV2.
- Patologie pregresse dei pazienti riscontrate prima dell'infezione da Covid-19.
- Terapie assunte in corso di infezione.
- Sintomi Post-Covid riscontrati dai pazienti nel tempo e oggettivati dai clinici (target in analisi).

Il campione, essendo costituito da soggetti che hanno aderito spontaneamente all'indagine, risulta essere auto selezionato anziché casuale. Potrebbe, quindi, soffrire di un errore sistematico, generato dal processo di reclutamento, denominato "Selection bias" che, se non opportunamente controllato, potrebbe distorcere i risultati dello studio. Il Selection bias, noto come distorsione da selezione, si verifica quando vi è una differenza sistematica tra le caratteristiche dei soggetti selezionati per lo studio e quelle di coloro che non sono stati inclusi, compromettendo così la rappresentatività del campione.

Nel corso della trattazione, con il termine paziente, si intenderà un soggetto che ha contratto l'infezione da SARSCOV2 e con paziente asintomatico un individuo che, nonostante la contrazione del virus, non ha sviluppato alcuna sintomatologia Long-Covid.

### 1.1 Descrizione variabili esplicative

La storia clinica dei pazienti è descritta dalle quattro macroaree sopra citate e racchiudono cinquantuno variabili esplicative.

Le informazioni personali comprendono le anagrafiche del paziente: età, sesso, peso, altezza e BMI, indice di massa corporea. Oltre a queste sono riportate le informazioni sull'intensità dell'infezione Covid-19, espressa dall'indice del Word Health Organization (WHO) e sull'eventuale ospedalizzazione del soggetto, rappresentata dalla variabile "Ex Ricovero".

Il secondo macro-gruppo, riguarda le patologie di cui, il paziente, soffriva prima di sviluppare l'infezione. Alcuni casi hanno presentato comorbidità, ossia la presenza o l'insorgenza simultanea di più entità patologiche tra le quali:

- PATOLOGIE POLMONARI: infezioni dei polmoni capace di compromettere il corretto funzionamento di questi.
- PATOLOGIE CARDIACHE: malattie e disturbi che interessano il cuore o alcune sue componenti.
- PATOLOGIE METABOLICHE: patologie che causano un'alterazione del metabolismo dei nutrienti.
- DIABETE: patologia metabolica in cui si verifica un aumento della concentrazione di zuccheri nel sangue.
- PATOLOGIE RENALI: affezione dei reni capace di compromettere il loro corretto funzionamento.
- NEOPLASIE: note anche come tumori indicano patologie in cui si verifica una crescita anomala di cellule nell'organismo.
- PATOLOGIE IMMUNOLOGICHE/AUTOIMMUNI: patologie caratterizzate da un disfunzionamento del sistema immunitario che induce l'organismo ad attaccare i propri tessuti non riconoscendoli come tali.
- EPATOPATIA: affezione del fegato capace di compromettere il suo corretto funzionamento. Si manifesta con steatosi, epatite alcolica o cirrosi epatica.

A seconda della patologia sofferta, i pazienti, prima dell'infezione acuta da SARSCOV2, hanno dichiarato di assumere farmaci, tra i quali:

- TAO/NAO: insieme di farmaci anticoagulanti orali in grado di ridurre il rischio della formazione di coaguli (trombi).
- ANTIAGGREGANTI (cardioaspirina): ulteriore tipologia di farmaci anticoagulanti orali.
- IPOLIPEMIZZANTE/STATINA: farmaci in grado di contrastare e ridurre il colesterolo al fine di diminuire il rischio di malattie coronariche.
- IPOGLICEMIZZANTI: farmaci che permettono la diminuzione della glicemia, concentrazione di glucosio nel sangue, consentendo il corretto funzionamento dell'ormone insulina. Comprendono antidiabetici orali e insulina.

- **CALCIO ANTAGONISTI:** classe di farmaci antipertensivi che agiscono bloccando i canali di calcio presenti sulla parete dei vasi arteriosi e del miocardio, riducendo la pressione arteriosa.
- **BETA-BLOCCANTI:** classe di farmaci antipertensivi.
- **ACE INIBITORI/SARTANI:** farmaci antipertensivi che agiscono sull'ormone dell'angiotensina II provocando un conseguente effetto ipotensivo.
- **DIURETICI:** farmaci antipertensivi che agiscono sul bilancio sodio-acqua, aumentando il volume della diuresi e riducendo così la volemia, quantità complessiva di sangue presente all'interno del sistema circolatorio.
- **ANTIARITMICI:** farmaci in grado di prevenire o interrompere le aritmie, alterazioni della normale regolarità di contrazione del cuore.
- **STEROIDI IMMUNOSOPPRESSORI:** farmaci utilizzati per ridurre l'attività del sistema immunitario in particolari condizioni come trapianti d'organo o in caso di malattie autoimmuni.
- **STEROIDI INALATORI:** steroidi immunosoppressori cortisonici spray utilizzati in caso di reazioni allergiche come asma, orticaria o shock anafilattici oppure per malattie da broncopneumopatia cronico ostruttiva (BPCO).
- **ANTIRETROVIRALI:** farmaci utilizzati per il trattamento di infezioni da virus HIV.
- **DMARDs:** farmaci che rallentano la progressione dell'artrite reumatoide, patologia autoimmune infiammatoria cronica che attacca i tessuti articolari.
- **ANTITUMORALI:** farmaci antineoplastici che vengono utilizzati per il trattamento dei tumori. Ne esistono di diversi tipi, classificabili in base al loro meccanismo di azione e alla loro struttura chimica.
- **NEURO PSICOATTIVI:** farmaci antiepilettici e benzodiazepine. I primi usati per il trattamento dell'epilessia, mentre i secondi sono una classe di farmaci ansiolitici con effetti ipnotici, anticonvulsivanti e anestetici.

Questi farmaci sono spesso utilizzati in modo combinato e, poiché potrebbero risultare fortemente associati tra loro e/o con la patologia per la quale sono stati prescritti, si riterrà opportuno, prima di applicare i metodi di regressione, condurre un'analisi della collinearità in modo da depurare i risultati da possibili distorsioni.

L'ultimo macrogruppo di variabili esplicative descrive le terapie prescritte e somministrate dai medici per contrastare l'infezione acuta da SARS-CoV2, quali:

- **OSSIGENOTERAPIA:** rappresenta la somministrazione di ossigeno al paziente. La somministrazione può avvenire in diversi modi e con diverse intensità.
- **STERODI/CORTICOSTEROIDI:** terapia in cui, attraverso l'assunzione di medicinali di sintesi, si imita l'azione degli ormoni naturali. Hanno proprietà antinfiammatorie e regolano l'attività di metabolismo e sistema immunitario.
- **REMDESEVIR:** Terapia nucleotidica che consiste nel bloccare l'enzima RNA polimerasi e l'eccessiva produzione di ATP nei soggetti affetti da COVID-19.
- **IDROSSICLOROCHINA (HCQ):** terapia indicata per curare le malattie reumatiche e la malaria ma adattata a curare il COVID-19, infatti essa offre un'attività antinfiammatoria e immunomodulatoria.
- **INIBITORI DI PROTEASI HIV:** terapia che blocca la capacità degli enzimi di metabolizzare le proteine, mantenendole in uno stato inattivo. In questo modo viene limitata la capacità riproduttiva delle cellule COVID-19.
- **TERAPIA ANTIBIOTICA:** terapia utilizzata per la cura del COVID-19. Presenta effetti benefici per malattie polmonari infiammatorie e ha un'ottima capacità di inibire la replicazione dei batteri patogeni.
- **EPARINA:** terapia anticoagulante efficace per contrastare le infezioni respiratorie acute dovute dal COVID-19.

## **1.2 Descrizione variabili dipendenti**

Le variabili risposta in studio sono identificate dai sintomi post-Covid riferiti dai pazienti nel tempo e oggettivati dai clinici. I soggetti, infatti, attraverso un colloquio o un questionario scritto, comunicavano il proprio stato di salute permettendone il monitoraggio.

Nel database sono raccolte le informazioni riguardanti la presenza o assenza dei sintomi e, nel primo caso, viene specificata la durata con annessa data di inizio ed eventuale data di fine.

I sintomi post-Covid analizzati sono:

- **DISPNEA:** difficoltà respiratorie.
- **TELOGEN EFFLUVIUM:** perdita temporanea di capelli.



- ASTENIA: stanchezza.
- MIALGIE: dolori muscolari.
- PALPITAZIONI: anomalie del battito cardiaco.
- MANCANZA OLFATTO
- MANCANZA GUSTO
- AMNESIA: perdite di memoria.
- CEFALEA: dolori alla testa/emicranie.
- ANSIA E PANICO
- INSONNIA: difficoltà o anormale brevità del sonno.
- ALTRO: qualsiasi altro sintomo riscontrato non indicato precedentemente.

La fase di analisi si aprirà, nel capitolo successivo, con la pulizia e la selezione delle informazioni presenti in studio.

## Capitolo 2

### **Pre-Processing variabili dipendenti, esplicative e Componenti Principali**

In seguito ad un'analisi esplorativa dei dati e una revisione della letteratura, si è osservato come la maggior parte dei soggetti avessero riscontrato molteplici sintomi post-Covid. Per tale ragione si è deciso di accorparli in macrocategorie. Come metodologia si è scelto di adottare approccio data-driven di riduzione della dimensionalità applicato alle durate dei sintomi. Lavorare con le durate ha permesso di affrontare l'analisi tenendo in considerazione sia la presenza della sintomatologia che l'importanza e il contributo di ciascuna sul fenomeno del Long Covid.

Prima della realizzazione dei “macro-sintomi” è fondamentale e necessario effettuare una pulizia puntuale delle variabili risposta.

#### **2.1 Pre-Processing variabili dipendenti**

Il processo di pulizia delle variabili risposta si è aperto con l'eliminazione della covariata “Altro” e delle sue associate, ovvero durata, data d'inizio e di fine sintomo. Questa, infatti, racchiudeva una serie di sintomatologie specifiche per ogni paziente, non attribuibili con certezza al Long Covid, per questo motivo è stata esclusa dallo studio.

Nei pazienti liberi da sintomi post-Covid la durata della malattia risulta essere assente anziché nulla, per cui si è proceduto ad assegnarle valore zero. Questa conversione ha permesso, quindi, di considerare l'intero spettro dei pazienti in studio che, altrimenti, sarebbe risultato incompleto.

Verificando la correttezza della codifica delle durate, tramite il confronto tra la data di inizio e di fine sintomo, si è notato che, per alcuni pazienti, non è stata riportata la data di fine follow-up. Questo ha reso ignota la durata della sintomatologia e, pertanto, in assenza della variabile risposta, i soggetti sono stati esclusi dall'analisi. Al contrario, le durate che presentavano delle incongruenze, sono state corrette.

In seguito alle modifiche apportate, le variabili dipendenti in analisi sono risultate le seguenti: Durata Dispnea, Durata Telogen Effluvium, Durata Astenia, Durata Mialgie,

Durata Palpitazioni, Durata Mancanza Olfatto, Durata Mancanza Gusto, Durata Amnesia, Durata Cefalea, Durata Ansia e Panico e Durata Insonnia.

## **2.2 Costruzione componenti principali**

Per la creazione dei macro-sintomi è stato adottato un approccio data-driven di riduzione della dimensionalità, in particolare il metodo delle componenti principali.

L'obiettivo consiste nell'individuare delle aree sintomatologiche comuni del Long Covid, in modo da riassumere i sintomi e studiarne i determinanti.

Queste aree sono costituite da variabili fortemente associate tra loro e rappresentano degli indicatori sintetici denominati Macro-Sintomi.

Algebricamente, tale approccio ha permesso di ricavare una nuova rappresentazione dei pazienti affetti da disturbi post-Covid passando da uno spazio  $p$ -dimensionale, rappresentato dalle durate delle sintomatologie, ad uno spazio inferiore  $q$ -dimensionale, rappresentato dalle aree sintomatologiche. La costruzione delle  $p$  componenti principali, ottenute dalle combinazioni lineari delle undici variabili di partenza, ha permesso di tener conto del variare complessivo di tutte le covariate facendo emergere i nessi che vi sono tra sintomi e individui.

Per costruzione l'informazione riprodotta dalle componenti principali, a differenza delle variabili originali, non è equidistribuita. Poiché le prime trasformate permettono di riprodurre la quota maggiore del contenuto informativo totale, estraendo le prime  $q$  componenti è possibile spiegare gran parte della variabilità degli undici sintomi.

La costruzione delle componenti principali è realizzabile in due modalità: tramite la matrice di varianze e covarianze ( $\Sigma$ ), se le variabili originarie sono espresse con la stessa unità di misura e hanno lo stesso ordine di grandezza, o, in alternativa, tramite la matrice di correlazione ( $R$ ), applicabile in caso di problemi di incommensurabilità. Si è deciso pertanto di svolgere le analisi con la matrice  $\Sigma$  al fine di esaltare la magnitudine, l'importanza e i contributi di ogni sintomo nello studio del long-Covid. L'applicazione di questo approccio è stata possibile in quanto, le dipendenti in studio, presentano le medesime unità di misura e ordini di grandezza.

Tabella 1: Percentuale di varianza spiegata dalle prime tre componenti estratte

|                               | <b>1<sup>st</sup> Principal Component</b> | <b>2<sup>nd</sup> Principal Component</b> | <b>3<sup>rd</sup> Principal Component</b> |
|-------------------------------|---|---|---|
| <b>Standard Deviation</b>     | 233,906                                   | 106,239                                   | 946,530                                   |
| <b>Proportion of Variance</b> | 0,4703                                    | 0,0970                                    | 0,0770                                    |
| <b>Cumulative Proportion</b>  | 0,4703                                    | 0,5673                                    | 0,6444                                    |

Per scegliere il numero ottimale di componenti principali da mantenere in analisi, sono stati utilizzati due criteri: quello della percentuale di varianza cumulativa spiegata e quello dello scree-plot.

Combinando queste due regole d'arresto si è deciso di estrarre le prime tre trasformate lineari, incorrelate tra loro, in grado di spiegare complessivamente il 64% della variabilità totale.

Per identificare i macro-sintomi, e quindi ottenere una struttura a blocchi, dove ad ogni variabile sia associata una e una sola trasformata, è stato necessario ruotare le componenti.

Sono stati adottati diversi metodi di rotazione, quelli ortogonali “Varimax” e “Quartimax”, e uno obliquo, “Oblimin”, che in termini interpretativi hanno portato ai medesimi risultati. Successivamente nella trattazione sono stati, quindi, riportati solamente le evidenze ottenute dal metodo “Varimax”, in quanto, rispetto alle altre rotazioni, riportava una struttura a blocchi maggiormente definita.

Attraverso la costruzione della matrice di correlazione tra le trasformate ruotate e le variabili dipendenti in esame, è stato possibile valutare l'associazione tra ogni componente e i singoli disturbi al fine di identificare e denominare i Macro-Sintomi.

Tabella 2: Correlazioni tra variabili e componenti ruotate con metodo varimax

| $\rho(y_i, x_j)$  | 1 <sup>st</sup> Principal Component | 2 <sup>nd</sup> Principal Component | 3 <sup>rd</sup> Principal Component |
|-------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Dispnea           | 0,7110                              | -0,1768                             | -0,5948                             |
| Telogen Effluvium | 0,4903                              | -0,2216                             | -0,5412                             |
| Astenia           | 0,9064                              | -0,2832                             | -0,6583                             |
| Mialgie           | 0,8915                              | -0,3239                             | -0,5999                             |
| Palpitation       | 0,5998                              | -0,2111                             | -0,6683                             |
| Mancanza Olfatto  | 0,3059                              | -0,9454                             | -0,2905                             |
| Mancanza Gusto    | 0,2741                              | -0,9163                             | -0,3450                             |
| Amnesia           | 0,5075                              | -0,2177                             | -0,8480                             |
| Cefalea           | 0,4472                              | -0,2312                             | -0,6794                             |
| Ansia e Panico    | 0,4785                              | -0,2071                             | -0,7116                             |
| Insomnia          | 0,5315                              | -0,2800                             | -0,6009                             |

Osservando la matrice di correlazione tra le componenti ruotate e i singoli sintomi, è possibile notare la presenza di tre gruppi distinti di variabili.

Risulta, quindi, possibile denominare i tre macro-sintomi identificati dall'associazione tra le variabili e le tre componenti estratte:

- il Sintomo Fisico composto da Dispnea, Astenia e Mialgia
- il Sintomo Sensoriale composto da Mancanza dell'Olfatto e del Gusto
- il Sintomo Psicologico composto da Amnesia, Cefalea, Telogen\_effluvium, Palpitazioni, Insomnia e Ansia e panico.

La costruzione dei tre macro-sintomi è stata inoltre supportata dall'interpretazione di un indicatore statistico di consistenza interna, denominato "Alpha di Crombach". Tale indice permette di valutare il grado di coerenza delle variabili che caratterizzano ogni componente. Il suo utilizzo, infatti, ha confermato i risultati della matrice di correlazione riportati nella Tabella 2.

Il metodo delle componenti principali evidenzia una problematica legata al contributo, seppur minimo, delle variabili meno associate nella spiegazione del macro-sintomo. Tale limite impedisce che le tre componenti descrivano solo e soltanto le variabili che le caratterizzano.

Per superare questa problematica è stata pertanto eseguita un'ulteriore analisi delle componenti principali. Questa è stata applicata separatamente ad ogni macro-sintomo individuato in precedenza, comprendendo, però, per ciascuna trasformata, solo e soltanto i sintomi maggiormente correlati alle componenti iniziali (blocchi di variabili in Tabella 2).

Tabella 3: Varianza spiegata dalla prima componente per ogni macro-sintomo

|                              | Componente Fisica | Componente Sensoriale | Componente Psicologica |
|------------------------------|-------------------|-----------------------|------------------------|
| <b>Cumulative Proportion</b> | 0,7467            | 0,8757                | 0,4993                 |

In tabella 3, viene riportata la quota di varianza totale spiegata dalla prima componente principale estratta da ognuna delle tre analisi distinte.

La prima, relativa al Macro-Sintomo Psicologico, spiega il 50% della variabilità totale delle sintomatologie: Amnesia, Cefalea, Telogen\_effluvium, Palpitazioni, Insonnia e Ansia e panico. Per quanto riguarda il Macro-Sintomo Sensoriale, la prima componente spiega l'88% della variabilità di Mancanza di Gusto e Mancanza di Olfatto. Infine, quella Fisica spiega il 75% della variabilità dei sintomi Dispnea, Astenia e Mialgia. Complessivamente le prime componenti relative ad ogni macro-sintomo spiegano il 64,07% del contenuto informativo totale.

Una volta ottenuti i punteggi delle componenti su tutti gli individui, poiché essi presentavano valori negativi, è stato necessario traslarli positivamente in modo da renderli logicamente interpretabili in termini di durate. In questo modo i punteggi delle componenti si distribuiscono in un range tra il valore zero, indicante i soggetti senza sintomi Long Covid, e valori positivi, esprimenti le varie durate.

Tabella 4: Correlazione tra variabili e la prima componente relativa ad ogni macro-sintomo

|                          | Componente Psicologica | Componente Sensoriale | Componente Fisica |
|--------------------------|------------------------|-----------------------|-------------------|
| <b>Dispnea</b>           | 0,5553                 | 0,2153                | 0,7400            |
| <b>Telogen Effluvium</b> | 0,5878                 | 0,2062                | 0,4523            |
| <b>Astenia</b>           | 0,6498                 | 0,2906                | 0,9184            |
| <b>Mialgie</b>           | 0,6261                 | 0,2974                | 0,8779            |
| <b>Palpitation</b>       | 0,7179                 | 0,2258                | 0,5538            |
| <b>Mancanza Olfatto</b>  | 0,2866                 | 0,9461                | 0,3094            |
| <b>Mancanza Gusto</b>    | 0,3272                 | 0,9237                | 0,2884            |
| <b>Amnesia</b>           | 0,7915                 | 0,2488                | 0,5569            |
| <b>Cefalea</b>           | 0,6733                 | 0,2300                | 0,4582            |
| <b>Ansia e Panico</b>    | 0,7225                 | 0,2147                | 0,4756            |
| <b>Insonnia</b>          | 0,6565                 | 0,2444                | 0,4907            |

Osservando la matrice di correlazione tra la prima componente relativa ad ogni macro-sintomo e tutte le undici variabili prese in esame (Tabella 4) si può notare come ogni variabile risulti correlata maggiormente al macro-sintomo di appartenenza.

L'esecuzione di tre analisi delle componenti principali distinte comporta la caduta dell'assunto di incorrelazione tra le trasformate lineari.

Tabella 5. Correlazioni tra componenti estratte

| Correlazioni           | Componente Fisica | Componente Sensoriale | Componente Psicologica |
|------------------------|-------------------|-----------------------|------------------------|
| Componente Fisica      | 1                 | 0.32                  | 0.717                  |
| Componente Sensoriale  |                   | 1                     | 0.326                  |
| Componente Psicologica |                   |                       | 1                      |

Nella tabella 5, infatti, è possibile osservare tra la componente psicologica-sensoriale e quella sensoriale-fisica delle basse correlazioni rispettivamente pari a 0,32 e 0,326. Al contrario, si evidenzia una correlazione elevata tra la componente psicologica e fisica pari a 0,7.

Il valore di correlazione, sopra riportato, sembrerebbe indicare la presenza di un'associazione tra le variabili che compongono la componente fisica e quelle relative alla componente psicologica suggerendo l'esistenza di una sola componente di natura psico-fisica.

L'approfondimento a supporto di tale ipotesi, attraverso la costruzione del grafico del cerchio delle correlazioni tra le componenti e le singole variabili e quello tra le componenti e i singoli individui, hanno permesso di osservare quanto le variabili siano legate alle componenti, quanto siano legate tra di loro e come i soggetti si distribuiscono nello spazio delle trasformate. Tale indagine ha introdotto nuovi quesiti allo studio:

- Esiste un legame tra il sintomo psicologico e quello fisico?
- I soggetti che hanno avuto almeno un sintomo di natura psicologica hanno sviluppato sintomi fisici?
- Sarebbe corretto condurre l'analisi utilizzando una sola componente psico-fisica?

Come scelta metodologica si è deciso, dopo un confronto col personale medico coinvolto, di analizzare i tre macro-sintomi distintamente, ponendo l'attenzione sui quesiti introdotti.

## 2.3 Pre-Processing variabili esplicative

Dopo aver effettuato il pre-processing e la costruzione delle componenti principali sulle variabili dipendenti, è stato necessario eseguire il processo di pulizia anche sulle variabili esplicative.

La prima operazione è stata la ricodifica della covariata “Colore”, indicante il grado di intensità dell’indice Word Health Organization (WHO), convertendo ogni gradazione di colore in un valore numerico da 0 a 3, per ordine crescente di gravità dell’infezione.

Andando ad analizzare la percentuale di valori mancanti su ogni covariata, la colonna “Kg persi durante il ricovero” è stata rimossa in quanto tale variabile presentava una percentuale di missing data superiore al 50%.

Concentrandoci sulle covariate relative alle informazioni personali del paziente si è deciso di controllare e ricalcolare i valori dell’indice di massa corporea (BMI), sfruttando le informazioni sul peso e l’altezza. Successivamente le covariate “Peso” e “Altezza” sono state eliminate in quanto collineari con la variabile “BMI”.

Passando alle variabili riferite ai farmaci, assunti nella fase pre-covid, e alle terapie, somministrate durante la fase acuta del virus, sono state eliminate le covariate “Nessun Farmaco” e “Terapia durante l’infezione”. Queste, infatti, sono risultate ridondanti in quanto esprimibili come combinazioni lineari delle variabili appartenenti ai due rispettivi macro-gruppi. Allo stesso modo la covariata “Almeno un antipertensivo” è apparsa collineare con tutti gli altri farmaci antipertensivi presenti in analisi, in quanto, quasi tutto lo spettro dei medicinali cardiaci, è spiegato dalle variabili appartenenti a tale gruppo. Per questo motivo è stata introdotta una nuova covariata “Altri Antipertensivi” ottenuta dalla differenza fra la variabile “Almeno un antipertensivo” e la somma delle modalità delle altre variabili dello stesso gruppo farmacologico.

Successivamente si è svolta un’operazione di ricodifica delle variabili “O2 terapia”, “O2 terapia ( $\leq 2$  L/min)”, “O2 terapia (6 –2 L/min)” e “O2 terapia ( $> 6$  L/min)”.

La variabile “O2 terapia” che inizialmente si distribuiva in un range da 1 a 6, dove il valore unitario esprimeva il mancato utilizzo di strumenti per la somministrazione di ossigeno, è stata traslata in un range tra 0 e 5. Invece, le variabili indicanti il grado di intensità di somministrazione dell’ossigeno, sono state accorpate in un’unica variabile denominata “O2 intensità” distribuita in un range tra 0 e 3, dove il valore nullo indica la mancata assunzione della terapia. Effettuando un controllo incrociato tra la variabile “O2



intensità” e “O2 terapia” sono stati considerati come missing values i valori sulle due covariate per quei pazienti trattati con la somministrazione dell’ossigeno ma con un livello d’intensità nullo e viceversa.

In seguito ad un controllo sulle modalità di tutte le variabili in esame sono stati corretti i livelli delle covariate “Polmonari”, “Ex-ricovero”, “Metaboliche”, “Patologie Immunologiche”, “Remdesevir” e “Cardiache”, in quanto presentavano degli errori di imputazione.

La fase di pulizia è terminata con l’analisi della zero-variance che ha portato l’eliminazione della variabile “Salrilumab, Baricitinib” poiché degenerare.

## Capitolo 3

### Analisi descrittive

Le analisi descrittive, eseguite a scopo esplorativo e conoscitivo del campione, sono state utilizzate come supporto nello studio dei potenziali nessi esistenti tra macro-sintomi e possibili determinanti. Il campione descritto è costituito da 1368 soggetti suddivisi per ondate infettive, 575 appartenenti alla prima e 793 alla seconda. Successivamente al processo di pulizia si sono mantenute in studio 52 variabili, 11 delle quali riportano le sintomatologie post-Covid.

#### 3.1 Analisi delle sintomatologie per ondate

In primo luogo, è stata effettuata un'analisi grafica al fine di studiare la ripartizione dei soggetti nelle diverse aree sintomatologiche individuate.

Grafico1: Partizionamento del campione per sintomatologie prima ondata

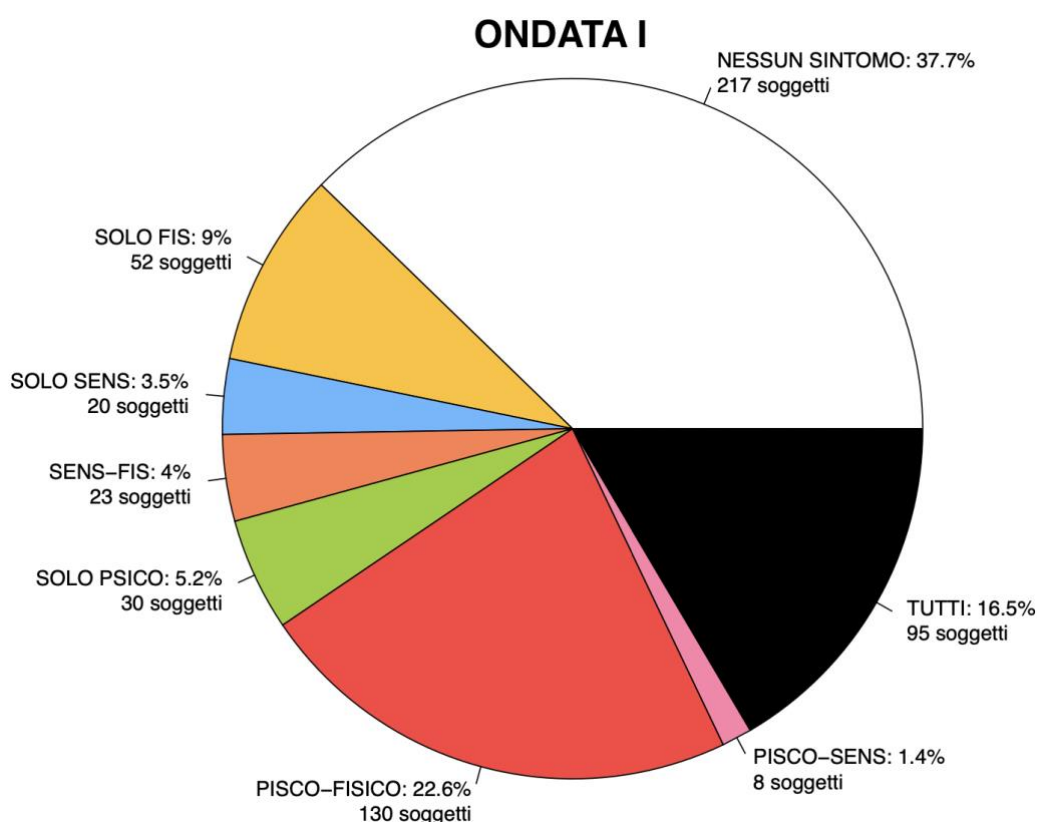
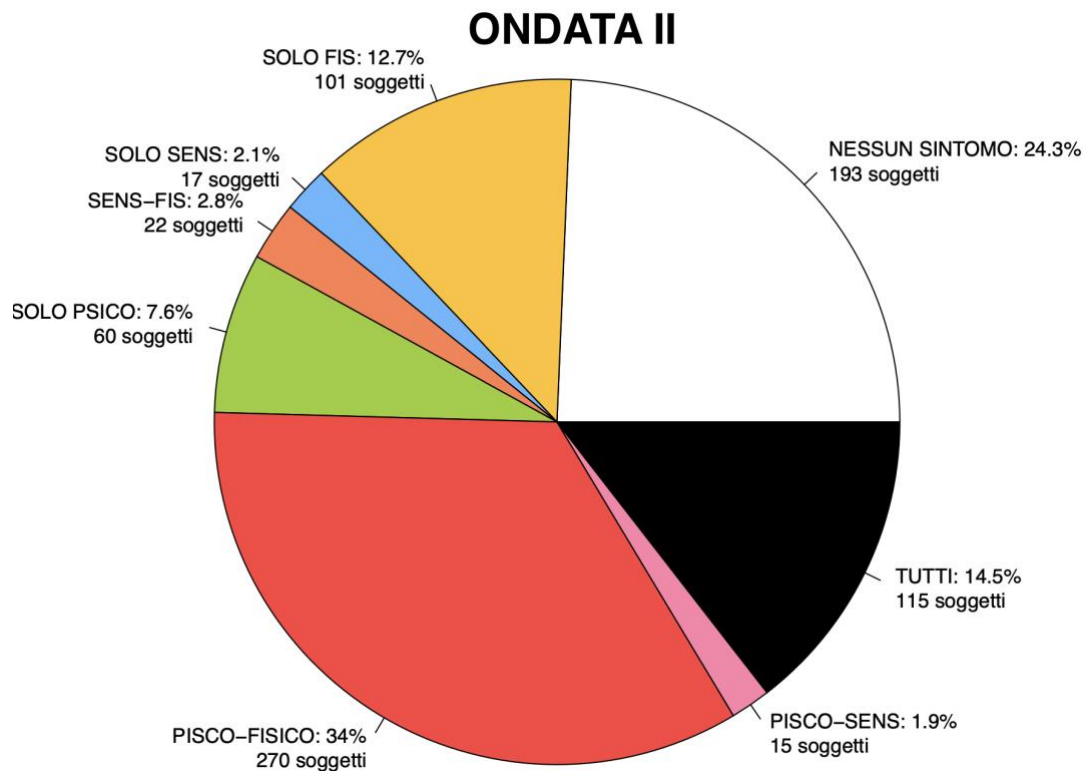


Grafico 2: Partizionamento del campione per sintomatologie seconda ondata



Osservando i grafici risulta evidente come il campione in studio sia costituito, in entrambe le ondate, da due sottopopolazioni: coloro che possiedono almeno un sintomo (porzioni colorate) e coloro che non ne hanno mai sviluppato uno (porzione bianca).

Gran parte della popolazione è ripartita in tre macro-gruppi: gli asintomatici, i pazienti che presentano sintomi di natura psico-fisica e coloro che hanno sviluppato tutti i sintomi in esame.

Si osserva come, passando dalla prima alla seconda ondata, il numero dei soggetti asintomatici sia diminuito del 13% passando dal 37.7% al 24.3%, mentre, la porzione dei soggetti con tutti sintomi è rimasta pressoché invariata; al contrario, il numero di soggetti con sintomatologie psico-fisiche è aumentato del 11%.

### 3.2 Analisi profili dei soggetti per ondate

In secondo luogo, si è analizzato il cambiamento dei profili clinici dei pazienti nelle due ondate Covid. Nelle tabelle sottostanti sono state riportate le sole variabili che, nel confronto, sono risultate statisticamente significative.

Tabella 6: Analisi sintomatologie nelle due ondate infettive

|                          |               | ONDATA 1                | ONDATA 2               | P-VALUE |
|--------------------------|---------------|-------------------------|------------------------|---------|
| TOTALE                   |               | 575                     | 793                    |         |
| DURATA DISPNEA           | NUMEROSITA'   | 172 [29.91 %]           | 294 [37.07 %]          | 0.007   |
|                          | MEDIA (SD)    | 62.6 (132)              | 42.8 (72.8)            |         |
|                          | MEDIANA (IQR) | 185.50 [60.50, 343.75]  | 96.50 [62.00, 158.75]  | 0.1203  |
| DURATA TELOGEN EFFLUVIUM | NUMEROSITA'   | 129 [22.43 %]           | 215 [27.11 %]          | 0.057   |
|                          | MEDIA (SD)    | 46.6 (113)              | 33.4 (65.9)            |         |
|                          | MEDIANA (IQR) | 153.00 [91.00, 330.00]  | 106.00 [69.00, 160.50] | 0.2199  |
| DURATA ASTENIA           | NUMEROSITA'   | 263 [45.74 %]           | 410 [51.70 %]          | 0.034   |
|                          | MEDIA (SD)    | 115 (172)               | 69.8 (88.6)            |         |
|                          | MEDIANA (IQR) | 218.00 [92.00, 387.50]  | 121.00 [73.50, 192.75] | 0.4746  |
| DURATA MIALGIE           | NUMEROSITA'   | 179 [31.13 %]           | 287 [36.19 %]          | 0.059   |
|                          | MEDIA (SD)    | 89.3 (166)              | 48.9 (82.2)            |         |
|                          | MEDIANA (IQR) | 328.00 [112.00, 438.50] | 119.00 [69.50, 192.50] | 0.9373  |
| DURATA PALPITAZIONI      | NUMEROSITA'   | 119 [20.70 %]           | 172 [21.69 %]          | 0.707   |
|                          | MEDIA (SD)    | 54.9 (132)              | 27.5 (63.3)            |         |
|                          | MEDIANA (IQR) | 245.00 [112.00, 390.00] | 107.00 [71.00, 167.75] | 0.6238  |
| DURATA MANCANZA OLFATTO  | NUMEROSITA'   | 131 [22.78 %]           | 151 [19.04 %]          | 0.105   |
|                          | MEDIA (SD)    | 46.8 (120)              | 22.1 (59.4)            |         |
|                          | MEDIANA (IQR) | 114.00 [61.00, 370.00]  | 93.00 [43.50, 182.00]  | 0.03098 |
| DURATA MANCANZA GUSTO    | NUMEROSITA'   | 122 [21.22 %]           | 146 [18.41 %]          | 0.222   |
|                          | MEDIA (SD)    | 38.4 (108)              | 21.5 (61.1)            |         |
|                          | MEDIANA (IQR) | 98.50 [39.50, 361.25]   | 92.00 [43.25, 175.75]  | 0.1251  |
| DURATA AMNESIA           | NUMEROSITA'   | 124 [21.57 %]           | 238 [30.01 %]          | < 0.001 |
|                          | MEDIA (SD)    | 75.4 (160)              | 42.5 (77.5)            |         |
|                          | MEDIANA (IQR) | 367.50 [244.75, 483.50] | 127.00 [83.00, 193.75] | 0.128   |
| DURATA CEFALEA           | NUMEROSITA'   | 87 [15.13 %]            | 124 [15.64 %]          | 0.857   |
|                          | MEDIA (SD)    | 39.7 (116)              | 20.4 (59.8)            |         |
|                          | MEDIANA (IQR) | 264.00 [92.00, 394.00]  | 115.50 [62.00, 186.00] | 0.8127  |
| DURATA ANSIA PANICO      | NUMEROSITA'   | 102 [17.74 %]           | 160 [20.18 %]          | 0.289   |
|                          | MEDIA (SD)    | 48.0 (126)              | 27.0 (64.4)            |         |
|                          | MEDIANA (IQR) | 275.00 [100.25, 426.50] | 125.00 [74.00, 184.00] | 0.6806  |
| DURATA INSONNIA          | NUMEROSITA'   | 71 [12.35 %]            | 198 [24.97 %]          | < 0.001 |
|                          | MEDIA (SD)    | 39.8 (125)              | 31.5 (67.5)            |         |
|                          | MEDIANA (IQR) | 350.00 [145.50, 506.50] | 106.50 [65.00, 165.75] | < 0.001 |

Tabella 7: Analisi terapie somministrate durante la fase infettiva nelle due ondate

|         |                                   | TOTALE                        | ONDATA 1    | ONDATA 2    | P-VALUE |
|---------|-----------------------------------|-------------------------------|-------------|-------------|---------|
| TERAPIE | O2 TERAPIA                        | 0                             | 329 (57.2%) | 368 (46.4%) | < 0.001 |
|         |                                   | NASO CANNULA (1)              | 68 (11.8%)  | 116 (14.6%) |         |
|         |                                   | VENTURI (2)                   | 61 (10.6%)  | 112 (14.1%) |         |
|         |                                   | RESERVOIR (3)                 | 15 (2.6%)   | 12 (1.5%)   |         |
|         |                                   | VENTILAZIONE NON INVASIVA (4) | 63 (11.0%)  | 150 (18.9%) |         |
|         |                                   | IOT (5)                       | 26 (4.5%)   | 25 (3.2%)   |         |
|         | O2 INTENSITÀ                      | 0                             | 325 (56.5%) | 367 (46.3%) | < 0.001 |
|         |                                   | <=2L/min                      | 45 (7.8%)   | 66 (8.3%)   |         |
|         |                                   | 6-2 L/min                     | 29 (5.0%)   | 58 (7.3%)   |         |
|         |                                   | >6 L/min                      | 155 (27.0%) | 287 (36.2%) |         |
|         | STEROIDE                          | 0                             | 497 (86.4%) | 271 (34.2%) | < 0.001 |
|         |                                   | 1                             | 62 (10.8%)  | 519 (65.4%) |         |
|         | REMDESEVIR                        | 0                             | 518 (90.1%) | 663 (83.6%) | < 0.001 |
|         |                                   | 1                             | 39 (6.8%)   | 127 (16.0%) |         |
|         | HCQ                               | 0                             | 340 (59.1%) | 785 (99.0%) | < 0.001 |
|         |                                   | 1                             | 217 (37.7%) | 5 (0.6%)    |         |
|         | INIBITORI PROTEASI HIV            | 0                             | 417 (72.5%) | 789 (99.5%) | < 0.001 |
|         |                                   | LOPINAVIR                     | 134 (23.3%) | 0 (0%)      |         |
|         |                                   | LOPINAVIR E DARUNAVIR         | 4 (0.7%)    | 0 (0%)      |         |
|         |                                   | DARUNAVIR                     | 2 (0.3%)    | 0 (0%)      |         |
|         | TOCILIZUMAB SARILUMAB BARICITINIB | 0                             | 524 (91.1%) | 788 (99.4%) | < 0.001 |
|         |                                   | 1                             | 32 (5.6%)   | 2 (0.3%)    |         |
|         | EPARINA                           | 0                             | 411 (71.5%) | 287 (36.2%) | < 0.001 |
|         |                                   | 1                             | 146 (25.4%) | 502 (63.3%) |         |

Tabella 8: Analisi profili clinici dei pazienti stratificati per ondate

|                          |               | ONDATA 1                | ONDATA 2               | P-VALUE |
|--------------------------|---------------|-------------------------|------------------------|---------|
|                          | TOTALE        | 575                     | 793                    |         |
| DURATA DISPNEA           | NUMEROSITA'   | 172 [29.91 %]           | 294 [37.07 %]          | 0.007   |
|                          | MEDIA (SD)    | 62.6 (132)              | 42.8 (72.8)            |         |
|                          | MEDIANA (IQR) | 185.50 [60.50, 343.75]  | 96.50 [62.00, 158.75]  | 0.1203  |
| DURATA TELOGEN EFFLUVIUM | NUMEROSITA'   | 129 [22.43 %]           | 215 [27.11 %]          | 0.057   |
|                          | MEDIA (SD)    | 46.6 (113)              | 33.4 (65.9)            |         |
|                          | MEDIANA (IQR) | 153.00 [91.00, 330.00]  | 106.00 [69.00, 160.50] | 0.2199  |
| DURATA ASTENIA           | NUMEROSITA'   | 263 [45.74 %]           | 410 [51.70 %]          | 0.034   |
|                          | MEDIA (SD)    | 115 (172)               | 69.8 (88.6)            |         |
|                          | MEDIANA (IQR) | 218.00 [92.00, 387.50]  | 121.00 [73.50, 192.75] | 0.4746  |
| DURATA MIALGIE           | NUMEROSITA'   | 179 [31.13 %]           | 287 [36.19 %]          | 0.059   |
|                          | MEDIA (SD)    | 89.3 (166)              | 48.9 (82.2)            |         |
|                          | MEDIANA (IQR) | 328.00 [112.00, 438.50] | 119.00 [69.50, 192.50] | 0.9373  |
| DURATA PALPITAZIONI      | NUMEROSITA'   | 119 [20.70 %]           | 172 [21.69 %]          | 0.707   |
|                          | MEDIA (SD)    | 54.9 (132)              | 27.5 (63.3)            |         |
|                          | MEDIANA (IQR) | 245.00 [112.00, 390.00] | 107.00 [71.00, 167.75] | 0.6238  |
| DURATA MANCANZA OLFATTO  | NUMEROSITA'   | 131 [22.78 %]           | 151 [19.04 %]          | 0.105   |
|                          | MEDIA (SD)    | 46.8 (120)              | 22.1 (59.4)            |         |
|                          | MEDIANA (IQR) | 114.00 [61.00, 370.00]  | 93.00 [43.50, 182.00]  | 0.03098 |
| DURATA MANCANZA GUSTO    | NUMEROSITA'   | 122 [21.22 %]           | 146 [18.41 %]          | 0.222   |
|                          | MEDIA (SD)    | 38.4 (108)              | 21.5 (61.1)            |         |
|                          | MEDIANA (IQR) | 98.50 [39.50, 361.25]   | 92.00 [43.25, 175.75]  | 0.1251  |
| DURATA AMNESIA           | NUMEROSITA'   | 124 [21.57 %]           | 238 [30.01 %]          | < 0.001 |
|                          | MEDIA (SD)    | 75.4 (160)              | 42.5 (77.5)            |         |
|                          | MEDIANA (IQR) | 367.50 [244.75, 483.50] | 127.00 [83.00, 193.75] | 0.128   |
| DURATA CEFALEA           | NUMEROSITA'   | 87 [15.13 %]            | 124 [15.64 %]          | 0.857   |
|                          | MEDIA (SD)    | 39.7 (116)              | 20.4 (59.8)            |         |
|                          | MEDIANA (IQR) | 264.00 [92.00, 394.00]  | 115.50 [62.00, 186.00] | 0.8127  |
| DURATA ANSIA PANICO      | NUMEROSITA'   | 102 [17.74 %]           | 160 [20.18 %]          | 0.289   |
|                          | MEDIA (SD)    | 48.0 (126)              | 27.0 (64.4)            |         |
|                          | MEDIANA (IQR) | 275.00 [100.25, 426.50] | 125.00 [74.00, 184.00] | 0.6806  |
| DURATA INSONNIA          | NUMEROSITA'   | 71 [12.35 %]            | 198 [24.97 %]          | < 0.001 |
|                          | MEDIA (SD)    | 39.8 (125)              | 31.5 (67.5)            |         |
|                          | MEDIANA (IQR) | 350.00 [145.50, 506.50] | 106.50 [65.00, 165.75] | < 0.001 |

Dalle seguenti analisi descrittive, confrontando le due ondate, si evince come i profili clinici dei soggetti in analisi siano cambiati significativamente a seconda del periodo di infezione del virus.

Nella prima ondata di infezione da SARSCOV2 l'età media dei soggetti che costituiscono il campione risulta essere statisticamente inferiore rispetto a quella dei soggetti della seconda ondata, rispettivamente pari a 53 e 58 anni.

Per quanto riguarda le terapie, si evince che il virus sia stato trattato diversamente nelle due ondate. Osservando le differenze in termini di trattamenti e durate, le terapie adottate nella seconda fase sono apparse più efficaci nonostante la maggior fragilità dei soggetti evidenziata dall'alto numero di farmaci assunti per la cura di patologie cardiache, diabetiche, metaboliche e dall'aumento del numero medio delle Comorbidità. In questa fase, inoltre, si è verificato un aumento dell'intensità del virus e del numero delle

sintomatologie post-Covid, sebbene queste presentino durate ridotte in media di circa la metà. Tale evidenza porta a domandarsi se questi risultati siano dovuti alla maggior cagionevolezza dei pazienti in studio o se le due ondate siano state caratterizzate dalla presenza di un virus di diversa natura. L'ausilio di strumenti analitici inferenziali permetterà nel seguito della trattazione di valutare se la diversità del Long-Covid, nelle due ondate, sia stata determinata dalla natura dell'infezione o da caratteristiche proprie dei soggetti.

## Capitolo 4

### Modelli di Self-Selection

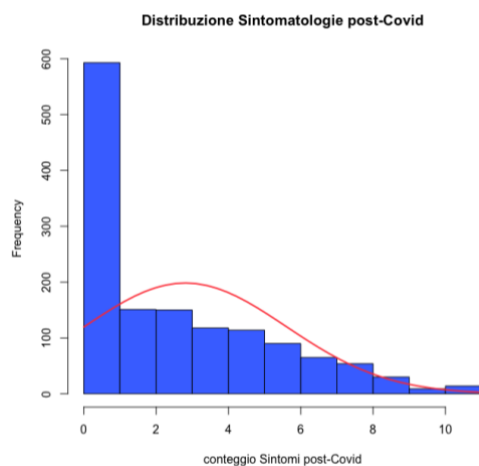
Il campione in studio, come descritto nel primo capitolo della trattazione, è un campione autoselezionato e non casuale. L'autoselezione non garantisce la rappresentatività della popolazione e, se non fosse tenuto in considerazione, i risultati ottenuti porterebbero a delle conclusioni falsate, distorte. Si sarebbe in presenza di una Distorsione da Selezione, un errore introdotto dall'estrazione non randomica degli individui. Affinché ciò non si verifichi è necessaria l'applicazione dei modelli di regressione "Self-Selection", modelli robusti a questo genere di bias. In particolare, per questo studio, sarà utilizzato il Modello Heckit.

#### 4.1 Censored Sample e Correzione di Heckman

Il campione in studio, oltre ad essere un campione auto selezionato, è anche un "Censored Sample". Un Campione Censurato è un insieme di osservazioni il cui valore di outcome è di due differenti tipi, zero se la sintomatologia è assente e positivo viceversa.

Il Modello di Heckman, basato sull'omonima correzione, permette di gestire la distorsione da campioni selezionati in modo non casuale. Concettualmente, permette la stima parallela di due equazioni ed è per questo che è definito anche modello Two-Step. Al contrario, se venissero applicati i modelli tradizionali, la stima dei determinanti potrebbe a risultati distorti e inconsistenti.

Grafico 3: Distribuzione delle sintomatologie dei soggetti in analisi



## 4.2 Modello Heckit

Il modello Heckit, come scritto in precedenza, è un modello Two-Step, composto da due equazioni di stima.

La prima, “Selection Equation”, permette, mediante il modello Probit, di esplicitare la propensione di un soggetto a sviluppare le sintomatologie da post-Covid. L’equazione di selezione è la seguente:

$$z_i^* = \gamma_1 + \sum_{j=2}^r \gamma_j w_j + u_i, \quad i = 1, \dots, N$$

dove  $z_i^*$  è “Attitudine ad ammalarsi di Long Covid”, variabile dipendente dell’analisi,  $w$  le covariate dello studio ed  $N$  la totale ampiezza campionaria. Siccome  $z_i^*$  è una variabile latente e come tale è inosservabile, la variabile dipendente risulta quindi essere l’indicatrice  $z_i$ :

$$z_i = \begin{cases} 1 & z_i^* > 0 \\ 0 & \text{altrimenti} \end{cases}$$

La seconda, “Main Equation”, consente, mediante una regressione OLS e le covariate in analisi, di spiegare le sintomatologie Long Covid. L’equazione principale è quella che segue:

$$y_i = \beta_1 + \sum_{j=2}^r \beta_j x_j + e_i \quad i = 1, \dots, n, \quad N > n$$

dove  $n$  indica la sola parte sintomatica del campione e  $r$  il numero totale delle covariate in analisi.

I residui del modello OLS non sono, però, casuali. Infatti, essi contengono al loro interno una sorta di “variabile omessa” la propensione alla malattia, modellata nella prima equazione. Questo perché  $y_i$  è osservata solo quando  $z_i = 1$  ovvero i pazienti con sintomi sono solo una parte selezionata della totalità dei soggetti. Inoltre, le stime potrebbero soffrire di Selection Bias a causa dell’autoselezione campionaria. Se fosse presente una Distorsione da Selezione allora gli errori delle due equazioni sarebbero correlati, i residui dipenderebbero dal cattivo campionamento, i parametri di regressione sarebbero distorti e inconsistenti.



La correzione di Heckman permette di risolvere questo problema mediante una funzione di regressione condizionata:

$$E(y_i|z_i^* > 0) = \beta_1 + \sum_{j=2}^r \beta_j x_j + \beta_\lambda \lambda_i + e_i^*$$

Dove  $\lambda_i$  è l'Inverse Mills Ratio, la variabile stimata dal modello Probit che, nel modello OLS, consente di tener conto della selezione pazienti sintomatici. L'IMR permette di specificare nella regressione la propensione al Long Covid. Il suo coefficiente  $\beta_\lambda$  è pari a  $\rho\sigma_e$ , dove  $\rho$  esprime proprio l'associazione tra gli errori delle due equazioni. Il valore del parametro  $\rho$  segnala, e con l'ausilio dell'IMR, corregge le stime dalla presenza dell'eventuale Selection Bias. Il modello Heckit permette di ottenere, quindi, delle stime robuste al bias.

Le due equazioni del modello di Heckman sono stimate sotto i tre assunti fondamentali sotto riportati:

- $u_i \sim N(0,1)$
- $e_i \sim N(0,\sigma^2)$
- $corr(u_i, e_i) = \rho$

## Capitolo 5

### **Analisi Preliminari e Model Selection**

Lo studio ha come duplice obiettivo la valutazione dell'effetto delle covariate sulla presenza/assenza dei sintomi post-covid e, se presente, sulla sua intensità. Per tale ragione è stato necessario effettuare delle analisi preliminari specifiche in funzione dello scopo considerato. La trattazione, da questo punto in poi, verrà eseguita in maniera indipendente su ogni macro-sintomo Psicologico e Fisico stratificato per ondate. Il macro-sintomo Sensoriale è stato escluso dall'analisi in quanto clinicamente poco rilevante.

#### **5.1 Analisi Preliminari: studio della Collinearità, Separation/Quasi-Separation e Near-Zero-Variance**

La fase di analisi si apre con lo studio della collinearità eseguito sia sulle esplicative qualitative che su quelle quantitative. Per quanto riguarda il primo gruppo di variabili, la collinearità è stata verificata mediante la metrica del Chi-Quadro normalizzato. Tale indicatore ha segnalato la presenza di una forte associazione tra le variabili “O2 intensità” e “O2 terapia” determinando l'esclusione di quest'ultima. Gli indici TOL e VIF hanno permesso di eseguire questa verifica anche per le variabili quantitative conclusa con il mantenimento di tutte le covariate in esame.

Per studiare l'effetto dei determinanti sull'intero spettro dei pazienti, il controllo della separation e quasi-separation è fondamentale. L'analisi delle tabelle di contingenza relative alle due ondate, per entrambe le Macro-Sintomatologie, ha determinato, infatti, l'eliminazione delle seguenti covariate:

Tabella 9: Covariate mantenute in esame dall'analisi della separation

| MACRO SINTOMO PSICOLOGICO  |  | MACRO SINTOMO FISICO   |   |
|--|--|--|---|
| Prima Ondata   | Seconda Ondata   | Prima Ondata   | Seconda Ondata  |
| TAO NAO<br>Altri Antipertensivi<br>Steroidi Immunologici<br>DMARDs<br>Antitumorali<br>Epatopatia<br>Inibitori Proteasi HIV | Altri Antipertensivi<br>Steroidi Immunologici<br>DMARDs<br>Antitumorali<br>Epatopatia<br>HCQ<br>Tocilizumab Sarilumab<br>Baricitinib<br>Inibitori Proteasi HIV | TAO NAO<br>Altri Antipertensivi<br>Antiaritmici<br>Steroidi Immunologici<br>DMARDs<br>Antitumorali<br>Epatopatia<br>Inibitori Proteasi HIV | Altri Antipertensivi<br>Steroidi Immunologici<br>DMARDs<br>Antitumorali<br>HCQ<br>Tocilizumab<br>Sarilumab<br>Baricitinib<br>Inibitori Proteasi HIV |

Infine, il controllo della near-zero variance ha portato all'eliminazione delle seguenti variabili:

Tabella 10: Covariate mantenute in esame dall'analisi della near-zero-variance

| MACRO SINTOMO PSICOLOGICO  |  | MACRO SINTOMO FISICO   |   |
|--|--|--|---|
| Prima Ondata   | Seconda Ondata   | Prima Ondata   | Seconda Ondata  |
| TAO NAO<br>Altri Antipertensivi<br>Steroidi Immunologici<br>DMARDs<br>Antitumorali<br>Inibitori Proteasi HIV | Altri Antipertensivi<br>Steroidi Immunologici<br>DMARDs<br>Antitumorali<br>HCQ<br>Tocilizumab Sarilumab<br>Baricitinib<br>Inibitori Proteasi HIV | TAO NAO<br>Altri Antipertensivi<br>Steroidi Immunologici<br>DMARDs<br>Antitumorali | Altri Antipertensivi<br>Steroidi Immunologici<br>DMARDs<br>Antitumorali<br>HCQ<br>Tocilizumab<br>Sarilumab<br>Baricitinib<br>Inibitori Proteasi HIV |

## 5.2 Model Selection

La Model Selection è una procedura statistica che permette di validare quale sia la forma funzionale maggiormente adeguata a spiegare il fenomeno in studio. Essa è basata su diversi criteri di penalizzazione in grado di eseguire la cosiddetta "Variable Selection". La Selezione delle Variabili consiste nel comparare diversi modelli, differenti solamente

per numero di covariate, al fine di decretare il numero ottimale di esplicative da mantenere in esame. Tale processo è stato eseguito per entrambe le equazioni, Probit e OLS, del modello Heckit.

I criteri di penalizzazione, o anche chiamati criteri d'informazione, considerano congiuntamente la bontà d'adattamento del modello ai dati e la parsimonia dello stesso bilanciandone la complessità, numero di parametri, e la capacità esplicativa, adattamento. I criteri utilizzati nell'analisi sono quello Bayesiano proposto da Schwarz, BIC, e AIC suggerito da Akaike.

Il Bayesian Information Criterion BIC è definito nel modo seguente:

$$BIC = -2\hat{l}_k(\theta) + \#par_k \log(n)$$

dove  $\hat{l}_k$  è il valore della log-verosimiglianza del modello, nel punto di massimo, con k di variabili esplicative penalizzato per il numero dei parametri del modello ( $\#par_k$ ) rispetto al logaritmo di n, quantità delle osservazioni disponibili. Un valore  $\hat{l}_k$  elevato esprime, da parte del modello in esame, una notevole capacità esplicativa, trascurabile sarà la quota di varianza residua. Siccome il coefficiente della verosimiglianza è negativo i modelli più performanti saranno quelli aventi un valore BIC molto basso. Se l'aggiunta di un parametro incrementa di poco il valore della verosimiglianza, il termine BIC potrebbe peggiorare a causa del termine additivo di penalità. In questo modo l'indice selezionerà il modello più parsimonioso al netto della capacità esplicativa. Il criterio di informazione di Akaike, invece, è il seguente:

$$AIC = -2\hat{l}_k(\theta) + 2par_k.$$

Rispetto al criterio BIC questo non penalizza per il numero di osservazioni ma per il doppio del numero di parametri del modello. In generale, non sempre i due metodi portano alle stesse conclusioni, solitamente il criterio BIC è quello più parsimonioso. Anche in questo caso, la scelta del modello migliore cade su quello che presenta il valore più basso di tale indice.

A supporto dei criteri tradizionali di Variable Selection è stato introdotto il principio di Variable Importance degli Alberi Decisionali. Per Importanza delle Variabili si intende la quota di variabilità totale, della variabile dipendente, spiegata dalle covariate esplicative passando dai nodi padre ai nodi figli. Questa viene valutata mediante il decremento dell'eterogeneità di Gini, in caso di target binario, e della varianza, in caso di target continuo.

L'unione dei metodi descritti ha portato la selezione delle seguenti variabili:

Tabella 11: Covariate Macro-Sintomo Psicologico selezionate per il modello Heckit

| MACRO SINTOMO PSICOLOGICO   |   |   |  |
|---|---|---|--|
| Prima Ondata  |   | Seconda Ondata  |  |
| Modello Probit  | Modello OLS   | Modello Probit  | Modello OLS  |
| Sesso<br>Ipolepizzante Statina<br>Ca antagonisti<br>Beta Bloccanti<br>ACE inibitori Sartani<br>Steroidi Inalatori<br>Neuro Psicoattivi<br>Renali<br>Numero Comorbidità<br>HCQ<br>BMI<br>Età<br>O2intensità<br>Intensità | Intensità<br>Sesso<br>Ipolepizzante Statina<br>Ca antagonisti<br>HCQ<br>Tocilizumab Sarilumab<br>Baricitinib<br>ACE inibitori Sartani<br>O2intensità<br>Età<br>BMI<br>Steroide<br>Eparina | Sesso<br>Ipolepizzante Statina<br>Steroidi inalatori<br>Steroide<br>Neuro Psicoattivi<br>Diabete<br>Patologie<br>Immunologiche<br>Remdesevir<br>Antibiotico<br>BMI<br>Età<br>Intensità<br>Eparina | Età<br>Sesso<br>BMI<br>ACE inibitori Sartani<br>Steroidi Inalatori<br>Patologie<br>Immunologiche<br>Steroide<br>O2intensità<br>Intensità<br>Eparina<br>Numero Comorbidità<br>Cardiache<br>Antiaggregante |

Tabella 12: Covariate Macro-Sintomo Fisico selezionate per il modello Heckit

| MACRO SINTOMO FISICO  |   |  |   |
|---|---|--|---|
| Prima Ondata  |   | Seconda Ondata   |   |
| Modello Probit  | Modello OLS   | Modello Probit   | Modello OLS   |
| Intensità<br>Età<br>Sesso<br>BMI<br>Antiaggregante<br>Beta Bloccanti<br>Steroidi Inalatori<br>Metaboliche<br>Numero Comorbidità<br>HCQ<br>O2intensità | Intensità<br>Sesso<br>Ipolepizzante Statina<br>Diuretici<br>Steroidi Inalatori<br>Numero Comorbidità<br>HCQ<br>Tocilizumab Sarilumab<br>Baricitinib<br>Eparina<br>BMI<br>O2intensità<br>Remdesevir<br>Età | Sesso<br>Diuretici<br>Polmonari<br>Patologie Autoimmuni<br>Steroide<br>Remdesevir<br>Età<br>Numero Comorbidità<br>Intensità<br>BMI | Sesso<br>BMI<br>Ipolepizzante Statina<br>Polmonari<br>Patologie<br>Immunologiche<br>Antibiotico<br>Steroide<br>Numero Comorbidità<br>O2intensità<br>Eparina |

### 5.3 Studio delle Interazioni

Oltre alle variabili selezionate dalla Model Selection, sono stati inseriti, all'interno delle equazioni del modello Heckit, dei termini di interazione. Questi derivano da un attento e approfondito studio della letteratura scientifica farmacologica riguardante principi attivi, la loro interazione, la posologia, gli effetti indesiderati e le patologie a loro connessi.

Dalla ricerca bibliografica e rispetto le variabili mantenute in analisi, i caratteri interattivi presenti in studio sono i seguenti:

Tabella 13: Interazioni selezionate per i Macro-Sintomo Psicologico del modello Heckit

| MACRO SINTOMO PSICOLOGICO  |   |   |   |
|--|---|---|---|
| Prima Ondata   |   | Seconda Ondata  |   |
| Modello Probit   | Modello OLS                             | Modello Probit  | Modello OLS   |
| Ca Antagonisti*Beta<br>Bloccanti<br>Ca Antagonisti *ACE<br>Inibitori Sartani<br>Ca Antagonisti<br>*Neuro_Psicoattivi<br>Beta Bloccanti * ACE<br>Inibitori Sartani<br>Beta Bloccanti<br>*Steroidi Inalatori<br>ACE Inibitori Sartani<br>*Renali<br>ACE Inibitori Sartani<br>*Steroidi Inalatori | Ca Antagonisti*ACE<br>Inibitori Sartani | Steroidi<br>Inalatori*Diabete<br>Steroidi<br>Inalatori*Neuro<br>Psicoattivi | ACE Inibitori Sartani<br>*Antiaggregante<br>ACE Inibitori Sartani<br>*Steroidi Inalatori<br>Steroidi<br>Inalatori*Antiaggregante<br>Steroidi<br>Inalatori*Cardiache |

Tabella 14: Interazioni selezionate per i Macro-Sintomo Fisico del modello Heckit

| MACRO SINTOMO FISICO  |                     |                     |                     |
|---|---------------------|---------------------|---------------------|
| Prima Ondata  |                     | Seconda Ondata      |                     |
| Modello Probit  | Modello OLS         | Modello Probit      | Modello OLS         |
| Antiaggregante*Beta<br>Bloccanti<br>Beta<br>Bloccanti*Metaboliche | Nessuna Interazione | Nessuna Interazione | Nessuna Interazione |

## Capitolo 6

### Analisi dei risultati

Lo studio del Long Covid è stato possibile, come ampiamente descritto nei paragrafi precedenti, tramite l'applicazione dei modelli di regressione Heckit. Questi hanno permesso di identificare e valutare l'impatto dei determinanti sintomatologici sulla durata dei disturbi post-Covid. In particolare, l'equazione Probit ha permesso di analizzare le relazioni esistenti tra le variabili selezionate e la propensione a sviluppare i sintomi, mentre quella OLS ha consentito di valutare le associazioni presenti tra le covariate e i soli soggetti affetti da tali disturbi. Inoltre, importante segnalare che, tutte variabili, aventi un valore di significatività limite, non sono state escluse dallo studio. Questo perché, trattandosi di un'analisi esplorativa, l'obiettivo era quello di segnalare, mediante delle prime evidenze scientifiche, tutti i possibili determinanti sintomatologici. Evidenze che potranno essere certificate e confermate solo da uno studio maggiormente accurato con un campione costituito ad hoc.

I risultati delle analisi, indicati in tabella, saranno illustrati ampiamente nei prossimi paragrafi. Questi verranno letti sia longitudinalmente, evidenziando, all'interno di ogni Macro-Sintomo, le differenze tra le ondate infettive, che trasversalmente risaltando le discrepanze tra le sintomatologie di ogni variante.

Prima di proseguire nella trattazione è importante segnalare che, osservando la tabella, i coefficienti di correlazione  $\rho$  sono, in valore assoluto, molto bassi. Questo ad indicare che l'autoselezione campionaria non ha determinato alcun selection bias. Il campione in esame si dimostra essere rappresentativo, i risultati delle analisi sono generalizzabili all'intera popolazione in studio. Trascurabili anche gli effetti degli Inverse Mills Ratio, le variabili stimate dal modello Probit che, nei modelli OLS, consentono di correggere le stime tenendo conto della selezione dei pazienti sintomatici. Il loro coefficiente non è mai risultato significativo.

Tabella 15: Output modelli di Self Selection Heckit

| MACRO-SINTOMO PSICOLOGICO   |          |            |         |          |          |   |                           |         |         |          |          |       |     |
|---|----------|------------|---------|----------|----------|---|---------------------------|---------|---------|----------|----------|-------|-----|
| PROBIT EQUATION   |          |            |         |          |          |   |                           |         |         |          |          |       |     |
| PRIMA ONDATA  |          |            |         |          |          | SECONDA ONDATA  |                           |         |         |          |          |       |     |
| Tobit 2 model (sample selection model)<br>2-step Heckman/heckit estimation<br>434 observations (199 censored and 235 observed)<br>19 free parameters (df = 416) |          |            |         |          |          | Tobit 2 model (sample selection model)<br>2-step Heckman/heckit estimation<br>738 observations (307 censored and 431 observed)<br>20 free parameters (df = 719) |                           |         |         |          |          |       |     |
|   | Estimate | Std.       | t-value | Pr(> t ) |          |   | Estimate                  | Std.    | t-value | Pr(> t ) |          |       |     |
| (Intercept)   | 0,18     | 0,11       | 1,69    | 0,091    | .        | (Intercept)   | 0,31                      | 0,09    | 3,44    | 0,001    | ***      |       |     |
| Sesso: Maschi   | -0,61    | 0,13       | -4,61   | 0,000    | ***      | Sesso: Maschi   | -0,71                     | 0,10    | -6,87   | <0,001   | ***      |       |     |
| SteroidI<br>Inalatori1  | 0,88     | 0,41       | 2,18    | 0,030    | *        | Steroidi<br>Inalatori1  | 0,51                      | 0,30    | 1,69    | 0,091    | .        |       |     |
| Neuro<br>Psicoattivi1   | -0,41    | 0,26       | -1,54   | 0,124    |          | Neuro<br>Psicoattivi1   | 0,26                      | 0,16    | 1,61    | 0,109    |          |       |     |
| Ipolipemizzante<br>Statina1   | -0,72    | 0,22       | -3,24   | 0,001    | **       | Steroide1   | 0,31                      | 0,12    | 2,66    | 0,008    | **       |       |     |
| Calcio<br>Antagonisti1  | -0,52    | 0,28       | -1,87   | 0,062    | .        | Diabete1  | -0,31                     | 0,16    | -1,98   | 0,048    | *        |       |     |
| B_Bloccanti1  | 0,77     | 0,32       | 2,37    | 0,018    | *        | Patologie<br>Immuno1  | -0,78                     | 0,21    | -3,64   | 0,000    | ***      |       |     |
| ACE_inibitori<br>Sartani1   | 0,58     | 0,20       | 2,83    | 0,005    | **       | Remdesevir1   | 0,23                      | 0,14    | 1,63    | 0,103    |          |       |     |
| Renali1   | -0,67    | 0,40       | -1,66   | 0,098    | .        | Antibiotico1  | 0,19                      | 0,11    | 1,76    | 0,079    | .        |       |     |
| HCQ1  | 0,57     | 0,13       | 4,21    | <0,001   | ***      |   |                           |         |         |          |          |       |     |
| B_Bloccanti1*<br>ACE_inibitori<br>Sartani1  | -0,72    | 0,46       | -1,55   | 0,123    |          |   |                           |         |         |          |          |       |     |
| MAIN EQUATION   |          |            |         |          |          |   |                           |         |         |          |          |       |     |
| PRIMA ONDATA  |          |            |         |          |          | SECONDA ONDATA  |                           |         |         |          |          |       |     |
|   | Estimate | In<br>Days | Std.    | t-value  | Pr(> t ) |   | Estimate                  | In Days | Std.    | t-value  | Pr(> t ) |       |     |
| (Intercept)   | 457,59   | 1098,22    | 85,47   | 5,35     | 0,000    | ***   | (Intercept)               | 166,76  | 400,22  | 44,79    | 3,72     | 0,000 | *** |
| Sesso: Maschi   | -66,75   | -160,20    | 40,50   | -1,65    | 0,100    | .   | Sesso: Maschi             | -40,67  | -97,62  | 18,77    | -2,17    | 0,031 | *   |
| Età   | -2,12    | -5,09      | 1,27    | -1,67    | 0,096    | .   | Età                       | -0,79   | -1,89   | 0,44     | -1,79    | 0,074 | .   |
| Tocilizumab<br>Sarilumab<br>Baricitinib1  | -107,65  | -258,36    | 72,62   | -1,48    | 0,139    |   | BMI                       | 2,28    | 5,46    | 1,22     | 1,86     | 0,063 | .   |
| Eparina1  | 70,00    | 168,00     | 38,86   | 1,80     | 0,072    | .   | ACE_inibitori<br>Sartani1 | -24,76  | -59,43  | 13,87    | -1,79    | 0,075 | .   |
|   |          |            |         |          |          |   | O2intensità:<br>≤2L/min   | -14,80  | -35,52  | 20,78    | -0,71    | 0,476 |     |
|   |          |            |         |          |          |   | O2intensità:<br>2-6 L/min | -387,51 | -930,03 | 219,00   | -1,77    | 0,077 | .   |
|   |          |            |         |          |          |   | O2intensità:<br>>6 L/min  | -102,30 | -245,52 | 155,59   | -0,66    | 0,511 |     |
|   |          |            |         |          |          |   |                           |         |         |          |          |       |     |
| Multiple R-Squared: 0.0571, Adjusted R-Squared: 0.0365  |          |            |         |          |          | Multiple R-Squared: 0.0968, Adjusted R-Squared: 0.0796  |                           |         |         |          |          |       |     |
| DIAGNOSTICS   |          |            |         |          |          |   |                           |         |         |          |          |       |     |
| PRIMA ONDATA  |          |            |         |          |          | SECONDA ONDATA  |                           |         |         |          |          |       |     |
|   | Estimate | Std.       | t-value | Pr(> t ) |          |   | Estimate                  | Std.    | t-value | Pr(> t ) |          |       |     |
| invMillsRatio   | -54,59   | 71,86      | -0,76   | 0,45     |          | invMillsRatio   | -39,79                    | 34,74   | -1,15   | 0,25     |          |       |     |
| sigma   | 259,61   | NA         | NA      | NA       |          | sigma   | 115,01                    | NA      | NA      | NA       |          |       |     |
| rho   | -0,21    | NA         | NA      | NA       |          | rho   | -0,35                     | NA      | NA      | NA       |          |       |     |



| MACRO-SINTOMO FISICO  |          |         |         |          |          |   |                        |         |         |          |          |       |    |
|---|----------|---------|---------|----------|----------|---|------------------------|---------|---------|----------|----------|-------|----|
| PROBIT EQUATION   |          |         |         |          |          |   |                        |         |         |          |          |       |    |
| PRIMA ONDATA  |          |         |         |          |          | SECONDA ONDATA  |                        |         |         |          |          |       |    |
| Tobit 2 model (sample selection model)<br>2-step Heckman/heckit estimation<br>434 observations (161 censored and 273 observed)<br>24 free parameters (df = 411) |          |         |         |          |          | Tobit 2 model (sample selection model)<br>2-step Heckman/heckit estimation<br>738 observations (264 censored and 474 observed)<br>17 free parameters (df = 722) |                        |         |         |          |          |       |    |
|   | Estimate | Std.    | t-value | Pr(> t ) |          |   | Estimate               | Std.    | t-value | Pr(> t ) |          |       |    |
| (Intercept)   | 0,05     | 0,31    | 0,17    | 0,862    |          | (Intercept)   | 0,25                   | 0,09    | 2,78    | 0,006    | **       |       |    |
| Sesso: Maschi   | -0,32    | 0,14    | -2,36   | 0,019    | *        | Sesso: Maschi   | -0,48                  | 0,10    | -4,70   | <0,001   | ***      |       |    |
| Intensità: Media  | -0,43    | 0,21    | -2,04   | 0,042    | *        | Diuretici1  | 0,27                   | 0,16    | 1,64    | 0,101    |          |       |    |
| Intensità: Media-Alta   | -0,13    | 0,36    | -0,37   | 0,713    |          | Polmonari1  | 0,35                   | 0,16    | 2,12    | 0,034    | *        |       |    |
| Intensità: Alta   | -0,05    | 0,38    | -0,13   | 0,898    |          | Patologie Autoimmuni1   | 0,85                   | 0,32    | 2,67    | 0,008    | **       |       |    |
| Età   | 0,01     | 0,00    | 2,77    | 0,006    | **       | Steroidi1   | 0,36                   | 0,11    | 3,31    | 0,001    | ***      |       |    |
| Antiaggregante1   | -0,68    | 0,26    | -2,65   | 0,008    | **       | Remdesevir1   | 0,28                   | 0,14    | 1,93    | 0,054    | .        |       |    |
| B_Bloccanti1  | 0,57     | 0,26    | 2,19    | 0,029    | *        |   |                        |         |         |          |          |       |    |
| Steroidi Inalatori1   | 0,94     | 0,43    | 2,19    | 0,029    | *        |   |                        |         |         |          |          |       |    |
| Metaboliche1  | 0,55     | 0,22    | 2,56    | 0,011    | *        |   |                        |         |         |          |          |       |    |
| N_Comorbidità   | -0,13    | 0,06    | -2,24   | 0,026    | *        |   |                        |         |         |          |          |       |    |
| HCQ1  | 0,29     | 0,17    | 1,72    | 0,086    | .        |   |                        |         |         |          |          |       |    |
| O2intensità: <=2L/min   | -0,11    | 0,30    | -0,35   | 0,723    |          |   |                        |         |         |          |          |       |    |
| O2intensità: 2-6 L/min  | 0,67     | 0,40    | 1,65    | 0,100    | .        |   |                        |         |         |          |          |       |    |
| O2intensità: >6 L/min   | 0,14     | 0,39    | 0,35    | 0,730    |          |   |                        |         |         |          |          |       |    |
| MAIN EQUATION   |          |         |         |          |          |   |                        |         |         |          |          |       |    |
| PRIMA ONDATA  |          |         |         |          |          | SECONDA ONDATA  |                        |         |         |          |          |       |    |
|   | Estimate | In Days | Std.    | t-value  | Pr(> t ) |   | Estimate               | In Days | Std.    | t-value  | Pr(> t ) |       |    |
| (Intercept)   | 249,68   | 424,46  | 46,99   | 5,31     | 0,000    | ***   | (Intercept)            | 120,87  | 205,48  | 39,00    | 3,10     | 0,002 | ** |
| Sesso: Maschi   | -69,38   | -       | 30,89   | -2,25    | 0,025    | *   | Sesso: Maschi          | -37,38  | -63,54  | 14,89    | -2,51    | 0,012 | *  |
| N_Comorbidità   | 29,47    | 50,10   | 11,31   | 2,61     | 0,009    | **  | N_Comorbidità          | -6,36   | -10,82  | 2,88     | -2,21    | 0,027 | *  |
| Ipolipemizzante Statina1  | -98,82   | -       | 51,87   | -1,91    | 0,057    | .   | BMI                    | 3,52    | 5,99    | 1,18     | 2,98     | 0,003 | ** |
| Diuretici1  | -123,90  | -       | 75,70   | -1,64    | 0,102    |   | O2intensità: <=2L/min  | -30,00  | -51,00  | 19,90    | -1,51    | 0,132 |    |
| Eparina1  | 69,55    | 118,24  | 34,61   | 2,01     | 0,045    | *   | O2intensità: 2-6 L/min | -37,26  | -63,34  | 20,64    | -1,81    | 0,071 | .  |
|   |          |         |         |          |          |   | O2intensità: >6 L/min  | -18,50  | -31,44  | 14,21    | -1,30    | 0,194 |    |
|   |          |         |         |          |          |   |                        |         |         |          |          |       |    |
| Multiple R-Squared: 0.0742, Adjusted R-Squared: 0.0533  |          |         |         |          |          | Multiple R-Squared: 0.0741, Adjusted R-Squared: 0.0602  |                        |         |         |          |          |       |    |
| DIAGNOSTICS   |          |         |         |          |          |   |                        |         |         |          |          |       |    |
| PRIMA ONDATA  |          |         |         |          |          | SECONDA ONDATA  |                        |         |         |          |          |       |    |
|   | Estimate | Std.    | t-value | Pr(> t ) |          |   | Estimate               | Std.    | t-value | Pr(> t ) |          |       |    |
| invMillsRatio   | 47,38    | 60,16   | 0,79    | 0,43     |          | invMillsRatio   | -32,91                 | 39,85   | -0,83   | 0,41     |          |       |    |
| sigma   | 248,51   | NA      | NA      | NA       |          | sigma   | 113,81                 | NA      | NA      | NA       |          |       |    |
| rho   | 0,19     | NA      | NA      | NA       |          | rho   | -0,29                  | NA      | NA      | NA       |          |       |    |

## 6.1 Interpretazione dei punteggi dei Macro-Sintomi

I Macro-Sintomi sono stati costituiti mediante la tecnica delle Componenti Principali. Un limite della metodologia, applicata all'inferenza, è l'interpretabilità. Infatti, queste, essendo delle combinazioni lineari pesate delle variabili che le compongono, non sono più esprimibili nell'unità di misura e ordine di grandezza delle covariate originali.

Grafico 4: Correlazione Punteggi Componente Macro-Sintomo Fisico e somma durate singoli sintomi che lo compongono

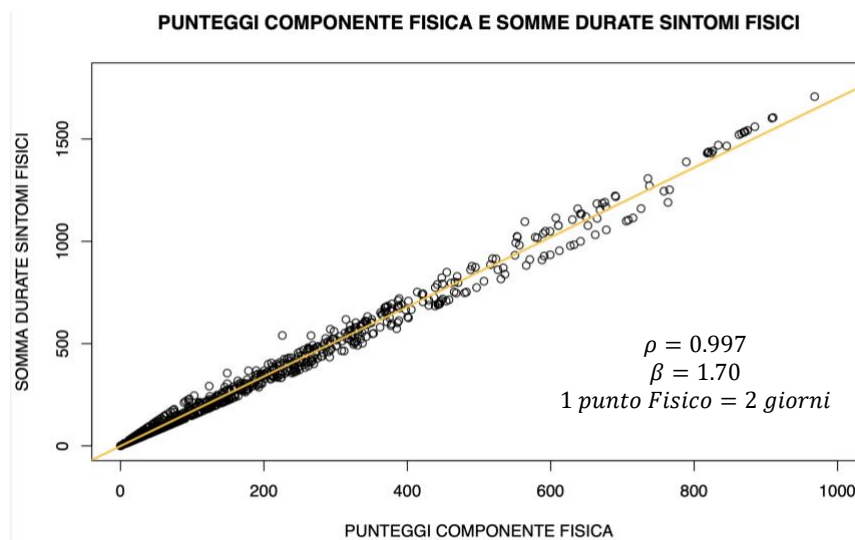
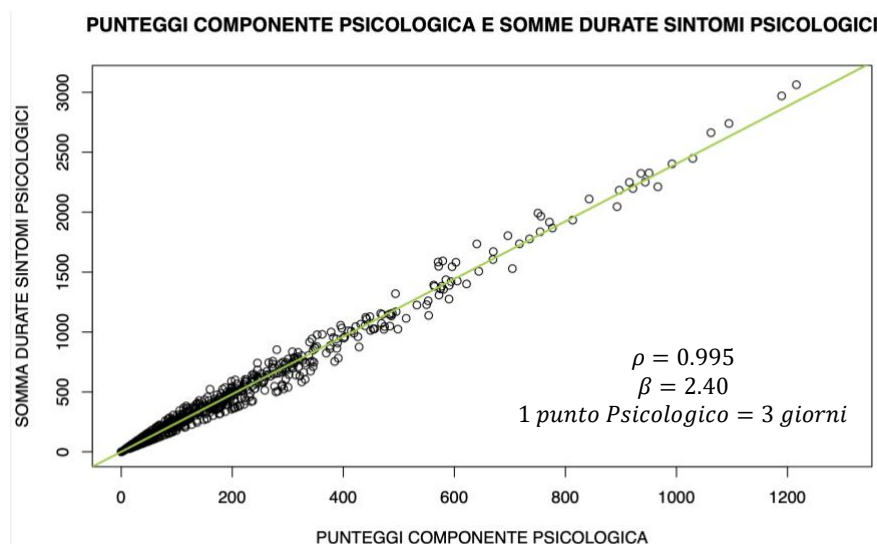


Grafico 5: Correlazione Punteggi Componente Macro-Sintomo Psicologico e somma durate singoli sintomi che lo compongono



Siccome ogni Macro-Sintomatologia è composta da una somma pesata dei singoli disturbi, regredendo il totale dei sintomi, in termini di durate, con il relativo punteggio di componente si ricava il parametro  $\beta$ . Questo ha permesso di convertire le affezioni psicologiche e fisiche interpretandole, come riportano le legende dei grafici 9 e 10, in

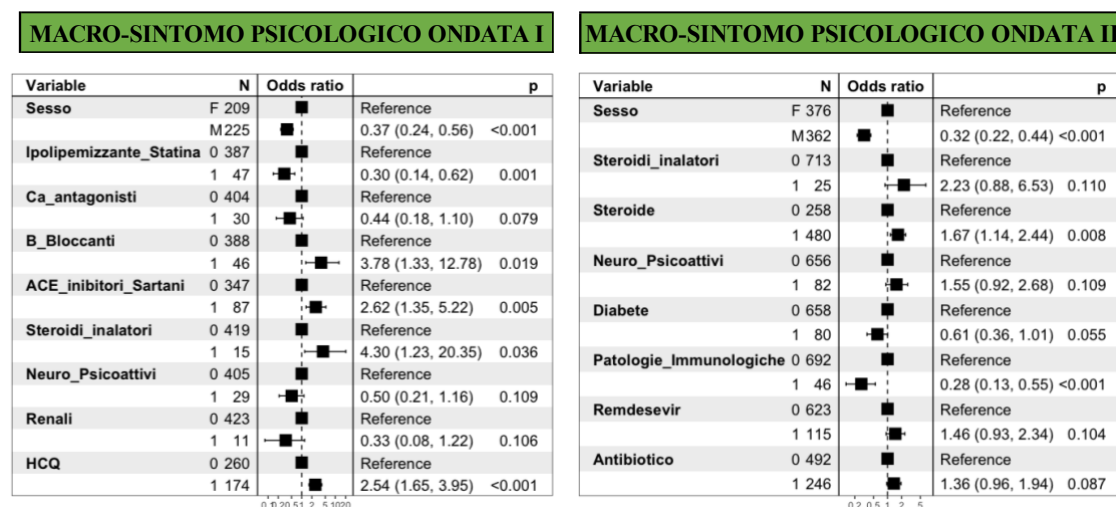
giorni. Ovviamente, per ogni soggetto, il punteggio convertito esprime la durata complessiva dei sintomi che lo compongono, ovvero la somma delle durate dei disturbi riscontrati dal paziente.

## 6.2 Analisi Macro-Sintomo Psicologico

I risultati dei modelli Probit e Logit, in termini di coefficienti di regressione, sono molto simili, quindi, per semplicità interpretativa, sono stati esposti quelli del modello Logistico, che esponenziati sono pari agli Odds Ratio (OR).

L'OR, rapporto tra Odds, è un indicatore che permette di valutare se il fattore in esame è un determinante della malattia e di segnalare se ne aumenta, o diminuisce, la probabilità di contrarla. Se questo assume un valore superiore all'unità, la presenza del fattore aumenta il rischio d'insorgenza della malattia, se assume un valore inferiore all'unità, è un fattore protettivo. Se invece è pari ad uno non esiste alcuna associazione e la variabile in esame non è un determinante sintomatologico.

Grafico 6: Odds ratio ricavati dalle stime del modello Logit del Macro-Sintomo Psicologico



Gli odds ratio evidenziano come i determinanti del sintomo Psicologico, eccezion fatta per “Sesso”, “Steroidi Inalatori” e “Neuro Psicoattivi”, sono molto diversi nelle due varianti. Analizzando i risultati è possibile affermare che le donne, in entrambe le ondate, hanno una propensione maggiore, rispetto gli uomini, a soffrire di disturbi psicologici post-Covid. Gli Steroidi Inalatori, come i farmaci Beta Bloccanti e ACE Inibitori Sartani, rappresentano un forte fattore di rischio. Anche il trattamento HCQ sembrerebbe aumentare le probabilità di riscontrare problemi psicologici. Al contrario, pazienti affetti da patologie Renali, utilizzatori di Calcio Antagonisti e Ipolipemizzante Statina

presentano, nella prima ondata, una minor probabilità di sviluppare la sintomatologia post virus. Infatti, queste, diminuendo l'attitudine alla malattia del 67%, 56% e 70%, rappresentano dei fattori protettivi.

I farmaci Neuro Psicoattivi, nelle due ondate infettive, giocano un ruolo differente. Nella prima rappresentano un fattore protettivo, abbattendo il rischio di malattia del 50%, mentre nella seconda lo aumentano del 55%. Questo, unito alle differenze, in termini di determinanti, tra le due varianti, sembra evidenziare che l'infezione abbia subito una mutazione tra le due ondate. Analizzando i determinanti della seconda ondata si osserva che le terapie Antibiotico, Remdesevir e Steroide aumentano il rischio di malattia, mentre i pazienti affetti da Diabete e Patologie Immunologiche riscontrano il 39% e il 72% in meno di probabilità di sviluppare le sintomatologie psicologiche.

I coefficienti lineari dei modelli OLS esprimono l'effetto delle covariate sulla variabile dipendente al netto delle altre, in termini di durate. Essi rappresentano, in caso di covariata continua, il contributo dell'incremento unitario mentre, in caso di covariata qualitativa, l'impatto del passaggio da un livello all'altro della variabile categoriale. Esprimono, dunque, l'effetto diretto del fattore sulla durata della sintomatologia in studio.

Tabella 16: Stime del modello OLS per il Macro-Sintomo Psicologico

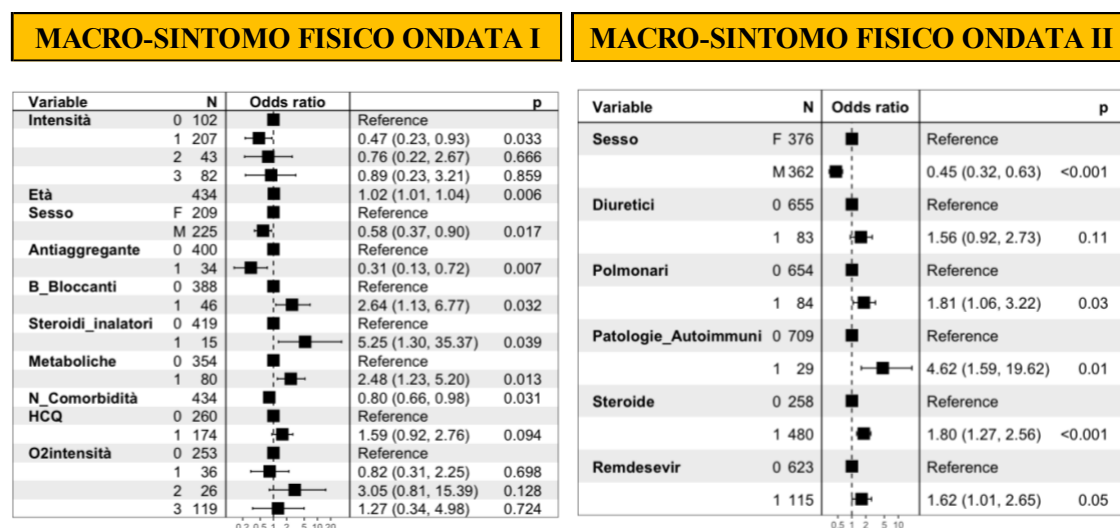
| MACRO-SINTOMO PSICOLOGICO                              |          |         |         |          |          |     |  |          |         |         |          |          |     |
|--|----------|---------|---------|----------|----------|-----|--|----------|---------|---------|----------|----------|-----|
| MAIN EQUATION  |          |         |         |          |          |     |  |          |         |         |          |          |     |
| PRIMA ONDATA   |          |         |         |          |          |     | SECONDA ONDATA   |          |         |         |          |          |     |
|  | Estimate | In Days | Std.    | t-value  | Pr(> t ) |     |  | Estimate | In Days | Std.    | t-value  | Pr(> t ) |     |
| (Intercept)  | 457,59   | 1098,22 | 85,47   | 5,35     | 0,000    | *** | (Intercept)  | 166,76   | 400,22  | 44,79   | 3,72     | 0,000    | *** |
| Sesso: Maschi  | -66,75   | -160,20 | 40,50   | -1,65    | 0,100    | .   | Sesso: Maschi  | -40,67   | -97,62  | 18,77   | -2,17    | 0,031    | *   |
| Età  | -2,12    | -5,09   | 1,27    | -1,67    | 0,096    | .   | Età  | -0,79    | -1,89   | 0,44    | -1,79    | 0,074    | .   |
| Tocilizumab<br>Sarilumab<br>Baricitinib1               | -107,65  | -258,36 | 72,62   | -1,48    | 0,139    | .   | BMI  | 2,28     | 5,46    | 1,22    | 1,86     | 0,063    | .   |
| Eparina1   | 70,00    | 168,00  | 38,86   | 1,80     | 0,072    | .   | ACE_inibitori<br>Sartani1                              | -24,76   | -59,43  | 13,87   | -1,79    | 0,075    | .   |
|  |          |         |         |          |          |     | O2intensità:<br>≤2L/min                                | -14,80   | -35,52  | 20,78   | -0,71    | 0,476    | .   |
|  |          |         |         |          |          |     | O2intensità:<br>2-6 L/min                              | -387,51  | -930,03 | 219,00  | -1,77    | 0,077    | .   |
|  |          |         |         |          |          |     | O2intensità:<br>>6 L/min                               | -102,30  | -245,52 | 155,59  | -0,66    | 0,511    | .   |
|  |          |         |         |          |          |     |  |          |         |         |          |          | .   |
| Multiple R-Squared: 0.0571, Adjusted R-Squared: 0.0365 |          |         |         |          |          |     | Multiple R-Squared: 0.0968, Adjusted R-Squared: 0.0796 |          |         |         |          |          |     |
| DIAGNOSTICS  |          |         |         |          |          |     |  |          |         |         |          |          |     |
| PRIMA ONDATA   |          |         |         |          |          |     | SECONDA ONDATA   |          |         |         |          |          |     |
|  | Estimate | Std.    | t-value | Pr(> t ) |          |     |  | Estimate | Std.    | t-value | Pr(> t ) |          |     |
| invMillsRatio  | -54,59   | 71,86   | -0,76   | 0,45     |          |     | invMillsRatio  | -39,79   | 34,74   | -1,15   | 0,25     |          |     |
| sigma  | 259,61   | NA      | NA      | NA       |          |     | sigma  | 115,01   | NA      | NA      | NA       |          |     |
| rho  | -0,21    | NA      | NA      | NA       |          |     | rho  | -0,35    | NA      | NA      | NA       |          |     |

Le Main Equation del Macro-Sintomo Psicologico evidenziano una diversa durata di base nelle due varianti, la prima è circa quattro volte superiore della seconda. Anche i determinanti sono molto differenti, eccezion fatta per il Sesso e l'Età. Le donne, infatti, soffrono circa 67 giorni in più degli uomini, mentre, all'aumentare dell'età, la durata dei sintomi calano di tre giorni per ogni anno del soggetto. I pochi determinanti in comune e le differenti durate di base, sembrerebbero rafforzare l'ipotesi che il virus, nella seconda ondata, abbia subito delle mutazioni. Analizzando i diversi determinanti, nella prima ondata del virus, le terapie Eparina e Tocilizumab Sarilumab Baricitinib hanno un effetto opposto. La prima aumenta le durate complessive dei disturbi psichici di 168 giorni, mentre la seconda le diminuisce di 259. Stessa cosa accade per il BMI, ACE Inibitori Sartani e O2Intensità. Nella seconda ondata, infatti, queste le aumentano di 6 giorni per ogni punto di massa corporea, mentre i farmaci antipertensivi e l'intensità dell'ossigenoterapia le diminuiscono di 60 e 931 giorni.

### 6.3 Analisi Macro-Sintomo Fisico

Il modello logistico, attraverso gli odds ratio, ha permesso di valutare, come per i disturbi psicologici, i determinanti e gli effetti di questi sul Macro-Sintomo Fisico.

Grafico 7: Odds ratio ricavati dalle stime del modello Logit del Macro-Sintomo Fisico



Anche in questo caso le similitudini tra le due ondate sono veramente poche, anzi ne è presente solo una a sottolineare l'ipotesi che il virus abbia subito, nella seconda ondata, una mutazione. L'analogia tra le varianti è che i maschi, come avveniva per il Macro-Sintomo psicologico, hanno una propensione minore a sviluppare le affezioni fisiche.

Nella prima ondata, per gli utilizzatori regolari degli Steroidi Inalatori, le probabilità di sviluppare la sintomatologia risultano essere circa cinque volte superiori rispetto a coloro che non li assumono. Stessa osservazione vale per i medicinali Beta Bloccanti e per le patologie Metaboliche. Al contrario, i farmaci Antiaggreganti, rappresentano un fattore protettivo. Pazienti con una forma lieve del virus hanno riscontrato un 53% in meno di probabilità di sviluppare i disturbi fisici e, contro intuitivamente, all'aumentare del numero delle comorbidità, la propensione alle affezioni fisiche cala del 20%.

Come accadeva nel Macro-Sintomo Psicologico, le cure HCQ, per la prima ondata, e quelle Steroide e Remdesevir, per la seconda, hanno comportato un incremento del rischio. Mentre, a differenza del caso psichico, la somministrazione di farmaci diuretici e la contrazione di patologie Polmonari e Autoimmuni, incrementano le probabilità di riscontrare sintomi post-Covid fisici.

Esattamente per come accadeva per il Macro-Sintomo Psicologico, la main equation permette di valutare, per coloro che hanno contratto affezioni fisiche, quali sono quei fattori che ne hanno prolungato o diminuito la durata.

Tabella 17: Stime del modello OLS per il Macro-Sintomo Fisico

| MACRO-SINTOMO FISICO                                   |          |         |         |          |               |     |  |          |         |          |         |          |     |
|--|----------|---------|---------|----------|---------------|-----|--|----------|---------|----------|---------|----------|-----|
| MAIN EQUATION  |          |         |         |          |               |     |  |          |         |          |         |          |     |
| PRIMA ONDATA   |          |         |         |          |               |     | SECONDA ONDATA   |          |         |          |         |          |     |
|  | Estimate | In Days | Std.    | t-value  | Pr(> t )      |     |  | Estimate | In Days | Std.     | t-value | Pr(> t ) |     |
| (Intercept)  | 249,68   | 424,46  | 46,99   | 5,31     | 0,000         | *** | (Intercept)  | 120,87   | 205,48  | 39,00    | 3,10    | 0,002    | *** |
| Sesso: Maschi  | -69,38   | -117,95 | 30,89   | -2,25    | 0,025         | *   | Sesso: Maschi  | -37,38   | -63,54  | 14,89    | -2,51   | 0,012    | *   |
| N_Comorbidità  | 29,47    | 50,10   | 11,31   | 2,61     | 0,009         | **  | N_Comorbidità  | -6,36    | -10,82  | 2,88     | -2,21   | 0,027    | *   |
| Ipolipemizzante Statina1                               | -98,82   | -167,99 | 51,87   | -1,91    | 0,057         | .   | BMI  | 3,52     | 5,99    | 1,18     | 2,98    | 0,003    | *** |
| Diuretici1   | -123,90  | -210,63 | 75,70   | -1,64    | 0,102         |     | O2intensità: <=2L/min                                  | -30,00   | -51,00  | 19,90    | -1,51   | 0,132    |     |
| Eparina1   | 69,55    | 118,24  | 34,61   | 2,01     | 0,045         | *   | O2intensità: 2-6 L/min                                 | -37,26   | -63,34  | 20,64    | -1,81   | 0,071    |     |
|  |          |         |         |          |               |     | O2intensità: >6 L/min                                  | -18,50   | -31,44  | 14,21    | -1,30   | 0,194    |     |
|  |          |         |         |          |               |     |  |          |         |          |         |          |     |
| Multiple R-Squared: 0.0742, Adjusted R-Squared: 0.0533 |          |         |         |          |               |     | Multiple R-Squared: 0.0741, Adjusted R-Squared: 0.0602 |          |         |          |         |          |     |
| DIAGNOSTICS  |          |         |         |          |               |     |  |          |         |          |         |          |     |
| PRIMA ONDATA   |          |         |         |          |               |     | SECONDA ONDATA   |          |         |          |         |          |     |
|  | Estimate | Std.    | t-value | Pr(> t ) |               |     | Estimate   | Std.     | t-value | Pr(> t ) |         |          |     |
| invMillsRatio  | 47,38    | 60,16   | 0,79    | 0,43     | invMillsRatio |     | -32,91   | 39,85    | -0,83   | 0,41     |         |          |     |
| sigma  | 248,51   | NA      | NA      | NA       | sigma         |     | 113,81   | NA       | NA      | NA       |         |          |     |
| rho  | 0,19     | NA      | NA      | NA       | rho           |     | -0,29  | NA       | NA      | NA       |         |          |     |

La durata di base del Macro-Sintomo Fisico, anche in questo caso, risulta superiore nella prima ondata e l'essere donna, in entrambe le varianti, influisce negativamente sui disturbi post-Covid. Le analogie tra i disturbi fisici e psicologici terminano con la

presenza, nella prima ondata, del trattamento Eparina, mentre nella seconda, delle variabili BMI e O<sub>2</sub>intensità. L'impatto dell'Eparina sui disturbi fisici è molto simile, per direzione ed intensità, a quello che accadeva per i disturbi psicologici, mentre, l'indice di massa corporea e l'ossigenoterapia incidono maggiormente sul Macro-Sintomo Fisico. Al contrario, i farmaci Diuretici e Ipolipemizzanti Statina, presenti solo nella prima ondata, riducono le durate dei sintomi di circa 211 e 168 giorni.

Oltre al Sesso, anche il Numero di Comorbidità risulta essere un determinante comune alle due varianti per il Macro-Sintomo Fisico. Nella prima, all'aumentare del numero delle patologie concomitanti, si registra un prolungamento delle sintomatologie di 50 giorni per comorbidità, mentre nella seconda accade esattamente l'opposto. Questa, unita alle differenti durate di base, sembrerebbe essere un'ulteriore evidenza della mutazione dell'infezione nella seconda ondata.

## Capitolo 7

### Conclusioni

L'identificazione e lo studio dei determinanti dei Macro-Sintomi hanno permesso di trarre le seguenti conclusioni. Relativamente al Long Covid, infatti, è possibile affermare che le donne sono i soggetti maggiormente colpiti. Gli Steroidi Inalatori, insieme ai Beta Bloccanti e la terapia HCQ, per la prima ondata, hanno determinato un forte aumento del rischio di sviluppare le sintomatologie. Stessa cosa per quanto riguardano le cure Steroide e Remdesivir nella seconda. Per concludere, l'indice di massa corporea (BMI), l'Eparina e l'Ossigenoterapia hanno comportato, per i soggetti affetti da sintomatologie post-SARSCOV2, un aumento delle durate di quest'ultime.

Da un'attenta analisi dei risultati prodotti dai modelli Heckit, sembrerebbe emergere che le due ondate infettive siano state caratterizzate da un virus di differente natura. Questo è supportato dalle nette divergenze empiriche, in termini di determinanti e durate, tra le ondate delle Macro-Sintomatologie. In particolare, variabili comuni come farmaci Neuro Psicoattivi e Numero di Comorbidità hanno prodotto, nelle due varianti, un effetto addirittura opposto. Per non parlare del fatto che la durata della prima fase, per il Macro-Sintomo Psicologico e per il Macro-Sintomo Fisico, è stata rispettivamente superiore di circa quattro e due volte.

Al contrario, le forti analogie tra i determinanti, all'interno delle rispettive varianti, sembrano supportare l'ipotesi della presenza di un'unica sintomatologia Psico-Fisica.

Tuttavia, sono solo delle prime evidenze empiriche e per certificare e supportare entrambe le ipotesi, sarebbe necessario condurre un nuovo studio maggiormente accurato. Uno spunto per un'indagine futura sarebbe quello di valutare l'impatto dei vaccini sui disturbi Long Covid, per analizzare se hanno avuto o meno un effetto protettivo. Inoltre, sarebbe opportuno eseguire lo studio su un campione nazionale per poter generalizzare i risultati all'intera popolazione della penisola, a differenza di quello in esame generalizzabile ai soli soggetti residenti nell'area dell'ospedale metropolitano.



## Bibliografia e Sitografia

- Hill, C. R., Griffiths, W. E., & Lim, G. C. (2018). *Principles of Econometrics* (5th ed.). Wiley.
- Pennoni F. *Dispensa di Modelli Statistici Analisi Statistica Multivariata*
- Martinella, V. (2022, May 9). *Long Covid nei malati di tumore: chi rischia di più?* CorrieredellaSera.  
[https://www.corriere.it/salute/sportello\\_cancro/22\\_maggio\\_09/long-covid-malati-tumore-b7aa30ca-c616-11ec-80ae-9f956e43a8f0.shtml](https://www.corriere.it/salute/sportello_cancro/22_maggio_09/long-covid-malati-tumore-b7aa30ca-c616-11ec-80ae-9f956e43a8f0.shtml)
- Martinella, V. (2022b, July 24). *Long Covid, i sintomi: le otto manifestazioni più frequenti che perdurano nel tempo.* Corriere della Sera.  
[https://www.corriere.it/salute/malattie\\_infettive/cards/otto-manifestazioni-piu-frequenti-long-covid-che-perdurano-tempo-cosa-sappiamo-disturbi-piu-comuni/post-covid-19-condition\\_principale.shtml?viewName=APERTURA\\_FOTO\\_FISSA](https://www.corriere.it/salute/malattie_infettive/cards/otto-manifestazioni-piu-frequenti-long-covid-che-perdurano-tempo-cosa-sappiamo-disturbi-piu-comuni/post-covid-19-condition_principale.shtml?viewName=APERTURA_FOTO_FISSA)
- Marrone, C. (2022, May 12). *Long Covid, almeno un sintomo in oltre la metà dei pazienti due anni dopo il ricovero.* Corriere della Sera.  
[https://www.corriere.it/salute/neuroscienze/22\\_maggio\\_12/long-covid-sintomi-pazienti-due-anni-dopo-d31e56ca-d132-11ec-b465-8b7c23727ee0.shtml](https://www.corriere.it/salute/neuroscienze/22_maggio_12/long-covid-sintomi-pazienti-due-anni-dopo-d31e56ca-d132-11ec-b465-8b7c23727ee0.shtml)
- Cervia C, Zurbuchen Y, Taeschler P, et al. *Immunoglobulin signature predicts risk of post-acute COVID-19 syndrome. Nat Commun. 2022;13(1):446. Published 2022 Jan 25. doi:10.1038/s41467-021-27797-1*
- Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. *Incidence, co-occurrence, and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors of COVID-19. PLoS Med.*

2021;18(9):e1003773. Published 2021 Sep 28. doi:

10.1371/journal.pmed.1003773

- Michael Clark. *Heckman Selection / Model Estimation by Example*. <https://m-clark.github.io/models-by-example/heckman-selection.html#>
- Bogdan Oancea (2017). *Heckman correction technique-a short introduction*.
- Rob Hicks (2018, September 17). *The Heckman Sample Selection Model*.  
[https://rlhick.people.wm.edu/stories/econ\\_407\\_notes\\_heckman.html](https://rlhick.people.wm.edu/stories/econ_407_notes_heckman.html)
- Taylor & Francis (2000). *A Heckman Selection-t Model*.  
[https://www.tandfonline.com/doi/full/10.1080/01621459.2012.656011?casa\\_token=HBaEH3xDXEIAAAAA%3AQgrYwdUejTnd1E9dur\\_Y9EzgRMQbh1Mo8VuSL5RY1WQk8mcX7c-kZpPI6FyN04XFVhXDpBP8No](https://www.tandfonline.com/doi/full/10.1080/01621459.2012.656011?casa_token=HBaEH3xDXEIAAAAA%3AQgrYwdUejTnd1E9dur_Y9EzgRMQbh1Mo8VuSL5RY1WQk8mcX7c-kZpPI6FyN04XFVhXDpBP8No)