



**POLITECNICO
MILANO 1863**

SCUOLA DI INGEGNERIA
CIVILE, AMBIENTALE
E TERRITORIALE

La sorveglianza ambientale dei reflui in Lombardia: stima della prevalenza di SARS-CoV-2 ed influenza e confronto con i dati clinici

TESI DI LAUREA MAGISTRALE IN
INGEGNERIA PER L'AMBIENTE E IL TERRITORIO –
ENVIRONMENTAL AND LAND PLANNING ENGINEERING

Autore: Eleonora Corti

Matricola: 233324

Relatore: Prof. Arianna Azzellino

Anno Accademico: 2024-2025

Ringraziamenti

Giunta al termine di questo percorso accademico, vorrei ringraziare tutte le persone che mi hanno accompagnata, non solo nella realizzazione di questo lavoro di tesi, ma anche in questi anni di studio.

Un ringraziamento va ai miei Professori, che hanno reso possibile lo sviluppo di questo elaborato.

Ringrazio tutta la mia famiglia.

Alla mia mamma e al mio papà, perché ci sono sempre per me e perché continuano a credere in me quando io non riesco a farlo.

A mia sorella, perché sa sempre come farmi sorridere e stare meglio, e allo zio, per la presenza e il sostegno.

Ai nonni, che purtroppo non sono qui fisicamente, ma sono certa che, in qualche modo, siano comunque presenti.

Un grazie va alle mie amiche di sempre e ai miei amici conosciuti in università. Grazie per aver condiviso momenti di felicità, di ansia e di studio.

Le persone che hanno condiviso questo percorso insieme a me sanno quanto siano importanti per me: ho sempre cercato di dimostrarlo e lo dimostrerò, ogni giorno.

Abstract

La sorveglianza ambientale delle acque reflue si è affermata negli anni '30 come approccio non invasivo ed economico per il monitoraggio della salute pubblica.

Conosciuta come Wastewater-Based Epidemiology (WBE), si è rivelata particolarmente efficace per tenere sotto controllo le malattie infettive diffuse da virus; è quindi un ottimo strumento per il rilevamento precoce delle epidemie prima che la popolazione mostri sintomi, il monitoraggio delle tendenze delle malattie e la valutazione delle misure di sanità pubblica (lockdown o campagne vaccinali) a livello locale, nazionale ed internazionale.

La ricerca di poliovirus infettivi nelle acque reflue era già una pratica consolidata in grado di fornire elementi probanti l'eradicazione della poliomielite, ottenuti tramite le campagne di vaccinazione.

Durante la pandemia di COVID-19 il ruolo della WBE è emerso maggiormente come strumento complementare ai metodi clinici, per individuare i casi asintomatici o non tracciati.

Ad oggi, però, emergono delle criticità legate all'accuratezza delle stime, per quanto riguarda la normalizzazione dei dati o la stima precisa del numero dei casi a partire dalla carica virale, che viene misurata nei reflui in ingresso agli impianti di depurazione.

Per questi motivi, con il seguente studio, si ottimizzano i metodi di calcolo e confronto con i dati clinici in modo tale da rendere la WBE uno strumento sempre più affidabile, che possa in futuro perfino sostituire la sorveglianza clinica tradizionale.

Parole chiave: La sorveglianza ambientale, COVID-19, normalizzazione.

Abstract in lingua inglese

Environmental surveillance of wastewater emerged in the 1930s as a non-invasive and cost-effective approach to public health monitoring.

Known as wastewater-based epidemiology (WBE), it has proven particularly effective for controlling infectious diseases spread by viruses. It is therefore an excellent tool for early detection of epidemics before the population shows symptoms, monitoring disease trends, and evaluating public health measures (lockdowns or vaccination campaigns) at the local, national, and international levels.

In the past, researchers have found infectious polioviruses in wastewater, which was crucial to demonstrate polio eradication due to vaccination campaigns.

During the COVID-19 pandemic, the role of WBE has emerged more as a complementary tool to clinical methods, to identify asymptomatic or untraceable cases.

However, critical issues have emerged regarding the accuracy of estimates, whether related to data normalization or precise estimation of the number of cases based on viral load, which is measured in wastewater entering wastewater treatment plants.

For these reasons, the following study optimizes the calculation methods and comparison with clinical data to make WBE an increasingly reliable tool, potentially even replacing traditional clinical surveillance in the future.

Key-words: Environmental surveillance, COVID-19, normalization.

Indice

Ringraziamenti.....	III
Abstract.....	V
Abstract in lingua inglese.....	VII
Indice.....	IX
1 Introduzione.....	1
1.1 Principi generali.....	1
1.2 Obiettivi dello studio	2
2 La sorveglianza ambientale delle acque reflue	4
2.1 Impianti di depurazione e laboratori virologici.....	4
2.2 Fasi della sorveglianza ambientale	6
2.2.1 Campionamento del campione	7
2.2.2 Trasporto del campione.....	12
2.2.3 PCR (polymerase chain reaction).....	13
2.2.4 Igiene e sicurezza.....	16
2.2.5 Valutazioni di stabilità	18
2.2.6 Stoccaggio e conservazione dei campioni.....	19
2.3 Ruoli e competenze nel sistema di sorveglianza ambientale.....	20
2.4 Etica e privacy	21
2.5 Finanziamenti	22
2.6 Potenzialità e limiti.....	23
2.7 Metodologie per le analisi dei dati WBE.....	24
2.7.1 Gli abitanti equivalenti	24
2.7.1.1 Calcolo degli abitanti equivalenti	25
2.7.2 La normalizzazione della carica virale.....	27
2.7.2.1 Metodi di normalizzazione	27

Indice	X
2.7.3 Shedding rate	29
2.7.4 Stima dei casi da carica virale e prevalenza	33
3 Il caso studio.....	35
3.1 Descrizione del dataset 2022.....	35
3.2 Normalizzazioni applicate al caso studio	42
3.2.1 Normalizzazione tramite azoto ammoniacale (N)	43
3.2.2 Normalizzazione tramite abitanti equivalenti (AE)	43
3.3 Applicazione della statistica al caso studio	45
3.3.1 Analisi di correlazione	45
3.4 Analisi di regressione.....	47
3.4.1 Modelli di regressione lineare semplice.....	47
3.4.2 Modelli di regressione lineare multipla	51
3.5 Stima dei casi da carica virale e prevalenza	56
3.6 Modelli di regressione lineare multipla per valutare la prevalenza WBE	58
3.7 Validazione delle metodologie su dati aggiornati e su virus influenza	63
3.7.1 Descrizione del dataset 24-25.....	63
3.7.2 Modelli di regressione lineare multipla per valutare la prevalenza WBE e il virus influenza	65
4 Conclusioni.....	75
4.1.1 Scelta del gene Migliore N_1	75
4.2.1 Stima della prevalenza.....	76
4.3.1 Dataset 24-25 ed estensione al virus influenza.....	76
Bibliografia	79
A Appendice	81
Lista di figure	127
Lista di tabelle.....	135

1 Introduzione

1.1 Principi generali

La sorveglianza ambientale delle acque reflue, nota anche come epidemiologia basata sulle acque reflue (Wastewater-Based Epidemiology, WBE), rappresenta un approccio innovativo e non invasivo per il monitoraggio della salute pubblica. Attraverso l'analisi di biomarcatori presenti nelle acque reflue, quali agenti patogeni e contaminanti ambientali, è possibile ottenere una panoramica anonima ed aggregata dello stato sanitario di una comunità. Questo metodo consente di individuare in modo tempestivo l'insorgenza di focolai, seguire l'evoluzione delle malattie infettive nel tempo e nello spazio, nonché monitorare la circolazione di ceppi virali emergenti o resistenti.

La WBE è particolarmente preziosa perché permette di rilevare sia casi sintomatici che asintomatici, offrendo un quadro più completo rispetto ai tradizionali sistemi di sorveglianza basati sui test clinici individuali. Inoltre, l'approccio si distingue per la sua economicità e rapidità, consentendo campionamenti frequenti su larga scala, utili per supportare le decisioni sanitarie pubbliche a livello locale e nazionale.

L'utilizzo della sorveglianza delle acque reflue ha radici storiche che risalgono agli albori dell'epidemiologia moderna: sin dagli anni '30 del Novecento, durante le epidemie di poliomielite, vennero identificati poliovirus nelle acque reflue, elemento che ha contribuito agli sforzi di eradicazione della malattia. Oggi i metodi di rilevazione sono molto diversi da quelli di allora e la diffusione di tecniche molecolari di indagine hanno reso la sorveglianza dei reflui più efficace. La recente pandemia di COVID-19 ha rilanciato l'importanza di questa metodologia, con l'istituzione di sistemi nazionali di monitoraggio come il National Wastewater Surveillance System (NWSS) negli Stati Uniti, dedicato al tracciamento del SARS-CoV-2 e successivamente ampliato per altri patogeni come il virus respiratorio sinciziale (RSV), virus dell'influenza aviaria e vaiolo delle scimmie.

Oltre alla rilevazione di virus, la sorveglianza ambientale delle acque reflue consente di monitorare la diffusione della resistenza antimicrobica, identificando la presenza di geni e batteri resistenti trasmessi attraverso le vie oro-fecali. Tale capacità fornisce un contributo prezioso per la definizione di strategie di controllo efficaci e mirate.

In prospettiva, la WBE rappresenta uno strumento strategico per la sanità pubblica, capace di fornire avvisi precoci, supportare la gestione tempestiva delle epidemie ed

integrare i dati raccolti dai sistemi di sorveglianza tradizionali. L'analisi dei campioni di acque reflue ha dimostrato di poter anticipare la comparsa di nuove varianti virali rispetto al rilevamento clinico, confermando il ruolo cruciale di questa metodologia nell'epidemiologia contemporanea.

L'adozione sistematica della sorveglianza ambientale all'interno dei sistemi sanitari nazionali richiede un coordinamento tra scienziati, epidemiologi e decisori politici per garantire che i dati raccolti siano tempestivi, accurati ed utili per orientare interventi efficaci a tutela della salute della popolazione.

1.2 Obiettivi dello studio

Negli ultimi anni, dunque, la sorveglianza ambientale basata sulle acque reflue (Wastewater-Based Epidemiology, WBE) ha dimostrato un enorme potenziale per integrare ed anticipare la sorveglianza clinica tradizionale [4]. Affinché questa metodologia diventi uno strumento affidabile, si devono superare alcune criticità, legate alla normalizzazione dei dati e alla stima del numero effettivo di individui infetti all'interno della popolazione servita da ciascun impianto di depurazione.

Lo studio ha, dunque, l'obiettivo di rafforzare la capacità predittiva e rappresentativa della WBE, migliorando l'accuratezza delle stime epidemiologiche derivate dai reflui.

In particolare, gli obiettivi specifici sono i seguenti:

- ottimizzare la normalizzazione della carica virale rilevata nei reflui, sostituendo il metodo tradizionale basato sulla stima nominale degli abitanti equivalenti (AE) serviti dall'impianto, con dei metodi più efficaci e di maggiore accuratezza;
- testare l'accuratezza delle stime sul SARS-CoV-2, valutandone la correlazione tra le cariche virali normalizzate e i casi clinici (da tampone), in funzione della normalizzazione utilizzata;
- analizzare più accuratamente l'associazione fra la carica virale nel refluo e i casi clinici, per ridurre l'impatto di fattori confondenti, per migliorare la capacità di spiegare la variabilità osservata nei casi clinici;
- stimare il numero di persone infette presenti nel bacino servito da ciascun impianto, a partire dalla carica virale rilevata nei reflui, associata allo "shedding rate" virale, ovvero la quantità media di materiale virale escreta da ciascun soggetto infetto. L'obiettivo è derivare un indice di infezione direttamente dalle acque reflue, che sia indipendente dalla sorveglianza clinica;
- trasformare il numero di infetti giornalieri per impianto in termini di prevalenza, ovvero in percentuale della popolazione che risulta infetta in quel dato giorno; in questo modo è possibile confrontare questa percentuale derivante dall'analisi del refluo con quella di tamponi positivi clinici;

- estendere l'intera metodologia al caso dell'influenza, analizzando la presenza e l'andamento dei geni virali specifici nelle acque in ingresso agli impianti di depurazione.

Si mira, dunque, al consolidamento del ruolo della WBE come strumento epidemiologico indipendente, per stimare con maggiore precisione la reale diffusione delle infezioni. Tale approccio, una volta validato, potrà rappresentare un supporto strategico per le autorità sanitarie nella gestione di epidemie future e nell'ottimizzazione delle politiche di prevenzione e contenimento.

L'elaborato è organizzato come segue. Nel *Capitolo 2* viene descritta la sorveglianza ambientale delle acque reflue, con riferimento agli impianti di depurazione e alle diverse fasi della sorveglianza, dal campionamento fino all'analisi in laboratorio. Nel *Capitolo 3* viene presentato il caso studio, dalla descrizione del dataset fino all'applicazione dei modelli di regressione, sia per l'anno 2022 sia per gli anni 2024-2025. Infine, nel *Capitolo 4* vengono discussi i risultati ottenuti, presentate le conclusioni dello studio e delineate possibili prospettive future, tra cui l'esplorazione di modelli alternativi. In *Appendice* vengono riportati i risultati completi dei modelli di regressione.

2 La sorveglianza ambientale delle acque reflue

2.1 Impianti di depurazione e laboratori virologici

Gli impianti di depurazione rappresentano una componente fondamentale delle infrastrutture urbane, garantendo il trattamento delle acque reflue prima della loro reimmissione nell'ambiente. Le acque reflue sono costituite per circa il 99,9 % da acqua e per lo 0,1 % da una miscela di solidi sospesi, materia organica, nutrienti, microrganismi, virus ed inquinanti di origine domestica ed industriale. È proprio questa piccola frazione a contenere le informazioni più rilevanti per il monitoraggio ambientale e sanitario [1].

Dal punto di vista epidemiologico, il punto di interesse è il refluo in ingresso, dove si concentra il carico organico e microbico proveniente dalla popolazione. È qui che vengono installati i sistemi di campionamento, che raccolgono frazioni di refluo durante l'arco della giornata per ottenere un campione rappresentativo. Tali sistemi devono essere mantenuti a temperatura controllata (4 °C) e puliti regolarmente per evitare contaminazioni [2].

Le acque reflue contengono escreti umani (urine, feci, secrezioni), che trasportano agenti patogeni eliminati dalle persone infette, sintomatiche o asintomatiche. Proprio questa caratteristica rende le acque reflue uno strumento potenzialmente potente per lo studio della circolazione di agenti infettivi all'interno di una comunità. Oltre ai patogeni, nei reflui si ritrovano materia organica, residui alimentari, detersivi, metalli pesanti e nutrienti come azoto e fosforo. La composizione media delle acque reflue urbane si distribuisce approssimativamente come segue [1]:

- 40 % da infiltrazioni nella rete fognaria;
- 15 % da acque meteoriche superficiali;
- 25 % da scarichi domestici;
- 20 % da scarichi industriali.

Queste percentuali possono variare in funzione del tipo di rete fognaria adottata. Nei sistemi separati, le acque reflue e quelle meteoriche viaggiano in condotte distinte: questo consente di mantenere una composizione più stabile del refluo, con una presenza ridotta di acque meteoriche e un maggiore controllo sui carichi inquinanti.

Al contrario, nei sistemi combinati, entrambe le tipologie di acqua confluiscono nella stessa rete, aumentando il rischio di sovraccarichi idraulici e di diluizione dei contaminanti in caso di precipitazioni intense.

In presenza di sistemi separati, la composizione del refluo rimane relativamente costante nel tempo, con percentuali tipiche vicine a: 30–35 % da infiltrazioni, 5 % da acque meteoriche, 35 % da scarichi domestici e 25 % da scarichi industriali [1]. Nei sistemi combinati, invece, le percentuali possono variare sensibilmente in base alle condizioni meteorologiche. Durante eventi piovosi, ad esempio, le acque meteoriche possono rappresentare oltre il 30 % del totale, con una conseguente riduzione proporzionale della quota domestica ed industriale. Ciò comporta una maggiore variabilità del refluo e può complicare le attività di campionamento e interpretazione dei dati, in particolare nell'ambito della sorveglianza epidemiologica [1].

La temperatura del refluo, influenzata da fattori stagionali ed urbanistici, varia generalmente tra i 10 °C e i 25–30 °C. Dal punto di vista chimico, le acque reflue contengono nutrienti come azoto totale (in forme TKN e TAN) e fosforo totale (sotto forma di ortofosfati e polifosfati). I carichi medi per abitante equivalente al giorno sono stimati in [1]:

- Azoto totale: 8–14 g/AE/giorno;
- Fosforo totale: 1,5–2,5 g/AE/giorno;
- COD (Chemical Oxygen Demand): 100–130 g/AE/giorno;
- BOD (Biochemical Oxygen Demand): 60–70 g/AE/giorno.

Il trattamento delle acque reflue si articola in tre fasi principali [1]:

1. Trattamento primario: per rimuovere sabbie, materiali grossolani e parte del materiale organico mediante griglie, grit chambers e sedimentazione (con la produzione di fango primario);
2. Trattamento secondario: di tipo biologico, effettuato solitamente con processi a fanghi attivi in vasche aerate, dove i microrganismi degradano la sostanza organica. Segue una sedimentazione secondaria per separare i fanghi (fango secondario);
3. Trattamenti terziari opzionali: includono disinfezione (UV, clorazione), rimozione avanzata dei nutrienti o microinquinanti, e filtrazioni finali.

La gestione dei fanghi prodotti richiede le seguenti fasi: digestione anaerobica per la stabilizzazione e produzione di biogas, seguita da ispessimento e disidratazione, per la riduzione del contenuto di acqua nei fanghi, e smaltimento finale (in discarica, incenerimento o uso agricolo) [1].

Per quanto riguarda il campionamento delle acque reflue è fondamentale scegliere il metodo, la localizzazione, la frequenza e il volume del campione in base all'obiettivo; è possibile valutare la qualità del refluo in ingresso o in uscita, stimare il tasso di eliminazione dei contaminanti o monitorare la presenza di patogeni nella popolazione.

Per il monitoraggio delle acque reflue il campionamento viene eseguito prima che le acque reflue scorrano all'interno dell'impianto di depurazione.

Un elemento cruciale per interpretare correttamente i dati raccolti è la conoscenza del bacino di utenza servito da ciascun impianto. Esso può variare da poche migliaia a milioni di abitanti equivalenti, influenzando direttamente la stima della carica virale.

Dopo il prelievo, i campioni vengono trasportati in laboratori virologici specializzati, in cui si effettuano analisi molecolari per la rilevazione e quantificazione del materiale genetico virale. La tecnica più diffusa è la RT-qPCR (reverse transcription quantitative PCR), che consente di convertire l'RNA virale in DNA e amplificarlo quantitativamente [3].

I laboratori devono seguire protocolli standardizzati e validati, spesso basati su normative internazionali.

La collaborazione tra impianti di depurazione, laboratori analitici e autorità sanitarie è quindi fondamentale per trasformare le acque reflue in una fonte di dati epidemiologici affidabile. Su queste basi si fonda l'approccio della Wastewater-Based Epidemiology (WBE).

2.2 Fasi della sorveglianza ambientale

La sorveglianza ambientale delle acque reflue è caratterizzata da diverse fasi collegate fra di loro, che garantiscono la qualità e l'affidabilità dei risultati [5]. Le principali fasi operative sono le seguenti:

- Pianificazione: definizione degli obiettivi della sorveglianza (es. rilevamento precoce, tracciamento di tendenze epidemiologiche, valutazione dell'efficacia degli interventi sanitari), individuazione dei siti di campionamento rappresentativi, scelta degli agenti patogeni da monitorare e definizione della frequenza di campionamento;
- Campionamento: raccolta dei campioni di acque reflue, secondo protocolli standardizzati che specificano la modalità (manuale o automatica), frequenza, volume, punto di prelievo e tipologia di matrice (liquido o fango). La rappresentatività del campione è cruciale per l'affidabilità dei risultati;
- Trasporto e conservazione: trasferimento dei campioni al laboratorio in condizioni controllate, tipicamente a 4 °C, con tempi compatibili con il materiale

genetico (idealmente entro 24 ore). È fondamentale rispettare le norme di biosicurezza e le linee guida per il trasporto di materiali potenzialmente infetti;

- Trattamento e preparazione del campione: processi preliminari come la concentrazione del patogeno, la chiarificazione e l'eventuale pastorizzazione;
- Analisi molecolare: estrazione dell'RNA o del DNA, amplificazione (ad esempio mediante PCR) e quantificazione del materiale genetico relativo ai patogeni target. In funzione delle finalità della sorveglianza, possono essere impiegate anche tecniche avanzate come il sequenziamento di nuova generazione (NGS);
- Elaborazione ed interpretazione dei dati: integrazione dei risultati analitici con informazioni epidemiologiche, cliniche e contestuali (come i dati di popolazione servita, la portata dell'impianto, le condizioni meteo o stagionali), al fine di fornire indicazioni utili per la salute pubblica.

Ciascuna di queste fasi richiede il rispetto di criteri metodologici rigorosi, l'impiego di protocolli condivisi e un coordinamento efficace tra laboratori, enti gestori del servizio idrico ed autorità sanitarie [5].

2.2.1 Campionamento del campione

Il campionamento rappresenta un pilastro fondamentale di ogni programma di sorveglianza ambientale delle acque reflue e deve essere pianificato ed implementato con attenzione per garantire rappresentatività, affidabilità e accuratezza dei risultati [5]. A seconda dell'obiettivo epidemiologico sono richiesti protocolli e strategie specifiche per assicurare la qualità dei dati.

Le strategie di campionamento variano in base a diversi fattori:

- popolazione bersaglio (generale vs vulnerabile, ad esempio ospedali o case di riposo);
- scopo della sorveglianza (di comunità vs ai punti di ingresso, ad esempio in aeroporto, routine vs emergenza);
- tipo di patogeno o sostanza da monitorare e relative dinamiche epidemiologiche.

È fondamentale sviluppare protocolli standardizzati per il campionamento delle acque reflue in diversi contesti applicativi, tra cui:

- localizzazioni specifiche e confinate (es. scuole);
- mezzi di trasporto e hub (es. aeroporti, porti);

- situazioni di emergenza (es. epidemie);
- ambienti decentralizzati (es. piccoli villaggi);
- campionamento dei fanghi.

La standardizzazione di queste pratiche è essenziale per garantire:

- affidabilità, comparabilità e validità dei dati;
- supporto a decisioni basate su evidenze in ambito sanitario, ambientale e normativo.

Le strategie e i protocolli comprendono anche aspetti come:

- frequenza di campionamento;
- volume e tipo di campione;
- localizzazione.

Il campionamento dell'acqua in ingresso al depuratore fornisce informazioni preziose sulla composizione delle acque reflue prima di qualsiasi trattamento, consentendo il tracciamento della presenza di:

- patogeni;
- residui farmaceutici;
- altri contaminanti derivanti dalla comunità;
- inquinanti chimici e altri contaminanti rilevanti per la salute pubblica, la cui concentrazione può essere influenzata dai trattamenti successivi. Ad esempio, nel caso del monitoraggio di agenti infettivi come SARS-CoV-2, i campioni possono essere raccolti agli ingressi degli impianti per stimare la prevalenza del virus nella comunità.

Si osserva che il campionamento dell'affluente grezzo viene spesso effettuato dopo le unità di trattamento preliminare, come le griglie e la rimozione della sabbia negli impianti meccanici di trattamento delle acque reflue. Questo consente di rimuovere materiali grossolani e sostanze inorganiche dense dalle acque raccolte prima della loro campionatura. Negli impianti di trattamento passivo, come i sistemi a lagunaggio o quelli decentralizzati a bassa intensità operativa, il campionamento può richiedere il prelievo a monte presso stazioni di pompaggio, per accedere ad acque reflue influenti non miscelate.

È necessario disporre di protocolli di campionamento standardizzati per raccogliere campioni rappresentativi che riflettano accuratamente le caratteristiche e la composizione delle acque reflue, così da garantire la rilevazione affidabile di agenti patogeni, genetici o chimici bersaglio e la migliore stima della quantità escreta presente

nella matrice acquosa. In questo contesto, è raccomandato l'uso di campionatori automatici, programmati per prelevare campioni composti di acque reflue influenti, ponderati sul flusso e raccolti ad intervalli prestabiliti, successivamente alla grigliatura e alla rimozione della sabbia. Tali campionatori permettono di raccogliere campioni rappresentativi in un determinato periodo, consentendo una valutazione accurata delle

caratteristiche dell'influente e l'identificazione di eventuali variazioni nella composizione delle acque reflue.

Strategie specifiche di campionamento in punti localizzati e/o confinati della rete fognaria sono fondamentali per ottenere dati affidabili per la valutazione sanitaria pubblica. Il campionamento in prossimità di strutture sanitarie, come ospedali o case di cura, può fornire informazioni dettagliate sulla presenza localizzata di patogeni o contaminanti di interesse per la salute pubblica. Ad esempio, campioni raccolti vicino agli ospedali possono aiutare a rilevare la presenza di batteri antibiotico-resistenti, elementi genetici o residui farmaceutici nelle acque reflue.

Le procedure standardizzate sono essenziali per ridurre i rischi legati all'accesso in ambienti confinati, come l'esposizione ad acque contaminate, gas tossici o pericoli fisici.

In questi contesti, si utilizzano spesso campionatori istantanei (grab samplers) per raccogliere campioni discreti in punti specifici della rete fognaria, data la difficoltà di accesso e la possibile assenza di elettricità o spazio sufficiente. Questi dispositivi manuali permettono agli operatori di raccogliere campioni singoli in un dato momento, offrendo una fotografia puntuale della qualità delle acque reflue in punti mirati. Tuttavia, anche in questi casi, quando possibile e sicuro, l'uso di campionatori automatici composti è preferibile per ottenere campioni più rappresentativi.

Il campionamento delle acque reflue provenienti da mezzi mobili, come aeroporti e aeromobili, si è dimostrato una strategia efficace per tracciare la diffusione globale dei patogeni, fornendo informazioni preziose per il rilevamento precoce di focolai epidemici.

Per quanto riguarda gli aeroporti, i protocolli e i dispositivi di campionamento utilizzati, sono simili a quelli impiegati presso gli impianti di trattamento delle acque reflue. Il campionamento delle acque reflue di bordo degli aeromobili richiede invece protocolli e strumentazione specifici, oltre a misure di sicurezza rafforzate. Sono stati sviluppati dispositivi dedicati, come trituratori per aerei e aeroporti, per consentire un campionamento efficace in questi ambienti particolari. I serbatoi degli aerei, infatti, contengono quantità molto ridotte di acqua, ma spesso con materiali solidi più concentrati: carta igienica, residui organici, fanghi.

In uno spazio così piccolo, i materiali non si miscelano bene, quindi un campione grezzo sarebbe molto disomogeneo e poco rappresentativo.

In questo contesto, è fondamentale sviluppare protocolli di campionamento armonizzati tra diverse giurisdizioni regionali, così da rendere confrontabili i campioni rappresentativi anche mediante tecniche come il sequenziamento genomico.

E' necessario fare ulteriori ricerche per capire come i risultati ottenuti dall'analisi delle acque reflue degli aerei si confrontino con quelli presi dagli impianti di trattamento o da altre fonti. Questo è fondamentale per valutare se il metodo sia efficace nell'identificare i virus, nel caratterizzarli e nel monitorare come si diffondono.

Studi teorici suggeriscono che la sorveglianza delle acque reflue degli aerei potrebbe aiutare ad individuare in anticipo la diffusione globale di nuovi focolai. Tuttavia, sono necessari studi

epidemiologici per capire meglio quanto questo metodo sia rappresentativo e utile nel sistema di sorveglianza globale.

Infine, è fondamentale che aeroporti, compagnie aeree ed operatori a terra collaborino e definiscano regole chiare per sostenere lo sviluppo di una rete di monitoraggio efficace, che possa funzionare sempre ed aumentare rapidamente l'attività in caso di emergenze sanitarie.

Lo sviluppo di protocolli per il campionamento delle acque reflue in situazioni di emergenza, come sversamenti chimici, catastrofi naturali, crisi umanitarie (ad esempio in aree di conflitto o tra popolazioni rifugiate) o focolai epidemici, è fondamentale per garantire una risposta rapida ed una valutazione accurata dei potenziali rischi. Approcci di campionamento standardizzati permettono la raccolta di dati tempestivi e affidabili, utili per supportare le operazioni di risposta, la valutazione del rischio, e la progettazione ed attuazione di eventuali misure di contenimento o bonifica.

Ad esempio, durante un'epidemia, il campionamento delle acque reflue in punti centrali o selezionati della rete può essere utile per monitorare la prevalenza e la distribuzione del patogeno nella popolazione, fornendo indicazioni utili agli interventi di sanità pubblica. In questi scenari, si possono impiegare campionatori automatici portatili, dotati di funzionalità di monitoraggio in tempo reale, oppure dispositivi per campionamento istantaneo (grab samplers). Questi strumenti possono essere rapidamente distribuiti sul territorio per raccogliere campioni compositi continui o campioni discreti, consentendo una valutazione rapida della composizione delle acque reflue e della presenza di patogeni o altri inquinanti.

Sono necessari protocolli specializzati e standardizzati anche per il campionamento delle acque reflue in aree prive di infrastrutture fognarie centralizzate, come comunità

rurali, località isolate o insediamenti informali, al fine di ottenere dati attendibili e significativi per la salute pubblica.

In tali contesti, i campioni possono essere prelevati da serbatoi di accumulo domestici. In alcune aree non servite da reti fognarie, le acque reflue vengono trasportate quotidianamente mediante autobotti, dai serbatoi di edifici e abitazioni ad impianti centralizzati (spesso sistemi di trattamento a lagunaggio). In questi casi, sono necessari metodi di campionamento standardizzati direttamente dalle autobotti, in modo da garantire la rappresentatività dei campioni ma sono necessarie ulteriori ricerche.

In aggiunta, strategie di campionamento presso strutture comunitarie, come scuole, ospedali o luoghi di lavoro di grandi dimensioni, potrebbero offrire nuove opportunità per migliorare la biosorveglianza ed orientare più efficacemente le decisioni in ambito di salute pubblica. In contesti non allacciati a fognature, possono essere utilizzati campionatori portatili istantanei o campionatori passivi per raccogliere campioni da sistemi decentralizzati. Questi dispositivi permettono il prelievo anche in ambienti dove il campionamento centralizzato non è fattibile.

Tuttavia, nella pratica, non è realistico campionare singolarmente le proprietà private. Le aree prive di fognature rappresentano una sfida rilevante per la sorveglianza ambientale. Una possibile soluzione consiste nel campionare presso luoghi di aggregazione, come scuole, luoghi di culto, centri di lavoro o altri centri comunitari, in modo da ottenere un campione rappresentativo di una porzione significativa della popolazione.

Il campionamento dei fanghi nell'ambito della sorveglianza ambientale delle acque reflue per scopi di salute pubblica riguarda la raccolta di fanghi per valutare la presenza e la distribuzione di agenti di interesse, come patogeni, determinanti genetici, sostanze chimiche o biomarcatori.

In particolare, si consiglia l'uso di fanghi primari, cioè i solidi sospesi che sedimentano durante la prima fase di rimozione dei solidi (sedimentazione primaria), in quanto non sono ancora stati sottoposti a trattamenti che potrebbero alterare o distruggere i bersagli di interesse. Rispetto all'affluente grezzo, il fango primario è più concentrato e può essere preferibile quando si cercano bersagli presenti in basse concentrazioni.

Tuttavia, l'uso di campioni di fango presenta anche sfide operative, che devono essere valutate caso per caso. Tra queste: la presenza di additivi chimici introdotti negli impianti o il ricircolo di flussi di refluo trattati che possono interferire con l'analisi.

È quindi necessaria l'adozione di procedure standardizzate di campionamento per raccogliere campioni rappresentativi di fango negli impianti di trattamento, a supporto di una valutazione affidabile dei patogeni o degli altri bersagli prioritari. I dispositivi più comuni sono i campionatori a carota (core samplers), utilizzati per

raccogliere carote intatte dai bacini di accumulo fanghi o dagli impianti stessi. Questi strumenti consentono di valutare la variabilità spaziale e le caratteristiche dei fanghi, inclusa la presenza di patogeni, metalli pesanti e contaminanti organici.

Nel contesto della sorveglianza ambientale per la salute pubblica, il campionamento dei fanghi rappresenta un ambito promettente ma ancora poco esplorato, e sono necessari ulteriori studi per svilupparne appieno il potenziale [5].

2.2.2 Trasporto del campione

Il trasporto dei campioni di acque reflue al laboratorio di analisi rappresenta una fase critica, che richiede un'attenta pianificazione. Le procedure di trasporto devono essere adattate a ciascun sistema di sorveglianza, tenendo conto di variabili quali il numero e la posizione dei laboratori coinvolti, il volume e la frequenza dei campioni, i vincoli temporali, l'efficienza economica e le considerazioni legate alla biosicurezza [5].

Di seguito sono riportate le principali linee guida relative al trasporto dei campioni nel contesto della sorveglianza epidemiologica basata sulle acque reflue:

1. Modalità di trasporto:

a. Nei progetti pilota o negli studi di ricerca, il personale di laboratorio può occuparsi direttamente del prelievo e del trasporto. Nei sistemi di sorveglianza su larga scala, è necessario predisporre mezzi di trasporto organizzati;

b. La spedizione tramite servizi di consegna rapida è una soluzione pratica per lunghe distanze, ma comporta aspetti normativi legati al trasporto di sostanze infettive e può non essere sostenibile economicamente a lungo termine;

c. Per distanze più brevi, è possibile utilizzare un veicolo dedicato al trasporto dei campioni, opzione efficiente dal punto di vista economico ma che richiede una coordinazione accurata;

d. L'utilizzo di rotte di trasporto già esistenti, come quelle impiegate per campioni clinici o per il monitoraggio ambientale di routine, può rappresentare un'alternativa economicamente vantaggiosa, ma necessita di un coordinamento approfondito.

2. Tempistiche di trasporto: idealmente, l'analisi dei campioni dovrebbe avvenire lo stesso giorno del prelievo o entro 24 ore, per limitare la degradazione dell'RNA. Si raccomanda quindi l'uso di servizi di spedizione in giornata o con consegna notturna.

3. Imballaggio:

a. I campioni devono essere trasportati in contenitori di plastica, disinfettati esternamente e confezionati nel rispetto delle normative sul trasporto di sostanze infettive;

b. È essenziale l'uso di imbottiture adeguate, un contenitore secondario impermeabile e la disinfezione dell'involucro esterno;

c. I contenitori devono essere a tenuta stagna e ben fissati durante il trasporto, e il personale coinvolto deve essere formato nella gestione di sostanze potenzialmente infettive.

4. Refrigerazione:

a. I campioni devono essere trasportati a bassa temperatura, ma non congelati, idealmente mantenuti a 4°C dal momento del prelievo fino all'analisi. Se il raffreddamento non è possibile, i campioni devono essere protetti dal surriscaldamento;

b. Nei servizi di spedizione, è possibile ordinare la spedizione refrigerata a 4°C. In alternativa, si possono utilizzare furgoni refrigerati o contenitori isotermitici con ghiaccio;

c. La valutazione delle condizioni ambientali esterne è raccomandata per monitorare la temperatura dei campioni durante il trasporto.

5. Documentazione ed etichettatura:

a. Ogni campione deve essere chiaramente etichettato con le seguenti informazioni: ID del campione, luogo, data e ora del prelievo, tipo di campione, nome dell'operatore, e dati relativi alla catena di custodia;

b. L'etichettatura esterna dell'imballaggio deve rispettare le normative per il trasporto di sostanze infettive, inclusa la presenza di un simbolo di rischio biologico.

In sintesi, il trasporto dei campioni di acque reflue verso il laboratorio di analisi richiede una pianificazione attenta che consideri modalità di trasporto, tempi, imballaggio, refrigerazione e tracciabilità documentale, al fine di preservare l'integrità del campione e garantire risultati analitici affidabili [5].

2.2.3 PCR (polymerase chain reaction)

Dopo il prelievo e il trasporto dei campioni, il processo di sorveglianza ambientale prosegue con la fase pre-analitica di preparazione del campione in laboratorio, che comprende le operazioni di concentrazione e l'estrazione dell'RNA o del DNA. Tali passaggi sono finalizzati a rendere possibile la successiva rilevazione e quantificazione dei patogeni mediante analisi molecolare.

Si utilizzano tecnologie di misurazione ed analisi genetiche e chimiche per monitorare la presenza di patogeni e contaminanti nelle acque reflue, tra cui la PCR e il metodo NGS.

Per concentrare i patogeni, per il miglioramento della capacità di rilevamento nella sorveglianza ambientale delle acque reflue, poiché la concentrazione dei patogeni spesso è molto bassa, vengono impiegati due metodi differenti: quelli progettati per la concentrazione della frazione liquida del campione e quelli per la concentrazione della frazione solida [5].

I metodi per concentrare la parte liquida comprendono l'ultrafiltrazione (filtri fini che trattengono i patogeni), la filtrazione con membrane cariche (filtri che attirano elettricamente i patogeni), e le sfere magnetiche (piccole particelle che aderiscono ai patogeni e poi vengono raccolte con una calamita).

I metodi per concentrare la parte solida comprendono la flocculazione (si aggiungono sostanze che aiutano l'aggregazione dei patogeni in gruppi più grossolani, che poi si possono raccogliere), la precipitazione con PEG (si utilizza una sostanza chimica che fa cadere i patogeni sul fondo) e l'ultracentrifugazione (si fa girare il campione a velocità molto alta per isolare i patogeni).

Il metodo utilizzato dipende dalla tipologia di campione (se è solo frazione liquida o anche fango), dai patogeni in analisi e quindi dal comportamento di patogeni nell'acqua (se tendono a rimanere nel liquido oppure si attaccano alle particelle solide) e dai costi.

A seconda del metodo, si possono ottenere risultati differenti, quindi è importante tenerne conto nell'analisi dei dati.

Una volta che vengono concentrati i patogeni, viene applicata l'estrazione degli acidi nucleici, ovvero del materiale genetico.

L'estrazione degli acidi nucleici avviene in tre fasi:

- Lisi: avviene la rottura delle cellule e viene liberato il DNA o l'RNA, tramite reagenti chimici o calore;
- Isolamento: il materiale genetico viene separato dal resto del campione, tramite filtri o solventi per raccogliere solo il DNA/RNA;
- Purificazione: si purifica il materiale genetico da sostanze che potrebbero interferire con le analisi successive, ad esempio enzimi o detriti.

È fondamentale estrarre la tipologia corretta di materiale genetico; viene estratto l'RNA per virus come SARS-CoV-2 o influenza e il DNA per batteri o altri tipi di virus.

Nella sorveglianza ambientale vengono utilizzati diversi metodi di PCR per il rilevamento e la quantificazione del materiale genetico di patogeni: la PCR convenzionale, la PCR quantitativa (qPCR), la PCR digitale e la PCR con trascrizione inversa (RT-PCR).

Innanzitutto, se nel campione si estrae RNA si deve convertire in DNA prima di applicare la PCR, poiché l'RNA non può essere amplificato direttamente; questo processo avviene tramite un enzima, la trascrittasi inversa, che converte RNA in cDNA (DNA complementare).

Ogni virus presenta sequenze caratteristiche, ovvero geni, che vengono definite come "impronte digitali", che permettono la loro rilevazione nel campione.

Vengono, dunque, creati dei primer, DNA sintetico, che si legano solo a quelle determinate sequenze specifiche; se il virus è presente nel campione allora avviene il legame tra il virus e il primer e viceversa.

Una volta avvenuto il legame, l'enzima polimerasi, insieme al primer, amplifica il DNA (il DNA del patogeno è inizialmente basso e le copie sono poche) nel processo chiamato PCR.

Possono essere applicati diversi metodi di PCR:

- PCR convenzionale: avviene l'amplificazione del DNA, e alla fine della reazione il DNA amplificato viene inserito nel gel agarosio. Si fa passare corrente elettrica, e il DNA, essendo carico negativamente, migra verso il polo positivo. Inoltre, i frammenti di dimensione più piccola si muovono più velocemente e viceversa, perciò la separazione avviene per dimensione. Il gel contiene un colorante fluorescente che si lega al DNA, e sotto una lampada UV, se il patogeno è presente, si vede una banda luminosa di una certa lunghezza;
- PCR quantitativa: mentre il DNA si amplifica viene utilizzata una sonda che emette luce, e a seconda della sua intensità, si può quantificare la quantità di carica virale. Viene utilizzata generalmente per il SARS-CoV-2;
- PCR digitale: il campione viene diluito per avere una concentrazione bassa di molecole di DNA/RNA, dopodiché viene diviso in goccioline, in modo tale che per ognuna sia presente una copia del gene o nessuna copia. In ogni gocciolina avviene la PCR e, se la copia non è presente, non avviene l'amplificazione. Le goccioline vengono analizzate tramite un sistema di rilevamento e si contano le copie esatte presenti nel campione;
- RT-PCR: avviene la conversione da RNA a DNA tramite la trascrittasi inversa, e poi si applica la PCR convenzionale o la PCR quantitativa.

Un altro metodo che può essere utilizzato è la NGS (Next-Generation Sequencing), che sequenza tutto il materiale genetico presente all'interno di un campione.

Quindi, a differenza della PCR che utilizza primer specifici per amplificare solo sequenze genetiche note dall'organismo target per capire se un certo virus è presente e in che quantità, la NGS identifica anche specie sconosciute e varianti nuove, fornendo anche informazioni quantitative.

Per il monitoraggio delle acque reflue, il metodo più conveniente è la PCR poiché le analisi sono più rapide, mirate e meno costose.

La NGS è più impiegata per indagini esplorative oppure per ricerca [5].

2.2.4 Igiene e sicurezza

Le considerazioni di biosicurezza devono essere tenute presenti in tutte le fasi del processo di campionamento ed analisi delle acque reflue. È fondamentale definire protocolli di biosicurezza specifici per il sistema di sorveglianza, comprendenti indicazioni su [5]:

- Vaccinazioni raccomandate per il personale;
- Utilizzo di dispositivi di protezione individuale (DPI);
- Procedure per la manipolazione sicura dei campioni;
- Requisiti per gli spazi di laboratorio;
- Pratiche di disinfezione e gestione dei rifiuti;
- Formazione specifica per operatori di campo e tecnici di laboratorio.

La sicurezza degli operatori è essenziale durante la raccolta dei campioni, poiché le acque reflue non trattate possono contenere agenti patogeni, prodotti chimici pericolosi ed altri contaminanti potenzialmente dannosi. Pertanto, è importante:

- Indossare DPI adeguati, inclusi guanti, tute protettive, mascherine/respiratori e visiere protettive, a seconda del contesto;
- Ridurre al minimo il rischio di contatto diretto con le acque reflue e di inalazione di aerosol;
- Applicare rigorose misure di igiene personale e decontaminazione;
- Garantire che il personale sia formato sulle procedure di emergenza in caso di esposizione accidentale.

Queste misure sono fondamentali per proteggere il personale coinvolto e garantire la sicurezza e l'affidabilità del sistema di sorveglianza ambientale.

I rischi per gli operatori che prelevano campioni sono simili a quelli di altri lavoratori negli impianti di trattamento delle acque reflue (WWTP). Le procedure standard per la protezione dei lavoratori sono generalmente sufficienti, ma potrebbero necessitare di adattamenti in caso di comparsa di patogeni particolarmente pericolosi nelle acque reflue. Durante il campionamento devono essere utilizzati dispositivi di protezione individuale (DPI) adeguati e pratiche di lavoro sicure.

Altri rischi per la sicurezza, come quelli legati ad infortuni fisici, devono anch'essi essere considerati.

I campioni devono essere raccolti in contenitori di plastica, con disinfezione esterna. In alcuni casi i campioni devono essere confezionati e spediti come sostanze infettive.

All'arrivo in laboratorio, i campioni devono essere aperti in un'area designata e trasferiti in un vassoio con bordi alti per evitare fuoriuscite.

Le analisi devono essere eseguite in laboratori con flusso d'aria unidirezionale. Devono essere predisposte aree separate per la vestizione e la svestizione dei DPI per evitare contaminazioni incrociate.

Le procedure lavorative all'interno del laboratorio devono essere unidirezionali, con tutta l'attrezzatura e i materiali consumabili pronti per il processo di analisi.

I campioni devono essere aperti e manipolati all'interno di una cappa a flusso laminare, con il livello di protezione adeguato in base alle metodologie utilizzate.

Molti laboratori effettuano la pastorizzazione dei campioni di acque reflue per ridurre il rischio di infezione. Il riscaldamento a 56°C per 30-90 minuti è in grado di inattivare i virus rivestiti da involucro.

Gli effetti del trattamento termico sulla concentrazione di RNA virale sono controversi e la pastorizzazione non è universalmente raccomandata. In caso di applicazione, devono essere utilizzati controlli di recupero della matrice.

Le superfici potenzialmente contaminate durante la lavorazione dei campioni devono essere disinfettate con prodotti approvati.

Soluzioni di ipoclorito allo 0,1% o prodotti a base di alcool con almeno il 70% di etanolo sono efficaci per eliminare SARS-CoV-2, mentre per altri patogeni sono necessarie specifiche valutazioni di rischio.

I campioni residui e tutti i rifiuti di laboratorio potenzialmente infettivi devono essere decontaminati mediante autoclave e smaltiti secondo le linee guida.

L'autoclave è un'apparecchiatura che utilizza vapore acqueo ad alta temperatura e pressione per sterilizzare strumenti, materiali e rifiuti potenzialmente contaminati da microrganismi patogeni.

Le pratiche di gestione dei rifiuti devono rispettare le normative locali o nazionali vigenti.

Il personale di campo e di laboratorio deve essere formato nella gestione corretta di campioni infettivi e nelle misure di sicurezza.

Tutti gli operatori devono avere un piano vaccinale aggiornato e utilizzare DPI adeguati in tutte le fasi di campionamento ed analisi.

Ogni laboratorio coinvolto nella sorveglianza deve effettuare una valutazione del rischio specifica per sito e attività.

Le misure di mitigazione, infine, devono essere definite in base ai rischi individuati nelle valutazioni [5].

2.2.5 Valutazioni di stabilità

I virus rivestiti da involucro generalmente decadono più rapidamente rispetto ai virus non rivestiti, anche se esistono differenze tra le famiglie virali [5].

La stabilità dei virus nelle acque reflue è influenzata da molteplici fattori fisici, chimici e biologici che agiscono durante il percorso dal rilascio virale fino all'analisi di laboratorio. Di seguito, si riportano i principali fattori:

- Fattori fisici: la luce solare rappresenta un importante agente ambientale che contribuisce all'inattivazione dei microrganismi, influenzando significativamente i tassi di decadimento sia dei virus rivestiti sia di quelli non rivestiti. Tuttavia, a causa della natura sotterranea dei sistemi fognari e dell'opacità delle acque reflue, l'effetto della luce solare sul decadimento virale nelle acque reflue risulta attenuato rispetto ad altri ambienti acquatici.

La dinamica del trasporto nelle reti fognarie e la ripartizione del virus nella matrice delle acque reflue (ad esempio virus legati a particelle solide o presenti in fase liquida) possono influenzare il tasso di decadimento: in alcuni casi, i virus adsorbiti alle particelle solide e trasportati sul fondo o in prossimità del letto della condotta mostrano tassi di decadimento maggiori rispetto a quelli trasportati in flusso dinamico.

La temperatura ha un impatto rilevante sulla stabilità virale, influenzando l'integrità dell'involucro virale e dell'RNA. Ad esempio, la concentrazione del genoma di SARS-CoV-2 decresce con cinetica di primo ordine all'aumentare della temperatura.

Temperature di congelamento, invece, favoriscono la persistenza del genoma virale per periodi prolungati, anche se lo stoccaggio prolungato a $-20\text{ }^{\circ}\text{C}$ può ridurre i carichi virali. Inoltre, il congelamento e scongelamento ripetuti possono portare alla degradazione del materiale genetico da parte di RNA. L'aggregazione o l'adsorbimento su materiale solido può proteggere i virus dall'inattivazione;

- Fattori chimici: la composizione chimica delle acque reflue, inclusi pH, salinità, sostanze organiche e altre sostanze chimiche, influisce sulla persistenza virale e sull'integrità dell'RNA nel campione. L'RNA di SARS-CoV-2, ad esempio, persiste più a lungo in acque pulite, così come altri virus. La composizione e la carica organica delle acque reflue variano in base al tipo di sistema fognario, all'origine dei reflui e alla distanza percorsa fino agli impianti di trattamento. Generalmente, un'elevata concentrazione di materia organica nelle acque reflue aumenta la persistenza delle popolazioni virali;
- Fattori biologici: la presenza di microrganismi antagonisti può ridurre la persistenza virale nell'acqua. Alcuni batteri vivi possono produrre sostanze antivirali, mentre i protozoi possono diminuire la concentrazione virale per adsorbimento o predazione. È stato dimostrato che la persistenza virale è più lunga nelle acque reflue autoclave (sterilizzate) rispetto a quelle non trattate.

Data la complessità e la molteplicità dei fattori coinvolti, valutare la stabilità virale in condizioni reali di campo risulta complesso. Il confronto simultaneo di campioni sottoposti a diverse condizioni può aiutare a chiarire i meccanismi di degradazione durante il trasporto e lo stoccaggio [5].

2.2.6 Stoccaggio e conservazione dei campioni

Una corretta conservazione dei campioni di acque reflue, sia liquide sia solide, è fondamentale per garantire la stabilità degli analiti, l'affidabilità delle analisi molecolari e la riproducibilità dei risultati [5]. In particolare, la conservazione dell'RNA virale risulta critica, poiché la sua degradazione può portare a una sottostima della carica virale e compromettere la tempestività e l'accuratezza del monitoraggio.

Le modalità di conservazione variano a seconda dello stato e del trattamento del campione:

- Campioni di acque reflue non trattate: devono essere raffreddati immediatamente dopo il prelievo e conservati a 4°C per un massimo di due settimane. Se il raffreddamento non è possibile, è raccomandata l'analisi entro 24 ore. Il congelamento, se necessario, può compromettere l'integrità dell'RNA, ed è pertanto preferibile suddividere i campioni in aliquote, per evitare scongelamenti ripetuti. Si ricorda che, per le analisi chimiche, la conservazione deve avvenire mediante congelamento;
- Campioni solidi: valgono gli stessi principi sopra esposti, con il vantaggio di richiedere minori volumi di stoccaggio, grazie al ridotto contenuto d'acqua;
- Campioni pretrattati: in questa fase, il campione può essere sottoposto a pre-filtrazione e/o centrifugazione per rimuovere i solidi sospesi. In alcuni casi si applica la pastorizzazione per ragioni di biosicurezza, sebbene non sia obbligatoria;
- Concentrati virali: devono essere conservati congelati, preferibilmente a -75/-80°C (o almeno a -20°C), specialmente se in piccoli volumi. Anche i filtri utilizzati per la concentrazione devono essere conservati alle stesse condizioni;
- Acidi nucleici estratti: l'RNA virale può essere stoccato a -20°C per brevi periodi, ma per la conservazione a lungo termine sono preferibili temperature ultrabasse [5].

2.3 Ruoli e competenze nel sistema di sorveglianza ambientale

La WBE coinvolge un gruppo di professionisti che lavorano insieme per raccogliere, analizzare e comprendere i dati in modo efficace [5].

I principali professionisti sono:

- Ingegneri ambientali: operano negli impianti di trattamento delle acque reflue e sono direttamente responsabili del prelievo dei campioni. Garantiscono, inoltre, che il campionamento e il trattamento avvengano in linea con le normative vigenti;
- Scienziati di laboratorio: analizzano i campioni di acque reflue per identificare patogeni. I biologi e microbiologi, ad esempio, svolgono un ruolo chiave nell'analisi genetica dei virus e patogeni;
- Esperti della sanità pubblica: interpretano i dati in un contesto di salute pubblica e prendono decisioni in base alle emergenze sanitarie;
- Esperti della privacy e della protezione dei dati: garantiscono che i dati vengano trattati in modo anonimo, sicuro e conforme alle normative di protezione dei dati. Tutelano, inoltre, la privacy individuale nei confronti delle persone monitorate [5].

2.4 Etica e privacy

L'applicazione della sorveglianza ambientale delle acque reflue solleva importanti questioni etiche e legate alla privacy [5]. Sebbene i dati ottenuti da questa forma di monitoraggio siano di natura aggregata e non siano, in genere, riconducibili a singoli individui, è comunque essenziale garantire il rispetto della normativa vigente in materia di protezione dei dati personali e della riservatezza collettiva. In questo contesto, gli esperti di privacy e protezione dei dati giocano un ruolo cruciale: si occupano di assicurare che i dati siano trattati in modo anonimo e sicuro, adottando misure tecniche ed organizzative adeguate, e valutano i potenziali rischi. Inoltre, vigilano affinché la privacy delle persone indirettamente coinvolte venga tutelata, anche nei casi in cui l'identificabilità non sia esplicita.

In linea generale, i dati generati dalla WBE non sono attribuibili ad individui o a piccoli gruppi di persone, come i singoli nuclei familiari. Tuttavia, la percezione pubblica della privacy può cambiare radicalmente in funzione del contesto in cui viene applicata. Ad esempio, l'utilizzo della WBE in ambienti circoscritti e statici, come carceri, scuole o strutture residenziali, può essere percepito come potenzialmente invasivo, poiché la popolazione osservata è facilmente identificabile.

Anche nel settore dell'aviazione sono emerse preoccupazioni: l'industria teme danni reputazionali o azioni legali nel caso in cui la presenza di un patogeno venga associata a un determinato volo o compagnia. Di conseguenza, le attività di sorveglianza in questo ambito devono prevedere particolari cautele per garantire l'anonimato del vettore.

Un ulteriore elemento critico è rappresentato dalla possibile ambivalenza d'uso dei dati: se da un lato la WBE è uno strumento prezioso per la sanità pubblica, come nel caso del monitoraggio di

SARS-CoV-2, dall'altro può essere impiegata per finalità più controverse, come la rilevazione dell'uso di droghe illecite. In questi casi, pur potendo orientare politiche pubbliche virtuose, come l'allocazione di risorse per programmi di trattamento delle dipendenze, può anche alimentare timori di sorveglianza repressiva.

Per tutte queste ragioni, è fondamentale adottare un approccio fondato su trasparenza, proporzionalità e coinvolgimento della comunità. La chiarezza nella comunicazione degli obiettivi della sorveglianza, l'adozione di strategie di anonimizzazione adeguate e il rispetto dei diritti collettivi contribuiscono a favorire la fiducia pubblica e l'accettazione sociale della WBE. Solo in questo modo è possibile coniugare efficacemente le potenzialità scientifiche di questo strumento con la tutela dei principi fondamentali di libertà e dignità individuale [5].

2.5 Finanziamenti

Il finanziamento della WBE per la salute pubblica può comportare diverse considerazioni [5]:

1. Investimento iniziale: avviare un programma di WBE da zero richiede un notevole investimento iniziale in infrastrutture, attrezzature e tecnologie per la raccolta, il trasporto, l'analisi e la comunicazione dei risultati. Tuttavia, l'utilizzo di infrastrutture esistenti, laddove disponibili, può ridurre i costi;
2. Costi operativi ricorrenti: mantenere o espandere i programmi di sorveglianza comporta costi operativi continui, tra cui personale, materiali di laboratorio, reagenti, analisi dei dati e integrazione con i sistemi di sorveglianza esistenti. Comprendere gli obiettivi della sorveglianza, i target monitorati, il livello di rischio associato a tali target e le loro dinamiche può contribuire alla flessibilità e sostenibilità del sistema;
3. Allocazione delle risorse: le agenzie sanitarie pubbliche e i governi devono destinare risorse per finanziare la WBE insieme ad altre iniziative di sanità pubblica, spesso in competizione per fondi limitati. Tuttavia, la WBE può contribuire ad un programma di sorveglianza multi-target più efficiente in termini di costi, migliorando l'efficienza allocativa tra diversi metodi di sorveglianza;
4. Sostenibilità del finanziamento: garantire un finanziamento a lungo termine è essenziale per la sostenibilità dei programmi di sorveglianza delle acque reflue. Si raccomanda l'integrazione della WBE con i sistemi di sorveglianza esistenti. Questo può richiedere il sostegno continuo da parte dei bilanci pubblici o la ricerca di fonti di finanziamento alternative;
5. Analisi costo-efficacia: condurre un'analisi costo-efficacia per dimostrare il valore e l'impatto della WBE nei risultati delle risposte sanitarie può aiutare ad assicurare fondi e sostegno, oltre che a stabilire priorità di investimento ed attuazione;
6. Equità ed accesso: garantire un accesso equo alla sorveglianza delle acque reflue tra le diverse comunità e regioni, in particolare nelle aree meno servite, può richiedere finanziamenti e

risorse aggiuntive. Approcci sentinella alternativi, come la sorveglianza delle acque reflue negli aeroporti e sugli aeromobili, possono generare grandi benefici per l'intera popolazione grazie ad azioni sanitarie precoci;

7. Coordinamento regionale e internazionale: di fronte a minacce sanitarie globali, come le pandemie, possono rendersi necessari meccanismi di coordinamento e

finanziamento a livello regionale e internazionale per supportare programmi WBE transfrontalieri e nei settori del trasporto e dell'aviazione [5].

2.6 Potenzialità e limiti

Di seguito, vengono riportati i vantaggi della sorveglianza delle acque reflue [5]:

- Rilevamento precoce delle epidemie: le acque reflue possono contenere materiale genetico come virus e patogeni e questo aiuta la rilevazione precoce di focolai. L'importanza della sorveglianza emerge in particolare quando i test clinici non sono diffusi o non vengono effettuati a causa della lievità dei sintomi;
- Monitoraggio delle tendenze delle malattie: studiare le acque reflue con una certa frequenza aiuta ad avere un quadro ampio sull'andamento di malattie ed a testare le misure della sanità pubblica;
- Valutazione delle misure di sanità pubblica: misurare i carichi virali nelle acque è utile per studiare l'efficacia degli interventi come i lockdown o le campagne vaccinali, per analizzare l'utilizzo di droghe;
- A differenza di altri strumenti utili per monitorare la salute della popolazione (tamponi o ricoveri ospedalieri), la sorveglianza ambientale delle acque reflue è meno soggetta a distorsioni dovute ai cambiamenti nei comportamenti delle persone o alle scelte politiche. È dunque un indicatore oggettivo e meno influenzabile da fattori esterni, e per questo considerato molto utile per valutare l'andamento reale di un'epidemia.

Nonostante i suoi vantaggi, diversi fattori possono influenzare l'inaffidabilità della sorveglianza ambientale come indicatore delle ondate pandemiche nella popolazione:

- Fattori ambientali, come le precipitazioni e l'aumento delle temperature, possono ridurre la sensibilità del rilevamento virale, diminuendo i livelli di RNA virale e/o riducendo la stabilità del virus nelle acque reflue, con conseguente degradazione dell'RNA di SARS-CoV-2;
- Le proprietà chimico-fisiche delle acque, quali il pH, concentrazione di ammoniacale, solidi totali, presenza di inibitori della PCR, dimensione della popolazione servita e percentuali variabili di reflui urbani, agricoli e industriali, possono influenzare l'utilizzo della WBE come strumento epidemiologico;
- La scelta dei target genici dell'RNA virale può incidere sul monitoraggio a lungo termine e sulla quantificazione accurata del carico virale. L'elevata frequenza di mutazioni genetiche e l'emergere di nuove varianti virali possono infatti ridurre l'efficienza di amplificazione dei frammenti virali. Le mutazioni possono ridurre la sensibilità o determinare falsi negativi per specifici target nelle analisi delle acque

reflue, dunque, una selezione accurata dei marker virali dell'RNA è fondamentale per garantire un rilevamento e una quantificazione affidabili;

- I dati sulle acque reflue non seguono sempre lo stesso andamento dei ricoveri o dei casi, probabilmente a causa di cambiamenti nei comportamenti (riduzione dei tamponi) oppure della presenza di nuove varianti meno gravi;
- A volte, lo stesso laboratorio può ottenere dei risultati molto discordanti in giorni ravvicinati a causa della variabilità dei dati, nonostante i metodi di normalizzazione;
- Per ottenere dei dati più affidabili, sono consigliati prelievi più frequenti, che portano però ad un incremento del costo e del lavoro [5].

2.7 Metodologie per le analisi dei dati WBE

2.7.1 Gli abitanti equivalenti

L'Abitante Equivalente (AE) costituisce l'unità di misura utilizzata nel campo della depurazione delle acque reflue per uniformare le stime degli abitanti residenti, di quelli fluttuanti (pendolari e turisti) e di quelli equivalenti derivanti dagli scarichi delle attività economiche [6].

Nel contesto della letteratura relativa al WBE, la pratica della "normalizzazione della popolazione" delle concentrazioni di SARS-CoV-2 (copie geniche per volume di campione di acque reflue) consente di tenere conto del numero di individui nel bacino idrografico servito da un determinato impianto di trattamento delle acque reflue (WWTP) e di esprimere i dati come carica virale giornaliera (copie geniche per persona collegata al sistema fognario al giorno).

La normalizzazione basata sulla popolazione può seguire approcci statici o dinamici [7]. Le stime statiche della popolazione possono essere ottenute utilizzando dati censuari che rappresentano la popolazione residente permanente servita dalla rete fognaria afferente ad un determinato WWTP, oppure facendo riferimento agli abitanti equivalenti (AE), che includono anche i carichi industriali, sebbene possano non riflettere accuratamente il carico effettivo in ingresso nel sistema fognario. Pertanto, le stime statiche della popolazione presuppongono una dimensione della popolazione costante nel tempo e non tengono conto delle fluttuazioni transitorie causate da nascite, decessi e migrazioni, o di quelle attribuibili al turismo stagionale o al pendolarismo lavoro-viaggio.

Le stime dinamiche, al contrario, tengono conto delle variazioni temporanee della popolazione, e quindi forniscono una stima più accurata giorno per giorno [7].

2.7.1.1 Calcolo degli abitanti equivalenti

Le analisi qui riportate sono parte dello studio condotto dal Dipartimento di Ingegneria Civile, Architettura, Territorio, Ambiente e di Matematica (DICATAM) dell'Università degli Studi di Brescia, che, in collaborazione con il Politecnico di Milano - Dipartimento di Ingegneria Civile e Ambientale (DICA), partecipa al protocollo di intesa sottoscritto dal Politecnico di Milano con Regione Lombardia, a supporto dell'applicazione della sorveglianza ambientale per la ricerca di SARS-COV2 nelle acque reflue [8].

Nella rete di sorveglianza oggetto del protocollo d'Intesa, sono stati raccolti i seguenti parametri su otto degli impianti presenti: portata (Q), solidi sospesi (SS), COD, BOD, azoto totale (N-tot), azoto ammoniacale (N-NH₄). In questo modo, si è potuto procedere con una migliore definizione della stima degli abitanti allacciati alla rete al momento della raccolta del campione [4]. È infatti possibile ottenere una stima dinamica degli Abitanti Equivalenti allacciati contestualmente al campione raccolto attraverso l'applicazione di coefficienti di emissione medi per abitante basati su questi parametri.

Di seguito vengono riportati i coefficienti di emissione giornaliera pro-capite (g/Ab/giorno) per i principali parametri:

- COD: 110 g/Ab/giorno;
- BOD: 60 g/Ab/giorno;
- N-tot: 12 g/Ab/giorno.

La stima dinamica degli abitanti equivalenti allacciati ad un impianto viene quindi ricavata con la seguente espressione:

$$AE \text{ dinamici} = \frac{\text{ccarico del parametro chimico nel refluo}}{\text{produzione media pro capite del parametro}} \quad [1]$$

Dove:

- Il carico nel refluo è espresso in g/giorno;
- La produzione media pro capite in g/Ab/giorno, da letteratura.

È possibile stimare gli AE dinamici con diversi parametri misurabili nel refluo, dopodiché la miglior correlazione fra carica virale presente nel refluo normalizzata per

gli AE appena ricavati e i casi clinici da tampone, rivelerà il miglior parametro per quel determinato impianto [4].

Lo studio di G. Bertanza, R. e Boiocchi però, ha evidenziato che valori normalmente assunti come produzione pro capite media dei principali contaminanti (BOD=60 g/AE/giorno; COD=110–120 g/AE/giorno; azoto (N) =12–15 g/AE/giorno; fosforo (P)=1,1–1,5 g/AE/giorno) spesso non trovano riscontro nelle misure svolte dai gestori degli impianti di depurazione [8]. Più precisamente, in molti casi, il carico medio giornaliero dei diversi contaminanti (calcolato su un periodo di 1–3 anni) è diverso (generalmente inferiore) rispetto al carico calcolato moltiplicando le produzioni pro capite per il numero di AE effettivamente serviti dall'impianto di depurazione (i quali derivano, a loro volta, dalla dimensione dell'agglomerato servito dall'impianto, considerando la percentuale effettiva dell'agglomerato dotata di fognatura).

Nello studio sopra citato [8] è stata svolta un'approfondita analisi statistica dei dati di 168 agglomerati, proprio allo scopo di verificare la corrispondenza tra i carichi stimati e quelli misurati. In estrema sintesi è emerso che nei casi in cui la dimensione dell'agglomerato era ritenuta particolarmente attendibile e il contributo di scarichi di origine non domestica trascurabile, i carichi pro capite medi di BOD, COD, N e P sono risultati, rispettivamente, pari a (tra parentesi la deviazione standard): 44,6 (5,9); 82,4 (11); 9,4 (1,1); 1,08 (0,13) g/giorno. Tutti questi valori risultano inferiori a quelli di norma considerati.

In una fase di studio preliminare si sono studiate le serie storiche dei parametri caratteristici (COD, BOD, SST, N-tot, P) dei 5 impianti per i quali sono state acquisite maggiori informazioni per verificare se, tra i vari parametri, fosse possibile identificarne uno (o più) che potesse essere correlato principalmente ai contributi in fognatura di natura civile. In altre parole, questa elaborazione è stata effettuata per individuare i parametri maggiormente stabili, e quindi presumibilmente meno affetti da variabili connesse, ad esempio, a scarichi di natura non prettamente domestica. I dati di concentrazione acquisiti sono stati quindi normalizzati rispetto alle rispettive medie annuali e si nota che i parametri più stabili sono N-tot e P.

Un'altra elaborazione preliminare è consistita nella verifica della correlazione della curva dei nuovi casi giornalieri con il numero di copie geniche (N_1), normalizzato rispetto alla concentrazione dei parametri caratteristici precedentemente identificati.

Queste elaborazioni preliminari hanno evidenziato l'opportunità di identificare, impianto per impianto, uno o più parametri che possano essere considerati buoni indicatori della componente civile degli scarichi, per ottenere una più corretta interpretazione dei dati di contaminazione microbiologica.

Per ottenere gli abitanti equivalenti dinamici si dividono i valori del carico giornaliero medio di ogni parametro (in kg/giorno, per i quattro parametri BOD, COD, N e P) per

un periodo sufficientemente lungo (di norma 2-3 anni) per la corretta dimensione nominale dell'agglomerato servito dall'impianto di depurazione.

Queste elaborazioni permettono di ricavare (per ogni impianto) quattro corrispondenti valori di produzione media pro capite, che vengono confrontati con i valori riportati nello studio di Bertanza e Boiocchi più sopra richiamati.

Perché un parametro possa essere ritenuto il migliore indicatore degli scarichi di natura domestica, il suo valore medio pro-capite deve risultare il più basso tra quelli calcolati per i quattro parametri. Infatti, un qualsiasi scarico industriale, anche se molto diluito, contribuisce ad aumentare il carico in arrivo all'impianto di depurazione.

Una volta definito il parametro da utilizzare per la stima dinamica del carico civile convogliato all'impianto di depurazione, si divide il carico giornaliero per la produzione pro capite media [8].

2.7.2 La normalizzazione della carica virale

Le concentrazioni di virus nelle acque reflue possono variare a causa di diversi fattori, quali la diluizione da acqua piovana o infiltrazione di falda, da differenze nel numero di abitanti equivalenti allacciati all'impianto di depurazione, dalla variazione nel volume di reflujo dovuto alla stagionalità, oppure dalla variazione di parametri chimici relativi alle acque (BOD, TSS ecc.) [4].

È possibile correggere queste variazioni utilizzando parametri di normalizzazione, che aggiustano i dati delle concentrazioni delle cariche virali misurate nel reflujo [4].

La normalizzazione, infatti, è un passaggio fondamentale nell'analisi dei dati di sorveglianza ambientale, poiché consente di confrontare in modo corretto i risultati ottenuti in periodi diversi o in aree geografiche differenti; in questo modo si riduce l'impatto di tali variabili, permettendo di distinguere le variazioni effettive nell'incidenza dell'infezione da quelle dovute a fattori confondenti.

2.7.2.1 Metodi di normalizzazione

Nello studio di Rasha Maal-Bared et al [9] si utilizzano i seguenti parametri di normalizzazione per aggiustare e confrontare le diverse concentrazioni di virus nelle acque:

- la portata in ingresso all'impianto;
- la popolazione allacciata all'impianto di depurazione;

- alcuni biomarcatori fecali, in particolare il PMMoV (Pepper Mild Mottle Virus) che è un virus vegetale che è sempre presente nelle feci umane.

Vengono sottoposti ad analisi dodici impianti di trattamento in Alberta (Canada) di cui vengono raccolti dati di SARS-CoV-2, parametri fisici o chimici (BOD, TSS, NH₃, TKN ecc.), concentrazioni di PMMoV ed il numero di infetti di SARS-CoV-2 da tampone.

Si cerca, quindi, l'esistenza di un metodo di normalizzazione che meglio correla i casi reali di COVID-19, per ciascun impianto [9]:

- Normalizzazione per il flusso dell'impianto: permette di confrontare impianti con flussi in ingresso differenti.

Il flusso di un impianto di depurazione può diluire oppure concentrare il virus. Per questo motivo si ricava il carico di SARS-CoV-2 al giorno moltiplicando la concentrazione media di SARS-CoV-2 per il flusso giornaliero.

In questo modo si ottiene il carico del virus che è indipendente dalla quantità di acqua in ingresso all'impianto [9];

- Normalizzazione per popolazione (statica e dinamica): permette di confrontare impianti di dimensioni o popolazione differenti, dividendo il carico di SARS-CoV-2 per la popolazione statica (da anagrafe), in modo da ottenere la quantità di virus per persona, che è più rappresentativa.

Per ottenere una stima più precisa del carico pro capite, si considera la popolazione dinamica che è allacciata ad un impianto nel tempo che, quindi, non è costante (turisti, pendolari ecc.).

Normalizzando quindi il carico di SARS-CoV-2 per la popolazione dinamica, si ottiene un valore più realistico del virus per persona, che è effettivamente presente ed allacciata a quell'impianto [9];

- Normalizzazione per PMMoV (Pepper Mild Mottle Virus): il PMMoV è un biomarcatore fecale che permette di correggere l'effetto delle precipitazioni oppure altre forme di diluizioni come l'infiltrazione.

Se la quantità di PMMoV è alta, il campione è concentrato e viceversa, se la quantità di PMMoV è bassa, il campione è diluito.

Normalizzando il carico di SARS-CoV-2 nelle acque per la concentrazione di PMMoV si ottiene quindi un valore più realistico della carica virale per abitante, indipendentemente dal fatto che l'impianto sia diluito o meno.

La normalizzazione dei dati di SARS-CoV-2 nelle acque reflue non sempre migliora la correlazione con i casi reali di COVID-19 rispetto ai valori grezzi; non esiste, quindi,

un metodo universale sempre migliore rispetto ad un altro ma la correlazione migliora solo in alcuni contesti specifici.

Gli autori dello studio consigliano di impiegare almeno due metodi di normalizzazione in parallelo per comprendere quale possa aggiungere realmente valore alla correlazione con i casi reali di COVID-19 ottenuti dal monitoraggio clinico [9].

2.7.3 Shedding rate

Lo waste/fecal shedding rate (tasso di escrezione virale nelle feci) è definito come la quantità di RNA virale rilasciata per grammo di feci da ogni individuo infetto, in media, a livello della popolazione (in \log_{10} copie geniche per grammo di feci oppure in miliardi di copie geniche/persona infetta/giorno) [10].

Risultati recenti [10] [11] suggeriscono che i tassi di secrezione fecale siano influenzati da vari fattori, inclusi la demografia della comunità e l'evoluzione della pandemia, specialmente in relazione a vaccinazione e varianti.

La circolazione delle varianti di SARS-CoV-2, in particolare, può avere un'influenza diretta sui tassi di secrezione fecale a causa di differenze nella gravità della malattia, nella trasmissibilità e/o nei carichi virali. Tuttavia, i tassi di secrezione fecale confrontati tra varianti rimangono relativamente sconosciuti, specialmente con l'emergere rapido e costante di nuove varianti [10] [11].

La quantificazione dei tassi di shedding fecale di SARS-CoV-2 negli individui infetti rappresenta un componente fondamentale finora mancante per i modelli di WBE, necessari per stimare e prevedere la prevalenza delle infezioni da SARS-CoV-2 nelle comunità.

Serve a calcolare, infatti, il numero reale di persone infette, che contribuiscono a quella determinata carica virale.

Secondo lo studio Wan Yang et al [11], i valori di shedding rate derivano dalla seguente relazione:

$$Wt = N * It * SR \quad [2]$$

Dove:

- Wt è il carico virale nelle acque reflue, normalizzato per la portata;
- N è il numero di persone allacciate all'impianto;

- It è la prevalenza di infezione, che viene stimata tramite un modello epidemiologico;
- SR è lo shedding rate.

La prevalenza, dunque, non viene misurata direttamente, ma viene stimata tramite un modello epidemiologico che considera dati ufficiali, quali casi, ricoveri e decessi e anche parametri quali tempo di incubazione, mobilità della popolazione, vaccinazioni, variante circolante ecc.

Ogni settimana, il modello produce una stima della prevalenza di infezione e lo shedding rate, viene quindi ricavato dall'inversione del modello.

Nella *tabella 2.1* e *tabella 2.2* vengono riportati i valori di shedding rate stimati:

Onda / Periodo	Shedding rate (miliardi copie/giorno per persona infetta, media [IC 95%])	Lag (giorni)	Numero campioni	R ² aggiustato
2° ondata (31/08/20– 26/06/21)	qPCR: 1,44 (1,35– 1,53) ddPCR: 2,40 (2,29–2,51)	0	3	0.84
Delta (27/06/21– 04/12/21)	qPCR: 1,13 (0,86– 1,40) ddPCR: 1,89 (1,45–2,33)	0	3	0.84
Periodo Omicron (05/12/21– 29/08/23)	qPCR: 0,60 (0,59– 0,61) ddPCR: 0,96 (0,76–1,16)	0	3	0.84

Tabella 2.1 – tabella shedding rate-studio Wan Yang et al-tutte le ondate di varianti

Onda / Periodo	Shedding rate (miliardi copie/giorno per persona infetta, media [IC 95%])	Lag (giorni)	Numero campioni	R ² aggiustato
2° ondata (31/08/20– 26/06/21)	qPCR: 1,44 (1,37– 1,52) ddPCR: 2,40 (2,28–2,52)	0	4	0.74
Delta (27/06/21– 04/12/21)	qPCR: 1,09 (0,91– 1,27) ddPCR: 1,81 (1,52–2,10)	-5	4	0.37
Periodo Omicron (05/12/21– 29/08/23)	qPCR: 0,60 (0,59– 0,61) ddPCR: 0,96 (0,76–1,16)	0	3	0.86
Omicron BA.1 (05/12/21– 05/03/22)	qPCR: 0,58 (0,56– 0,61) ddPCR: 0,78 (0,32–1,23)	0	3	0.91
Dopo BA.1 (06/03/22– 29/08/23)	qPCR: 0,72 (0,70– 0,75) ddPCR: 1,21 (1,17–1,25)	-5	4	0.78

Tabella 2.2 – tabella shedding rate-studio Wan Yang et al-stratificato per ondata/periodo

In particolare, i valori di shedding rate vengono stimati come coefficiente della regressione lineare fra il carico virale nel refluo e la prevalenza, con un intervallo di confidenza del 95%.

Ogni stima dello shedding rate, infatti, ha un certo margine di errore; ad esempio, avere l'intervallo di confidenza 1.35-1.53 al 95% significa che se lo studio venisse ripetuto molte volte, il vero valore medio dello shedding rate cadrebbe nel 95% dei casi all'interno di questo intervallo.

Viene riportato anche il numero di campioni, cioè le misure di refluo che vengono utilizzate per ottenere la stima del carico virale medio, e l'ondata del virus.

Infine, vengono anche riportati il tempo di lag, che è il ritardo o anticipo temporale impiegato per far coincidere al meglio la prevalenza e il carico virale, e l' R^2 che valuta la bontà del modello.

È molto importante sottolineare come gli shedding rate stimati con questo metodo dipendano esclusivamente dalle diverse ondate, ma non dalla tipologia di gene del SARS coV-2, ovvero N_1 , N_3 oppure ORF [11].

Schmitz et al [12] hanno stimato, invece, il tasso di secrezione fecale di RNA di SARS-CoV-2 per grammo di feci di un individuo infetto basandosi sulle concentrazioni note di RNA virale nei campioni positivi di acque reflue, per il gene N_1 , e sul fatto che il numero totale di individui infetti, che contribuiscono al carico virale totale nei campioni, era anch'esso noto.

In particolare, 13 dormitori per studenti nel campus principale dell'Arizona sono stati monitorati per tutto il semestre autunnale 2020.

Ogni campione di acque reflue è stato raccolto da tombini fognari almeno 2 volte a settimana e i campioni sono stati testati tramite RT-PCR.

Il tasso di secrezione fecale (FS), definito come copie genomiche per grammo di feci (c.g./g-feci), è stato calcolato come segue:

$$FS = \frac{VC * Q * f * h}{G * I} \quad [3]$$

Dove:

- VC è la concentrazione virale nel campione di acque reflue (copie geniche/L);
- Q è la portata delle acque reflue nel tombino (gpm);
- f è il fattore di conversione fra galloni e litri;
- h è il fattore di conversione fra minuti e giorni;

- G è la massa media di feci prodotta per persona adulta al giorno (media di 300 g);
- I è il numero totale di infetti che contribuiscono al campione di acque reflue in un intervallo di sei giorni di dati clinici (cioè dal giorno prima, del giorno del campionamento, e i quattro giorni successivi).

Viene considerato un intervallo di 6 giorni poiché chi è positivo il giorno prima contribuisce ancora al carico virale nel refluo, e chi ha iniziato ad infettarsi in un determinato giorno, può iniziare a rilasciare il virus solo i giorni successivi.

Se questa finestra temporale non venisse considerata, si rischierebbe di calcolare uno shedding rate troppo alto a causa dei pochi individui considerati.

La media e la deviazione standard dei valori di FS sono poi state calcolate.

Il tasso medio di secrezione fecale risulta essere di $7,30 \pm 0,67 \log_{10}$ copie genomiche per grammo di feci (c.g./g-feci) basato sul gene N₁.

Deviazione standard di $\pm 0,67$ significa che la maggior parte dei singoli valori di shedding rate ($\approx 68\%$) cade tra 6,63 e 7,97 \log_{10} c.g./g-feci [12].

2.7.4 Stima dei casi da carica virale e prevalenza

Una volta definiti i valori di shedding rate (in \log_{10} copie geniche per grammo di feci oppure in miliardi di copie geniche/persona infetta/giorno) è possibile calcolare il numero dei casi infetti, per ciascun impianto di depurazione, come segue:

$$\text{Carico} = Q * \text{carica virale} * 1000 \quad [4]$$

Dove:

- Carico è espresso in copie geniche/giorno;
- Q è la portata dell'impianto di depurazione (m^3/giorno);
- Il carico virale nelle acque è espresso in copie geniche/L;
- 1000 è il fattore di conversione da m^3 a L.

$$\text{Numero di infetti} = \frac{\text{carico}}{\text{shedding rate}} \quad [5]$$

Dove:

Lo shedding rate viene espresso come \log_{10} copie geniche per grammo di feci oppure in miliardi di copie geniche/persona infetta per giorno.

Nel caso in cui lo shedding rate venisse espresso per grammo di feci, si considerano 300 g medi di feci espulsi al giorno per adulto.

Viene eseguita la stessa procedura per tutti i geni presenti nel refluo.

In generale, le misure di frequenza delle malattie [13] possono descrivere:

- L'insieme di tutti i casi esistenti in un determinato momento in una determinata popolazione;
- Il verificarsi di nuovi casi.

A questo scopo si usano quindi due misure fondamentali: la prevalenza e l'incidenza.

La prevalenza viene definita come:

$$Prevalenza [\%] = \frac{\text{numero di casi infetti}}{AE \text{ allacciati}} * 100 \quad [6]$$

Dove:

- Numero di casi infetti sono stimati tramite lo shedding rate;
- AE allacciati all'impianto.

Poiché il fattore tempo è importante nel calcolo della prevalenza, questa misura è di tipo statico.

L'incidenza rappresenta la proporzione di individui che vengono colpiti dalla malattia in un determinato periodo di tempo.

Viene definita come:

$$Incidenza [\%] = \frac{\text{nuovi malati nell'intervallo di tempo}}{\text{individui a rischio nell'intervallo di tempo}} * 100 [7]$$

Questa misura è di tipo dinamico.

Sinteticamente, i termini probabilistici, la prevalenza quantifica la probabilità di avere la malattia, mentre l'incidenza misura la probabilità di sviluppare quella malattia.

3 Il caso studio

3.1 Descrizione del dataset 2022

Le analisi sviluppate in questo studio si basano su due dataset distinti ma strutturalmente simili, relativi a periodi temporali differenti.

Il primo dataset, che copre l'anno 2022, è stato utilizzato per sviluppare ed ottimizzare l'intero workflow di analisi, dalla normalizzazione della carica virale alla correlazione con i casi clinici, al calcolo dei casi infetti e alla prevalenza di SARS-CoV-2. Una volta verificata la solidità delle metodologie, queste sono state applicate ad un secondo dataset relativo agli anni 2024–2025, impiegato come base di validazione su un periodo più recente ed esteso anche al virus influenza. Di seguito viene descritto in dettaglio il dataset del 2022 e le principali variabili utilizzate. Il dataset relativo agli anni 2024-25 verrà descritto successivamente nel capitolo *Validazione delle metodologie su dati aggiornati e su virus influenza*.

Il dataset analizzato comprende i dati raccolti nell'ambito della sorveglianza ambientale delle acque reflue, relativi a più impianti di depurazione lombardi, campionati giornalmente tra il 1° gennaio 2022 e il 31 dicembre 2022.

Gli impianti inclusi nello studio sono i seguenti: Bergamo, Bresso, Canegrate, Città di Cremona, Como, Lonate Pozzolo, Milano Nosedo, Milano San Rocco, Monza, Pavia, Peschiera Borromeo, Sondrio, Varese, Brescia-Verziano, Vigevano.

In *Tabella 3.1* vengono riportati gli AE nominali per sito di prelievo.

Sito di prelievo	AE nominali	Sito di prelievo	AE nominali
Bergamo	220.000	Monza	600.000
Bresso	220.000	Pavia	132.912
Canegrate	137.950	Peschiera Borromeo	566.000
Citta di Cremona	180.000	Sondrio	49.500
Como	196.000	Varese	74.402
Lonate Pozzolo	450.000	Verziano	296.000

Milano Nosedo	1.250.000	Vigevano	57.925
Milano San Rocco	1.036.000		

Tabella 3.1 – AE nominali per sito di prelievo

Per ciascun campione sono stati determinati, tramite RT-qPCR, i valori di concentrazione (in copie geniche/L) di tre sequenze geniche del SARS-CoV-2:

- N_1 ;
- N_3 ;
- ORF.

Questi dati rappresentano la carica virale grezza presente nel refluo al momento del campionamento, target molecolari appartenenti a regioni distinte del genoma a RNA di SARS-CoV-2.

Per ogni campione sono stati misurati anche parametri di qualità delle acque reflue, fondamentali per la stima della popolazione servita:

- Q (portata in m³/giorno);
- SS (solidi sospesi, mg/L);
- COD (domanda chimica di ossigeno, mg/L);
- BOD (domanda biochimica di ossigeno a 5 giorni, mg/L);
- N-tot (azoto totale, mg/L);
- N-NH₄ (azoto ammoniacale, mg/L);
- P (fosforo totale, mg/L).

Questi parametri sono di particolare importanza poiché ogni abitante contribuisce mediamente con un determinato carico di sostanza organica, solidi sospesi e nutrienti. Confrontando i valori misurati nel refluo con i valori medi pro capite riportati in letteratura, è possibile stimare gli abitanti equivalenti (AE) effettivamente collegati alla rete fognaria al momento del campionamento in maniera grezza.

Il dataset contiene anche dati sanitari:

- Nuovi casi: numero giornaliero di casi con diagnosi COVID-19 confermati da tampone;
- % tamponi positivi su tamponi totali.

Questi dati provengono dalla rete di sorveglianza clinica della Regione e sono alimentati dalle segnalazioni delle autorità sanitarie e dagli ospedali.

Per quanto riguarda gli abitanti equivalenti, sono presenti diverse stime di popolazione servita:

- Abitanti equivalenti da anagrafe dell'impianto: valore teorico che non tiene conto delle fluttuazioni temporanee dovute a turismo, pendolarismo o alla distinzione fra scarichi civili ed industriali; pertanto, può essere poco rappresentativo in alcuni contesti. È costante giorno per giorno;
- AE Brescia: calcolata secondo la metodologia sviluppata nello studio dell'Università di Brescia, che utilizza parametri chimici meno influenzabili dagli scarichi industriali (generalmente l'azoto totale). La stima è più rigorosa;
- AE_{COD} , AE_{BOD} , AE_{NTOT} : stime grezze degli abitanti equivalenti calcolate secondo il metodo WBE-SARI. A partire dalle concentrazioni misurate in ingresso all'impianto (mg/L), è stato stimato il corrispondente carico giornaliero mediante moltiplicazione per la portata. Il carico così ottenuto è stato quindi rapportato ai valori medi pro capite riportati in letteratura (60 $g_{BOD_5}/AE/giorno$, 110 $g_{COD}/AE/giorno$ e 12 $g_{Ntot}/AE/giorno$), ottenendo una stima della popolazione equivalente associata al carico organico misurato. Non è costante giorno per giorno, ma la correzione specifica per gli scarichi industriali non è presente, infatti, si tratta di una stima più rapida ma meno robusta rispetto al metodo Brescia.

È opportuno riportare le seguenti considerazioni sui dati:

- I dati di AE sono disponibili solo nei giorni di campionamento del refluo;
- I dati virologici, chimici ed epidemiologici sono raccolti con frequenza diversa: in particolare, questi ultimi sono disponibili 1-2 volte a settimana, mentre i dati epidemiologici sono giornalieri.

La tabella 3.2 riporta una descrizione del dataset originale, contenente l'elenco delle colonne, le relative unità di misura e una breve descrizione del loro significato:

Nome colonna	Unità di misura	Descrizione
Data	–	Giorno di campionamento
Impianto	–	Nome del depuratore in cui è stato prelevato il campione.
N_1	copie geniche/L	Numero di copie della sequenza genica N_1 del SARS-CoV-2 nel campione di refluo.

N ₃	copie geniche/L	Numero di copie della sequenza genica N ₃ del SARS-CoV-2 nel campione di refluo.
ORF	copie geniche/L	Numero di copie della sequenza genica ORF del SARS-CoV-2 nel campione di refluo.
Q	m ³ /giorno	Portata giornaliera in ingresso all'impianto.
SS	mg/L	Solidi sospesi totali presenti nel refluo.
COD	mg/L	Domanda chimica di ossigeno, indicatore della materia organica ossidabile presente.
BOD	mg/L	Domanda biochimica di ossigeno a 5 giorni, indicatore della frazione biodegradabile della materia organica.
N-tot	mg/L	Azoto totale presente nel campione di refluo.
N-NH ₄	mg/L	Azoto ammoniacale presente nel campione di refluo.
P	mg/L	Fosforo totale presente nel campione di refluo.
Nuovi casi	numero di casi/giorno	Numero di nuovi casi confermati tramite tampone nella popolazione servita.
Abitanti equivalenti da anagrafe	abitanti equivalenti	Stima teorica di popolazione servita, non corretta per fluttuazioni temporanee o scarichi civili/industriali.
AE Brescia	abitanti equivalenti	Stima calcolata secondo il metodo Brescia, basata su parametri meno influenzati da scarichi industriali (generalmente N-tot).
AE _{BOD}	abitanti equivalenti	Stima di AE ottenuta con il metodo WBE-SARI usando il carico di BOD (60 g _{BOD5} /AE/giorno)
AE _{COD}	abitanti equivalenti	Stima di AE ottenuta con il metodo WBE-SARI usando il carico di COD (110 g _{COD} /AE/giorno)
AE _{NTOT}	abitanti equivalenti	Stima di AE ottenuta con il metodo WBE-SARI usando il carico di N-tot (12 g _{Ntot} /AE/giorno)
Tamponi positivi	%	Percentuale di tamponi positivi rispetto ai totali eseguiti.

Tabella 3.2 – Descrizione del dataset 2022

La *tabella 3.3* riporta la numerosità dei target molecolari dell'anno 2022.

Gene	Campioni totali	Campioni con dato	Mancanti [%]
ORF	5 390	1 028	80.9
N ₁	5 390	842	84.4
N ₃	5 390	843	84.4

Tabella 3.3 – numerosità dei geni-2022

- ORF è il gene con il numero più elevato di misure disponibili;

- N_1 e N_3 mostrano una disponibilità quasi identica.

Una percentuale così elevata di dati mancanti riflette la frequenza di campionamento pari a 1-2 volte a settimana per impianto, come riportato sopra.

Il numero di target molecolari per impianto è mostrato nella *tabella 3.4*.

Impianto	Campioni totali	ORF	N_1	N_3
Bergamo	366	78	69	74
Bresso	365	86	33	34
Canegrate	365	47	0	0
Città di Cremona	365	85	85	93
Como	365	71	90	90
Lonate Pozzolo	365	75	45	45
Milano Nosedo	365	78	98	88
Milano San Rocco	365	74	96	86
Monza	365	89	74	75
Pavia	365	35	43	43
Peschiera Borromeo	365	84	34	34
Sondrio	365	46	36	36
Varese	367	46	1	0
Verziano (Brescia)	365	96	93	100
Vigevano	277	38	45	45

Tabella 3.4 – numerosità dei geni per impianto-2022

La numerosità totale dei campioni è molto omogenea tra gli impianti principali, e la numerosità dei geni non lo è per tutti gli impianti:

- alcuni impianti presentano una buona copertura per tutti i target (es. Como, Milano, Cremona);
- altri impianti mostrano una forte asimmetria o assenza di alcuni geni (es. Canegrate, Varese).

In generale, la diversa numerosità dei campioni all'interno del dataset per ciascun target molecolare influenza gli errori delle stime e la robustezza dei coefficienti, di conseguenza, il confronto tra modelli sviluppati su geni diversi deve essere effettuato con cautela.

Analogamente la disponibilità dei target tra i diversi impianti potrebbe incidere sulle analisi condotte a livello del sito di prelievo.

Per visualizzare meglio la distribuzione dei target molecolari per impianto vengono costruiti i seguenti boxplot, riportati in *Figura 3.1*, *Figura 3.2* e *Figura 3.3*.

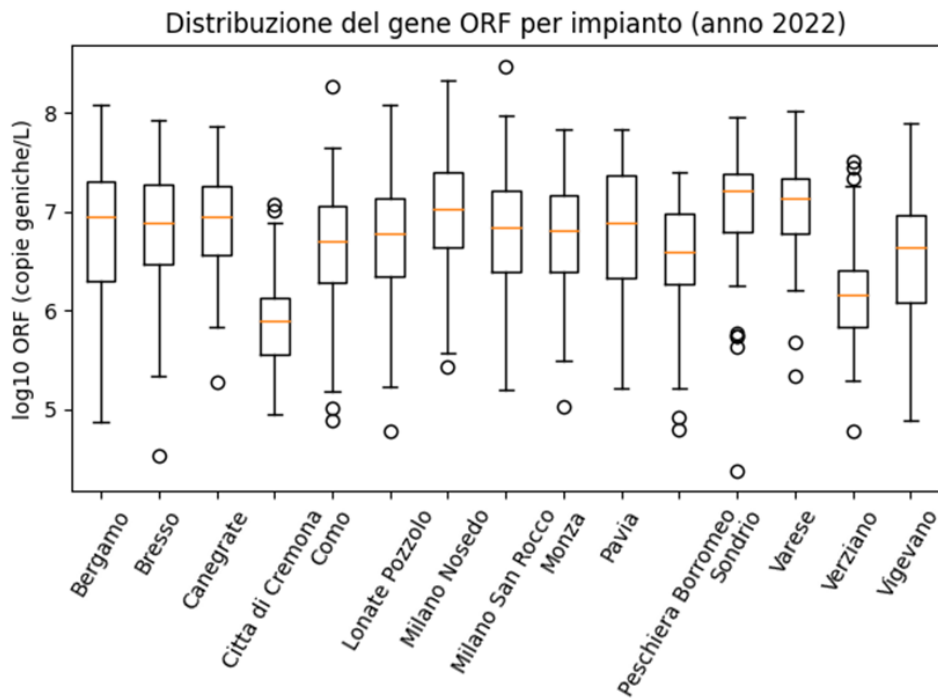


Figura 3.1 – boxplot gene ORF per impianto-anno 2022

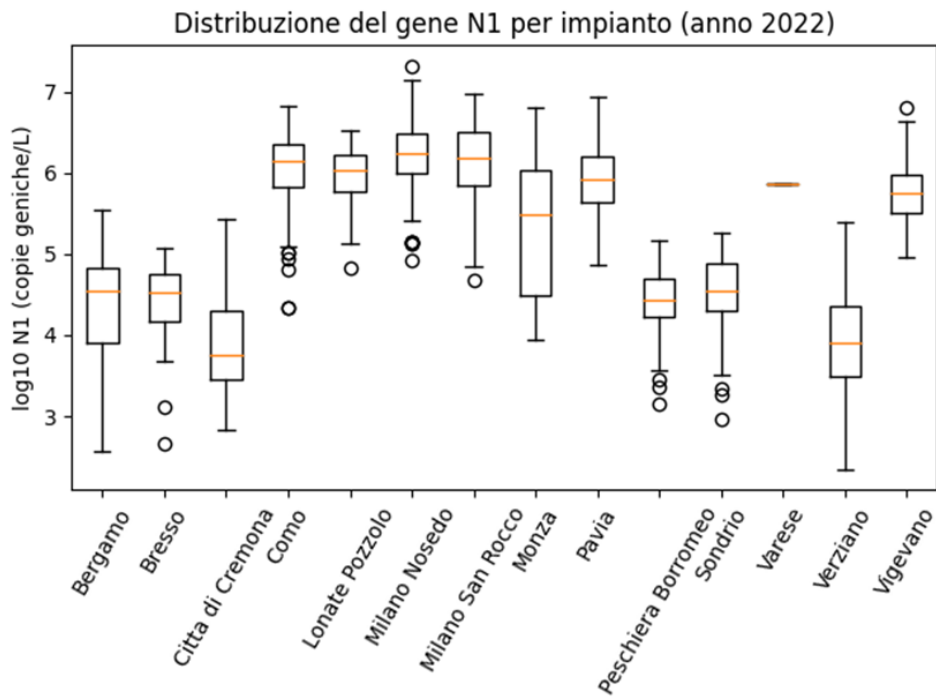


Figura 3.2 – boxplot gene N₁ per impianto-anno 2022

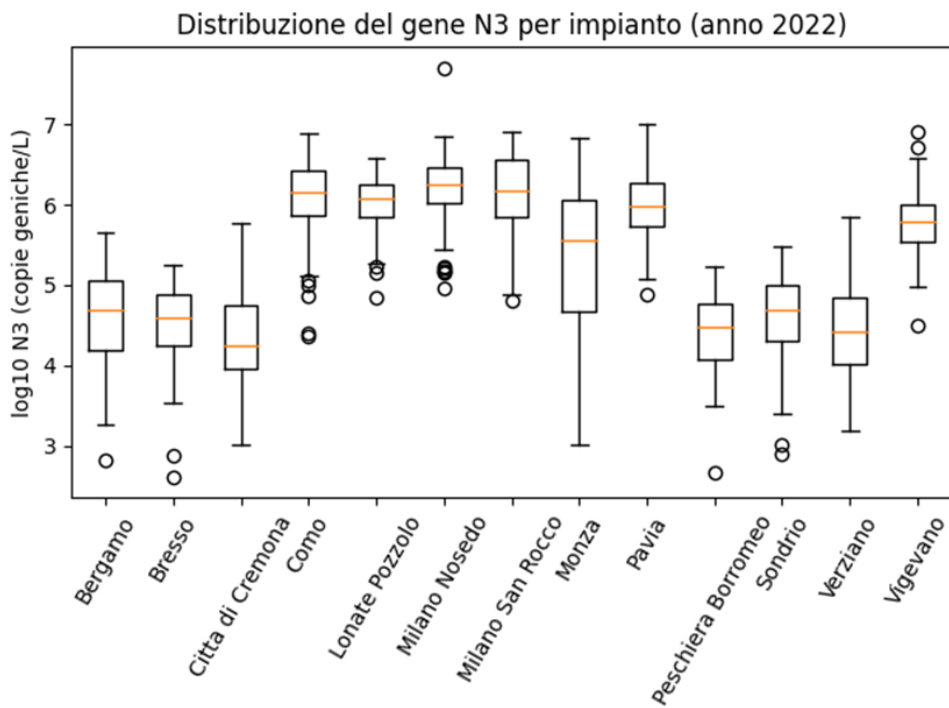


Figure 3.3 – boxplot gene N₃ per impianto-anno 2022

I boxplot evidenziano delle eterogeneità in termini sia di valore mediano che di variabilità delle concentrazioni virali tra impianti:

- I geni N_1 e N_3 hanno delle distribuzioni simili per quanto riguarda il valore mediano e la variabilità, mentre il gene ORF presenta dei valori mediani maggiori e in alcuni casi anche maggiore variabilità;
- In numerosi impianti si evidenzia un'asimmetria positiva della distribuzione, con baffi inferiori più estesi e presenza di outlier inferiori. La maggior parte delle osservazioni, dunque, si concentra su valori medio-bassi.

Vengono, infine, riportate in sintesi le principali statistiche descrittive per i tre geni ORF, N_1 ed N_3 , nella *Tabella 3.5*.

Per tutti e tre i geni, la media maggiore della mediana conferma l'asimmetria positiva delle distribuzioni e i valori massimi distanti dal valore mediano, probabilmente, dipendono dalla presenza di picchi epidemici.

Variabile	Media [copie/L]	Mediana [copie/L]	IQR	Min	Max
ORF	$1,22 \times 10^7$	$5,32 \times 10^6$	$1,32 \times 10^7$	$2,41 \times 10^4$	$3,00 \times 10^8$
N_1	$1,02 \times 10^6$	$2,25 \times 10^5$	$1,30 \times 10^6$	$2,23 \times 10^2$	$2,11 \times 10^7$
N_3	$1,04 \times 10^6$	$2,34 \times 10^5$	$1,31 \times 10^6$	$4,13 \times 10^2$	$5,08 \times 10^7$

Tabella 3.5 – statistiche descrittive variabili geni-anno 2022

3.2 Normalizzazioni applicate al caso studio

Le normalizzazioni sono state effettuate con l'obiettivo di identificare il metodo che meglio permette di correlare la carica virale presente nelle acque reflue con i nuovi casi di infezione. Questo approccio consente di distinguere le variazioni della carica virale dovute a reali cambiamenti nell'incidenza dell'infezione da quelle causate da fattori esterni, come fluttuazioni nella popolazione collegata all'impianto o nel volume delle acque reflue. Infatti, sebbene alcuni impianti abbiano dei valori mediani e variabilità simile, non è possibile confrontare le concentrazioni grezze poiché potrebbero essere influenzate da fattori diversi. È necessario, dunque, applicare la normalizzazione per rendere i dati più comparabili.

3.2.1 Normalizzazione tramite azoto ammoniacale (N)

La prima normalizzazione della carica virale grezza presente nel refluo, per tutti e tre i geni, viene effettuata tramite l'azoto ammoniacale (N) con la formula che segue:

$$\text{Carica virale normalizzata} = \frac{\text{carica grezza} \cdot 1000}{N/9.4} \quad [8]$$

Dove:

- Carica virale normalizzata: è espressa in (c.g./ab/giorno);
- Carica grezza: copie genomiche per litro di refluo (c.g./L);
- N: concentrazione di azoto ammoniacale (mgN/L);
- 9.4 gN/AE·giorno: carico pro capite medio di azoto ammoniacale;
- 1000: fattore di scala derivante dalla differenza fra la concentrazione espressa in mg/L e il coefficiente di emissione in g/d.

Il valore 9.4 proviene dallo studio di Bertanza e Boiocchi ed è la media del carico pro capite giornaliero di azoto ammoniacale calcolato su tutti gli impianti considerati. Questo valore permette di ottenere una misura della carica virale pro capite, rendendo i dati confrontabili tra impianti diversi e tra periodi differenti. Va sottolineato che tale valore si discosta dai valori tipici riportati in letteratura, generalmente compresi tra 12 e 15 g/(AE·giorno), evidenziando come la stima del carico pro capite possa variare in base alle caratteristiche locali dei reflui e dei sistemi fognari.

3.2.2 Normalizzazione tramite abitanti equivalenti (AE)

La seconda normalizzazione applicata nello studio viene effettuata tramite gli abitanti equivalenti stimati secondo la metodologia sviluppata da Brescia, per tutti e tre i geni.

La formula applicata per ottenere la carica virale normalizzata è la seguente:

$$\text{Carica virale normalizzata} = \frac{\text{carica grezza} \cdot Q}{AE \text{ brescia}} \quad [9]$$

Dove:

- Carica virale normalizzata: è espressa in (c.g./ab/giorno);
- Carica grezza: copie genomiche per litro di refluo (c.g./L);
- Q è la portata;
- AE Brescia: è la stima del numero di abitanti equivalenti per l'impianto eseguita da Brescia.

La metodologia UniBS prevede innanzitutto il calcolo delle produzioni pro capite medie dei principali parametri caratteristici (BOD, COD, N-tot e P-tot) a partire dai

carichi storici dell'impianto. Tali valori risultano generalmente inferiori a quelli normativi standard e coerenti con quelli riportati nello studio di Bertanza e Boiocchi (44,6; 82,4; 9,4; 1,08 g/ab/giorno). Il parametro che restituisce la produzione pro capite più bassa viene considerato il miglior indicatore della componente civile, in quanto presumibilmente meno influenzato da contributi di natura non domestica, e viene quindi utilizzato per la stima dinamica degli abitanti equivalenti. La stima giornaliera viene effettuata su carichi mediati mediante una media mobile 2+1 (media del giorno considerato e dei due adiacenti) al fine di ridurre la variabilità non epidemiologica. L'adozione di queste normalizzazioni consente di valutare quale approccio fornisca la migliore correlazione tra carica virale rilevata e nuovi casi.

Non esiste un metodo universalmente migliore: la scelta dipende dalle caratteristiche dell'impianto e dal tipo di variabilità non epidemiologica che si desidera ridurre.

Di seguito, in *Tabella 3.6* viene riportata una tabella di numerosità dei target molecolare, dopo la normalizzazione.

Gene	Carica grezza	Normalizzazione con N	Normalizzazione con AE
ORF	1028	378	340
N ₁	842	421	360
N ₃	843	409	361

Tabella 3.6 – numerosità dei geni grezzi e normalizzati-anno 2022

Teoricamente, una normalizzazione efficace dovrebbe ridurre la variabilità tra impianti dal punto di vista di diluizione o differente popolazione; se la variabilità osservata fosse soprattutto di natura epidemiologica, le normalizzazioni non dovrebbero modificare molto le distribuzioni.

3.3 Applicazione della statistica al caso studio

3.3.1 Analisi di correlazione

In questo paragrafo vengono mostrati i risultati delle analisi di correlazione tra i dati di sorveglianza ambientale e gli indicatori epidemiologici.

L'obiettivo è valutare se la concentrazione di SARS-CoV-2 nelle acque reflue rifletta l'andamento dei casi nella popolazione, e se questo possa effettivamente migliorare dopo aver applicato la normalizzazione.

In particolare, viene effettuata la correlazione di Pearson fra la carica virale grezza e normalizzata con i due metodi e i casi, per tutti i marcatori fecali (ORF, N_1 e N_3).

Nella *Tabella 3.7* viene riportata una tabella riassuntiva dei risultati delle correlazioni con r:

Gene	Nuovi casi
ORF grezzo	0.01
ORF normalizzato N	0.002
ORF normalizzato AE	-0.16
N_1 grezzo	0.55
N_1 normalizzato N	0.55
N_1 normalizzato AE	0.42
N_3 grezzo	0.29
N_3 normalizzato N	0.19
N_3 normalizzato AE	0.11

Tabella 3.7 – correlazione di Pearson fra dati clinici e ambientali-anno 2022

Nella *figura 3.4* si riportano gli scatter plot dei casi rispetto alle cariche, grezze e normalizzate.

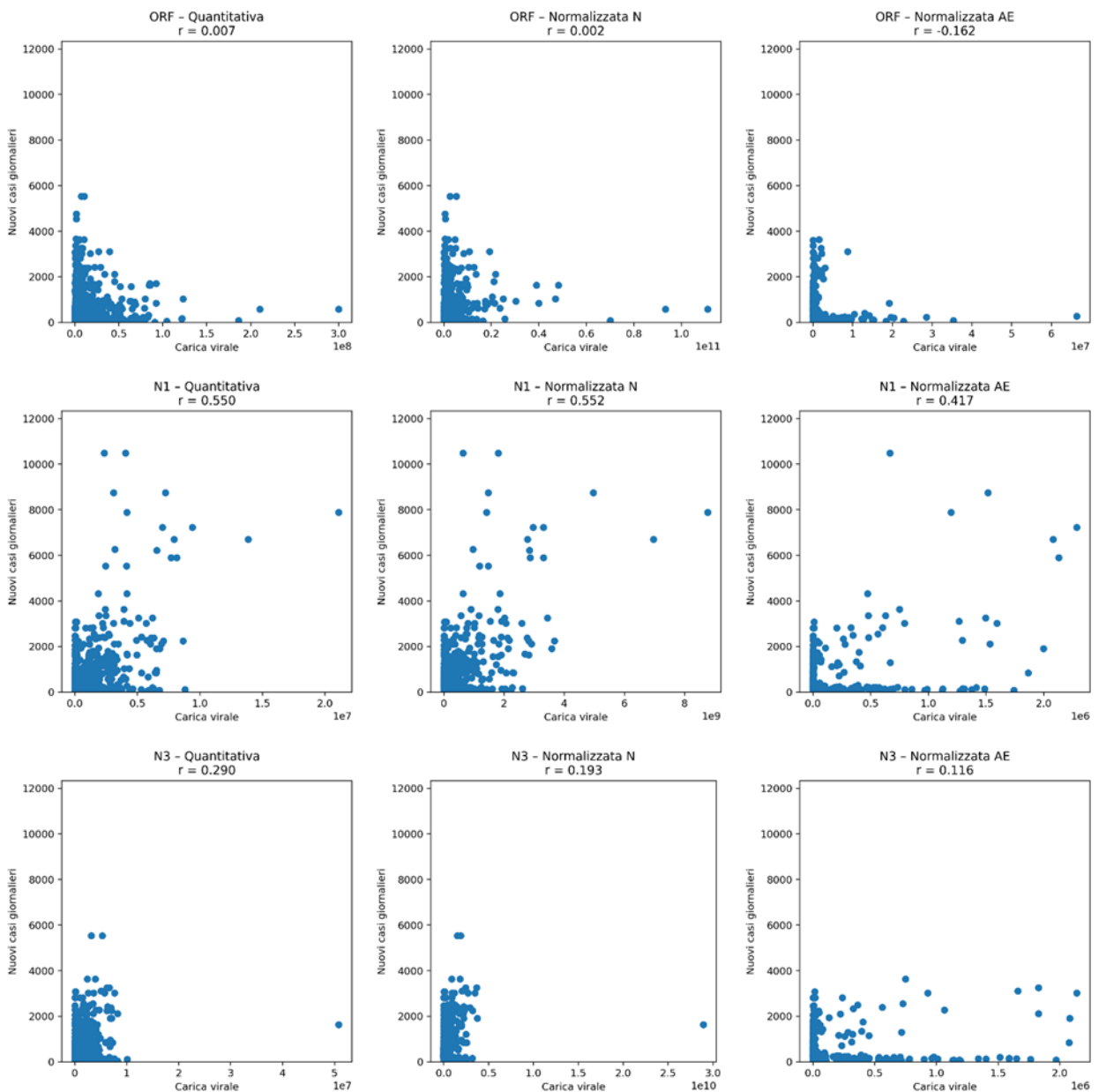


Figura 3.4 – scatter plot dei nuovi casi vs cariche grezze e normalizzate

Dai risultati ottenuti si può concludere che il gene N_1 mostra la relazione migliore con i nuovi casi giornalieri; infatti, presenta una correlazione moderata nella versione della carica grezza, la nuvola di punti suggerisce una tendenza crescente, anche se con della dispersione, che può essere plausibile con la variabilità biologica del segnale WBE.

Il gene N_3 presenta una correlazione più debole, con più dispersione e una relazione meno definita, mentre il gene ORF non evidenzia una correlazione significativa, con una minore sensibilità epidemiologica del target in questo contesto.

Il gene N_1 sembra essere quello più robusto.

Per quanto riguarda la normalizzazione con azoto ammoniacale, si può concludere che non si osserva un miglioramento apprezzabile della correlazione, rispetto alla carica grezza.

Questo comportamento suggerisce che il segnale grezzo non fosse influenzato molto dalla diluizione legato al carico organico e che il segnale epidemiologico fosse già abbastanza “pulito”.

La normalizzazione per gli abitanti equivalenti riduce, invece, la correlazione per tutti i geni; la riduzione della numerosità del campione sicuramente può influire sulla stabilità della stima e sulla significatività della correlazione. Inoltre, la stima degli abitanti equivalenti può introdurre rumore ulteriore dovuto a parametri chimici soggetti a fluttuazioni giornaliere e ci può essere un disallineamento con la reale dinamica epidemiologica.

3.4 Analisi di regressione

3.4.1 Modelli di regressione lineare semplice

Dopo aver valutato in fase preliminare la correlazione fra i dati ambientali e i dati sanitari, si è ritenuto opportuno valutare più approfonditamente la relazione fra i due, tramite l'impiego di modelli di regressione lineare semplice.

A differenza della correlazione, che fornisce esclusivamente una misura sintetica dell'associazione fra due variabili, la regressione lineare permette di modellare la relazione tra la carica virale nel refluo e i nuovi casi clinici.

Il modello di regressione lineare semplice assume la forma seguente:

$$Y = \beta_0 + \beta_1 * X + \varepsilon \quad [10]$$

Dove:

- Y è la variabile risposta, corrispondente ai nuovi casi clinici;
- X è la variabile indipendente, corrispondente alla carica virale nel refluo;
- β_0 è l'intercetta del modello;
- β_1 è il coefficiente di regressione;
- ε è l'errore.

Per completezza metodologica, sebbene il gene N_1 presenti la migliore correlazione con i nuovi casi, si procede eseguendo dei modelli di regressione semplice per tutti e tre i geni.

Nella Tabella 3.8 viene riportata una tabella riassuntiva dei parametri dei modelli di regressione lineare semplice.

Gene	Normalizzazione	N osservazioni	R ²	RMSE	Significatività β_1
N ₁	Grezzo	842	0.30	988	p<0.001
N ₁	Normalizzato N	421	0.30	1266	P<0.001
N ₁	Normalizzato AE	360	0.174	1096	P<0.001
N ₃	Grezzo	843	0.084	1150	P<0.001
N ₃	Normalizzato N	409	0.037	1400	P<0.01
N ₃	Normalizzato AE	361	0.013	1250	P=0.05
ORF	Grezzo	1028	0.00005	1200	n.s.
ORF	Normalizzato N	378	0.000004	1500	n.s.
ORF	Normalizzato AE	340	0.026	1100	P<0.05

Tabella 3.8 – parametri principali dei modelli di regressione lineare semplice-anno 2022

Dall'analisi dei modelli di regressione lineare semplice si evidenzia come il gene N₁, nella forma grezza, presenti la miglior capacità esplicativa, seguito dalla forma normalizzata con N.

I geni ORF e N₃ hanno una capacità predittiva inferiore, confermando ciò che era emerso dall'analisi esplorativa della correlazione.

Si sottolinea che in tutti i modelli di regressione lineare, non tutte le ipotesi sono soddisfatte:

- La relazione tra le variabili è approssimativamente lineare solo per il gene N₁, e la dispersione dei punti suggerisce che ci possono essere dei fattori che influenzano la relazione o la presenza di una possibile componente non lineare;
- Il test di Durbin–Watson ha restituito valori compresi tra 1,00 e 1,77 nei diversi modelli indicando una tendenza ad autocorrelazione positiva dei residui. Tale evidenza suggerisce una possibile dipendenza temporale non completamente catturata dal modello, coerente con la continuità dei fenomeni epidemiologici e dei segnali WBE. Tuttavia, i valori osservati non indicano una violazione estrema e l'analisi tramite regressione lineare è stata mantenuta a fini descrittivi e comparativi tra le diverse normalizzazioni, segnalando che l'eventuale autocorrelazione può influenzare l'affidabilità dei test di significatività (stima delle incertezze) più che i valori puntuali dei coefficienti;

- Dall'analisi dei residui si osserva una maggiore dispersione per valori elevati dei casi e questo può suggerire una possibile eteroschedasticità;
- I Q-Q plot evidenziano un discostamento dalla normalità, soprattutto nelle code della distribuzione e questo può essere attribuito alla presenza di outlier epidemiologici.

Sebbene non tutte le ipotesi siano soddisfatte, i modelli vengono impiegati con finalità esplorative e comparative dei geni, per valutare l'associazione lineare fra carica virale e nuovi casi.

Modello di regressione lineare semplice N_1 quantitativo-nuovi casi

Nella *Figura 3.5* viene riportato il diagramma di regressione:

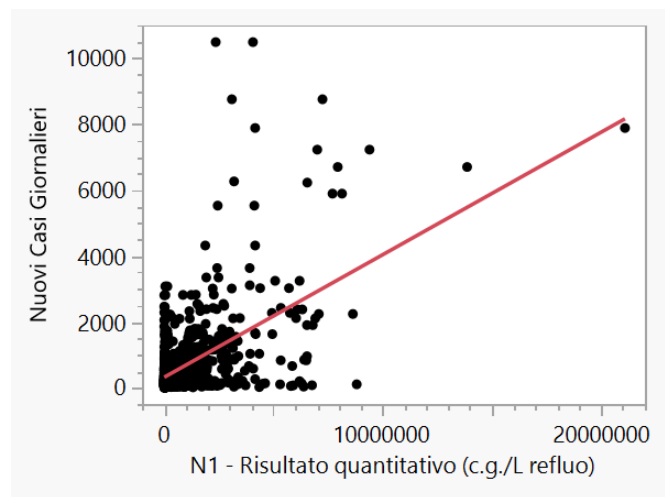


Figura 3.5 – diagramma di regressione-modello di regressione lineare semplice N_1 quantitativo-nuovi casi-anno 2022

Si evidenzia una relazione lineare positiva tra la concentrazione del gene N_1 e i nuovi casi, il coefficiente β_1 è positivo, il che indica, appunto, che all'aumentare della carica virale corrisponde un aumento dei casi.

Il modello spiega circa il 30% della variabilità, il che indica una relazione moderatamente forte e la rilevanza della varianza non spiegata.

Nella *Figura 3.6* viene riportato il Q-Q plot.

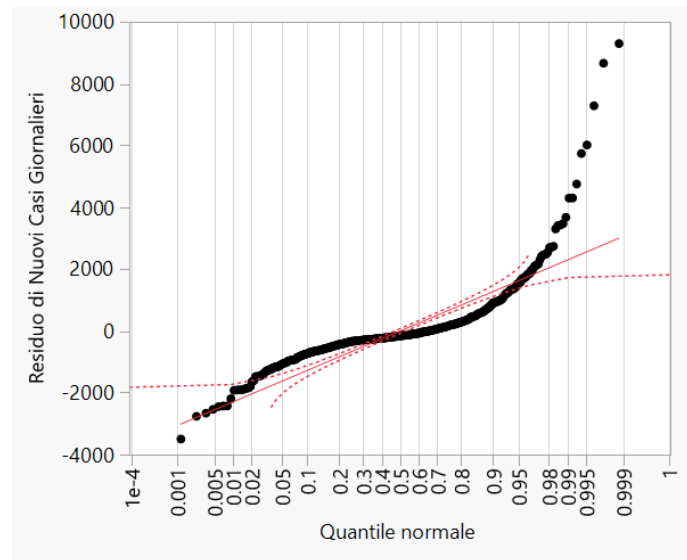


Figura 3.6 – diagramma dei quantili normali residui-modello di regressione lineare semplice N_1 quantitativo-nuovi casi-anno 2022

Il Q-Q plot evidenzia deviazioni della normalità nelle code della distribuzione con presenza di outlier; questo è compatibile con la natura asimmetrica delle cariche e anche dei casi giornalieri.

In *Figura 3.7* viene mostrato il diagramma dei residui rispetto ai valori previsti:

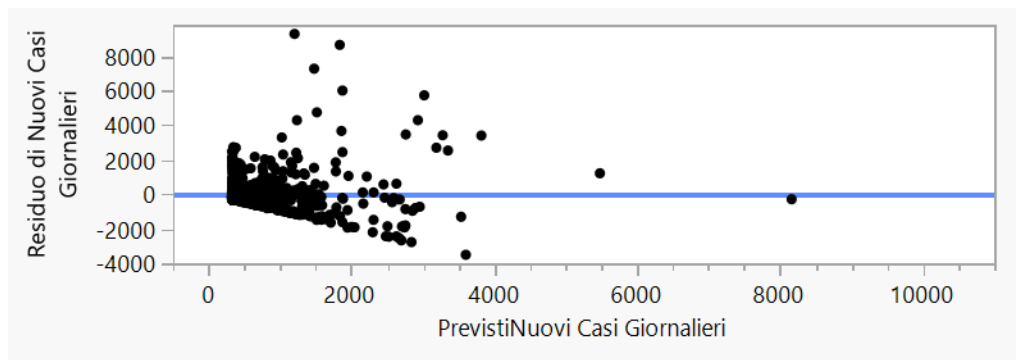


Figura 3.7 – diagramma dei residui vs previsti-modello di regressione lineare semplice N_1 quantitativo-nuovi casi-anno 2022

Si osserva una forma ad imbuto dei residui, la cui variabilità cresce per valori maggiori di nuovi casi previsti; questo suggerisce un possibile fenomeno di eteroschedasticità ed è coerente con la dinamica epidemica, caratterizzata da maggiore variabilità nelle fasi di picco dei casi.

Alcuni punti, dunque, mostrano valori estremi, associati a picchi epidemici. Tali punti non sono stati esclusi dall'analisi in quanto rappresentano eventi reali. Inoltre, il confronto tra R^2 e R^2 PRESS non evidenzia variazioni sostanziali, indicando che il

modello non è guidato da singole osservazioni influenti e che la relazione stimata risulta stabile rispetto alla rimozione di ciascuna osservazione (leave-one-out).

3.4.2 Modelli di regressione lineare multipla

Dopo aver analizzato i modelli di regressione lineare semplice, si è ritenuto opportuno utilizzare dei modelli di regressione lineare multipla che consentono di controllare fattori confondenti, quali il sito di prelievo e la variante virale, migliorando la capacità di spiegare la variabilità osservata nei casi clinici.

Il modello assume la formula seguente:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 \dots + \beta_K * X_k + \epsilon \quad [11]$$

Dove:

- Y è la variabile risposta che rappresenta i nuovi casi clinici;
- X_1 è la carica virale;
- $X_2 \dots X_k$ sono i predittori categoriali ovvero i fattori sito di prelievo e variante.

I modelli vengono eseguiti per tutti e tre i geni poiché questi ultimi potrebbero essere più stabili con l'aggiunta di fattori confondenti, sebbene ci si aspetti che il gene N_1 sia il più robusto.

In particolare, sono state valutate:

- Tutte le cariche virali quantitative in relazione ai nuovi casi clinici;
- Tutte le cariche virali normalizzate N in relazione ai nuovi casi clinici;
- Tutte le cariche virali normalizzate rispetto agli AE in relazione ai nuovi casi clinici.

In ciascuno di questi modelli sono stati inseriti due predittori categoriali:

- Sito di prelievo;
- Variante dominante.

Nella *Tabella 3.9* viene riportata una tabella riassuntiva dei parametri dei modelli di regressione lineare multipla.

Gene	Normalizzazione	n	R ²	RMSE	Carica	Variante	Sito	DW	Hp violate
N ₁	Quantitativo	842	0.78	560	P<0.001	P<0.001	P<0.001	1.14	Autocorr, eterosched
N ₁	Normalizzato N	421	0.76	748	P<0.001	P<0.001	P<0.001	1.14	Autocorr, eterosched
N ₁	Normalizzato AE	360	0.79	551	P<0.001	P<0.001	P<0.001	1.60	Autocorr moderata
N ₃	Quantitativo	843	0.65	420	P<0.001	P<0.001	P<0.001	1.03	Autocorr. forte
N ₃	Normalizzato N	409	0.56	573	P=0.006	P<0.001	P<0.001	0.96	Autocorr. forte
N ₃	Normalizzato AE	361	0.76	335	P<0.001	P<0.001	P<0.001	1.28	Autocorr.
ORF	Quantitativo	1028	0.67	406	P<0.001	P<0.001	P<0.001	1.22	Autocorr.
ORF	Normalizzato N	378	0.60	603	P<0.005	P<0.001	P<0.001	1.02	Autocorr. forte
ORF	Normalizzato AE	340	0.75	358	P=0.845	P<0.001	P<0.001	1.24	Autocorr.

Tabella 3.9 – parametri principali dei modelli di regressione lineare multipla (con sito e variante) -anno 2022

Dai risultati dei modelli si può concludere che:

- Le variabili categoriali aumentano la capacità esplicativa di tutti i modelli, rispetto ai modelli di regressione lineare semplice, evidenziando come parte della variabilità dei nuovi casi sia attribuibile sia ai siti che alle varianti;

- Il gene N_1 sembra essere il più robusto;
- Nonostante il miglioramento della capacità esplicativa, le ipotesi di autocorrelazione ed eteroschedasticità sono quasi sempre violate, coerentemente con la natura dei dati epidemiologici.

Nonostante il modello N_1 normalizzato AE mostri l' R^2 maggiore, nel corpo principale si sceglie di approfondire il modello N_1 quantitativo; la normalizzazione per gli AE, infatti, comporta una riduzione delle osservazioni e introduce una variabilità aggiuntiva legata agli abitanti equivalenti.

Modello di regressione lineare multipla N_1 quantitativo-nuovi casi

Nella *Tabella 3.10* è riportato il riepilogo della stima

N	842
R^2	0.78
R^2 aggiustato	0.77
RMSE	560.57

Tabella 3.10 – riepilogo della stima-modello di regressione lineare multipla N_1 quantitativo-nuovi casi (sito e variante) -anno 2022

Nella *Tabella 3.11* è mostrata la stima dei parametri

Variabile	β	p-value	Significativa
Intercetta	1271.2681	<0.0001	si
N_1 quantitativo (c.g./L)	0.0001347	<0.0001	si
Variante Dominante [BQX]	-904.3501	<0.0001	si
Variante Dominante [Omicron BA.1]	2827.0104	<0.0001	si
Variante Dominante [Omicron BA.2]	-863.7521	<0.0001	si
Sito [Bergamo]	-205.4261	0.0092	si
Sito [Bresso]	-70.3036	0.4911	no
Sito [Città di Cremona]	-210.6059	0.0040	si
Sito [Como]	-448.2219	<0.0001	si
Sito [Lonate Pozzolo]	-155.4118	0.0863	no

Sito [Milano Nosedo]	988.3763	<0.0001	s
Sito [Milano San Rocco]	1042.5459	<0.0001	si
Sito [Monza]	298.9708	<0.0001	si
Sito [Pavia]	-351.6168	0.0002	si
Sito [Peschiera Borromeo]	83.1730	0.4100	no
Sito [Sondrio]	-328.1437	0.0009	si
Sito [Varese]	-126.0709	0.8097	no
Sito [Verziano]	-96.4264	0.1809	no

Tabella 3.11 – tabella stima dei parametri-modello di regressione lineare multipla N_1 quantitativo-nuovi casi (sito e variante) -anno 2022

Dai risultati del modello si può affermare che:

- Tutte le covariate (carica, sito e variante) sono significative;
- Per quanto riguarda il sito, l'effetto risulta essere eterogeneo: minore numerosità dell'impianto, maggiore variabilità interna ai dati di quel sito, assenza di uno scostamento rispetto alla baseline;
- Il test globale F del modello è significativo;
- Il test di mancata stima è significativo, il che suggerisce che la relazione tra carica e nuovi casi non sia perfettamente lineare oppure che la variabilità temporale non sia completamente catturata dal modello;
- Nessun sito viene eliminato nell'analisi poiché la covariata sito è significativa;

Di seguito, in *Figura 3.8* viene riportato il grafico dei valori predetti dal modello rispetto a quelli osservati.

I casi osservati e predetti sono stati aggregati giornalmente (eseguendo una somma) sull'insieme degli impianti per rappresentare graficamente l'andamento temporale complessivo.

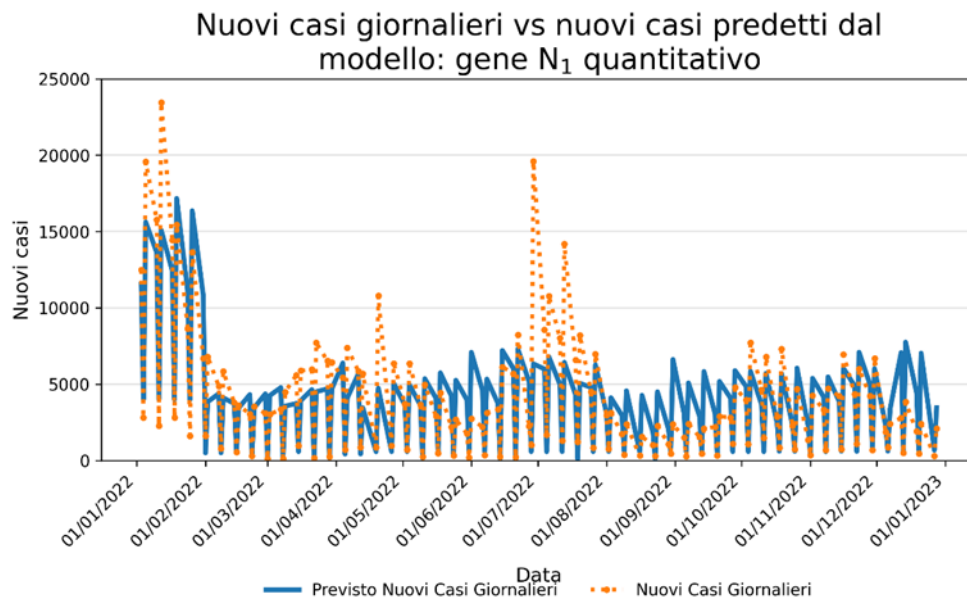


Figura 3.8 – nuovi casi giornalieri vs nuovi casi predetti dal modello: gene N_1 quantitativo-nuovi

Il grafico temporale dei valori osservati e predetti evidenzia una buona capacità del modello di seguire l'andamento generale delle onde epidemiche, però i picchi più estremi appaiono parzialmente sottostimati. Tale comportamento è coerente con la presenza di dinamiche non lineari tipiche dei fenomeni epidemici.

In *Figura 3.9* è riportato il diagramma dei residui rispetto ai previsti:

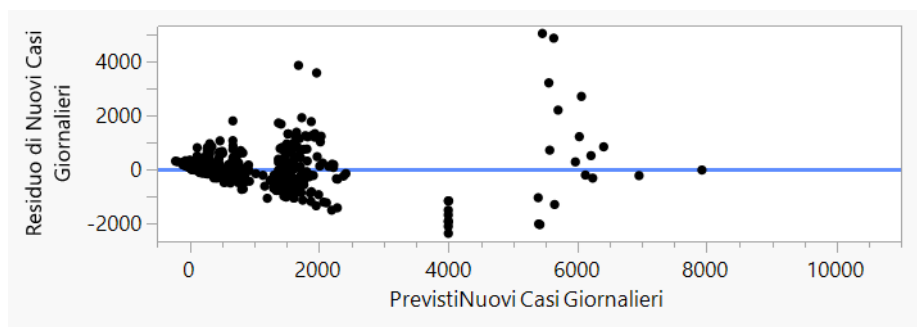


Figura 3.9 – diagramma dei residui rispetto ai previsti-modello di regressione lineare multipla N_1 quantitativo-nuovi casi (sito e variante) -anno 2022

Dal grafico dei residui si osserva una dispersione non completamente casuale attorno allo zero, con un aumento della variabilità al crescere dei valori previsti, suggerendo la possibile presenza di eteroschedasticità. Tale condizione suggerisce che la varianza dell'errore non sia costante, con possibili ripercussioni sull'affidabilità dell'inferenza statistica. La presenza di alcuni residui elevati, invece, può essere associata a picchi epidemici.

In Figura 3.10 viene mostrato il Q-Q plot:

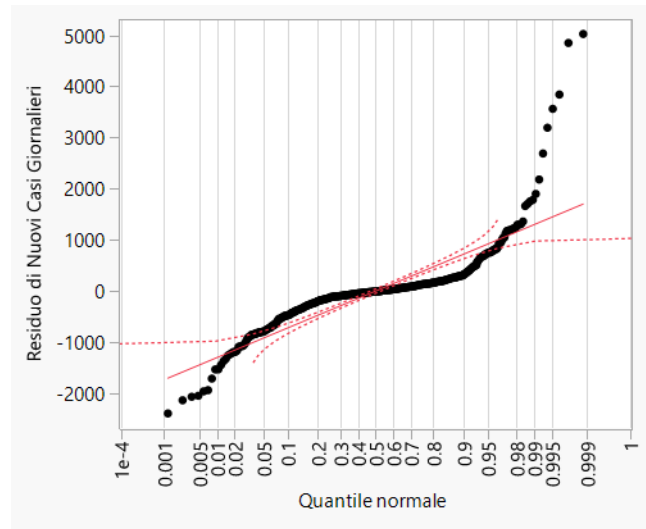


Figura 3.10 – diagramma dei quantili normali residui-modello di regressione lineare multipla N_1 quantitativo-nuovi casi (sito e variante) -anno 2022

Il Q-Q-plot evidenzia un discostamento dalla distribuzione normale in corrispondenza delle code della distribuzione dei residui, coerente con la presenza di picchi epidemici e di valori estremi. La parte centrale dei residui, invece, segue una distribuzione normale.

Non sono stati rimossi punti estremi dal dataset poiché i valori residui sono coerenti con i picchi epidemici; inoltre, la statistica PRESS mostra valori di R^2 simili a quello ordinario, il che significa che il modello non è influenzato da singole osservazioni.

In *Appendice* sono riportati i grafici di confronto fra i nuovi casi giornalieri e predetti dal modello, per il gene N_1 normalizzato.

3.5 Stima dei casi da carica virale e prevalenza

Dopo aver eseguito i modelli di regressione lineare, si è preceduto con la stima del numero di infetti tramite lo shedding rate, del solo gene N_1 poiché ritenuto il più robusto.

Di seguito viene riportato il valore di shedding rate utilizzato per le analisi per il gene N_1 :

$$\text{Shedding rate gene } N_1 = 7.30 \log_{10} \text{ c. g./g feci}$$

Si considerano 300 g di feci medie giornaliere per adulto come valore sensato per poter esprimere in maniera coerente lo shedding rate in relazione agli altri dati disponibili.

$$\text{Shedding rate gene } N_1 = 6 * 10^9 \text{ c. g./individuo infetto/Giorno}$$

È importante sottolineare che il valore di shedding rate sopra riportato non è riferito ad una particolare variante dominante del SARS-CoV-2.

Il numero di casi infetti si calcola tramite le seguenti formule:

$$\text{Carico } N_1 = N_1 \text{ quantitativo} * Q * 1000 \quad [12]$$

Dove:

- Carico N_1 : è espresso in [c.g./giorno];
- N_1 quantitativo in [c.g./L];
- Q è la portata dell'impianto di depurazione [m³/giorno].

$$\text{Numero casi infetti} = \frac{\text{carico } N_1}{\text{shedding rate } N_1} \quad [13]$$

Dove:

- Carico N_1 è espresso in [c.g./giorno];
- Shedding rate N_1 è espresso in [c.g./persona infetta/giorno];

È stato, quindi, analizzato l'andamento temporale degli infetti stimati nel tempo, per ciascun impianto.

A livello preliminare, l'andamento dei casi stimati tramite shedding rate (WBE) risulta complessivamente coerente con quello dei casi clinici: i picchi e i cali principali sono simili nella maggior parte degli impianti.

La curva dei casi WBE, però, risulta spesso più frastagliata rispetto a quella clinica poiché considera anche i casi non testati e/o asintomatici e risponde prima dei dati clinici ai cambiamenti rapidi. Questo perché lo shedding rate inizia prima della comparsa dei sintomi mentre i dati clinici arrivano in ritardo poiché la comparsa dei sintomi si verifica dal giorno tre.

Inoltre, spesso, la scala dei casi WBE non è confrontabile con quella dei casi clinici in alcuni impianti, probabilmente a causa dello shedding rate utilizzato, possibili errori nella popolazione servita o campionamenti discontinui.

Queste analisi preliminari, però, confermano che il segnale WBE contiene informazione epidemiologica, giustificando l'uso dei modelli nel capitolo successivo.

I grafici di confronto fra il numero di casi infetti N_1 tramite WBE e quelli clinici vengono riportati in *Appendice*.

Dopo aver calcolato il numero di infetti giornalieri, per impianto, si trasformano tali valori in termini di prevalenza, ovvero in percentuale della popolazione che risulta infetta in quel dato giorno.

In questo modo sarà possibile confrontare questa percentuale derivante dall'analisi del refluo con quella di tamponi positivi N_1 clinici.

La prevalenza viene ricavata tramite la seguente formula:

$$\text{prevalenza WBE} = \frac{\text{numero casi infetti } N_1}{AE \text{ media}} * 100 \quad [14]$$

Dove:

- La prevalenza WBE è espressa in termini di percentuale [%];
- AE media è la stima del numero di abitanti equivalenti per l'impianto eseguita da Brescia;
- Il fattore 100 viene impiegato per esprimere la prevalenza in termini di percentuale.

3.6 Modelli di regressione lineare multipla per valutare la prevalenza WBE

Dopo aver calcolato il numero di infetti tramite lo shedding rate e la prevalenza, si è proceduto con l'utilizzo di modelli di regressione lineare multipla per valutare la relazione presente tra la stima della percentuale degli infetti eseguita tramite tampone e quella stimata tramite monitoraggio WBE, calcolata tramite la prevalenza.

I modelli mostrano quanto bene la prevalenza clinica è spiegata dalla prevalenza WBE e quali fattori sono significativi.

Per poter effettuare questo tipo di analisi, è stata calcolata la media settimanale delle variabili per le quali erano disponibili più valori nella stessa settimana per lo stesso impianto, mantenendo una struttura "settimana per impianto". Il numero di impianti per settimana risulta essere relativamente stabile, evitando che le settimane con maggiore numerosità abbiano un'influenza maggiore sul modello.

Vengono eseguiti tre modelli con la seguente forma:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 \dots + \beta_K * X_k + \varepsilon \quad [15]$$

Dove:

- Y è la variabile risposta che rappresenta, nei tre modelli, la % di tamponi positivi N_1 al tempo t , la % di tamponi positivi N_1 anticipati di una settimana e la % di tamponi positivi N_1 anticipati di due settimane;
- X_1 è la prevalenza WBE;
- $X_2...X_k$ sono i predittori categoriali, ovvero la stagione e l'interazione fra la stagione e l'influenza, quando la percentuale di quest'ultima risulta maggiore del 3%.

La covariata legata all'influenza è stata inserita per poter valutare la relazione fra i due virus; si è scelto di utilizzare, come soglia di significativa presenza di influenza, una prevalenza del 3%.

Di seguito, in *Tabella 3.12* viene riportato il riassunto dei parametri più importanti dei modelli di regressione lineare multipla.

Modello	n	R ²	RMSE	Prevalenza WBE	Stagioni	Interazione influenza	DW	Hp violate
Nessun anticipo	495	0.29	5.28	significativa	Globalmente significativa	Globalmente significativa	0.59	autocorrelazione
Anticipo 1 settimana	481	0.40	4.81	significativa	Globalmente significativa	Globalmente significativa	0.71	autocorrelazione
Anticipo 2 settimane	468	0.50	4.36	significativa	significativa	significativa	0.88	autocorrelazione

Tabella 3.12-- parametri principali dei modelli di regressione lineare multipla-gene N_1 -prevalenza (stagione ed interazione influenza) -anno 2022

Dalle analisi dei modelli si può concludere che:

- La capacità esplicativa dei modelli migliora con l'aumentare dell'anticipo temporale, con riduzione dell'errore di stima; dunque, il segnale ambientale migliora quando si tiene conto del ritardo temporale fra eliminazione del virus e diagnosi clinica;
- La prevalenza WBE è sempre significativa, quindi conferma la presenza dell'informazione epidemiologica rilevante nel segnale ambientale;

- Le stagioni sono globalmente significative, e anche l'interazione con l'influenza;
- I modelli hanno tutti autocorrelazione positiva nei residui, (0.59, 0.71 e 0.88) che si attenua con l'anticipo temporale delle settimane.

Il modello con anticipo di due settimane risulta essere il migliore e viene riportato di seguito, mentre gli altri modelli si trovano in *Appendice*.

Modello di regressione lineare multipla: % tamponi positivi N_1 - anticipo 2 settimane

In *Tabella 3.13* è riportato il riepilogo della stima:

R ²	0.50
R ² corretto	0.49
RMSE	4.36
N osservazioni	468

Tabella 3.13-- riepilogo della stima-modello di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane -anno 2022

In *tabella 3.14* vengono riportate le stime dei parametri:

Termine	Stima	Errore std	Statistica T	p-value
Intercetta	21.324637	0.296299	71.97	<.0001*
Autunno inverno	6.5351744	0.489081	13.36	<.0001*
Estate	2.2521322	0.475792	4.73	<.0001*
Fine estate	-2.471387	0.443504	-5.57	<.0001*
Inverno	-3.693996	0.47217	-7.82	<.0001*
Prevalenza WBE	0.1940716	0.026822	7.24	<.0001*
Autunno inverno × Influenza>3%	-3.213996	0.986324	-3.26	0.0012*
Estate × Influenza>3%	-6.785842	1.711979	-3.96	<.0001*

Fine estate × Influenza>3%	0	-	-	-
Inverno × Influenza>3%	3.9017016	1.054816	3.70	0.0002*

Tabella 3.14--tabella stima dei parametri -modello di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane-anno 2022

L'interazione fine estate × Influenza>3% è nulla poiché l'influenza non è presente in questa stagione.

Nessuna covariata viene rimossa dal modello poiché globalmente tutte le covariate sono significative.

Di seguito, in *Figura 3.11*, viene riportato il grafico dei valori osservati rispetto a quelli predetti dal modello, facendo riferimento sia alla serie regionale della % dei tamponi positivi sui tamponi totali, sia alla media settimanale di tutti gli impianti presenti nella rete di sorveglianza.

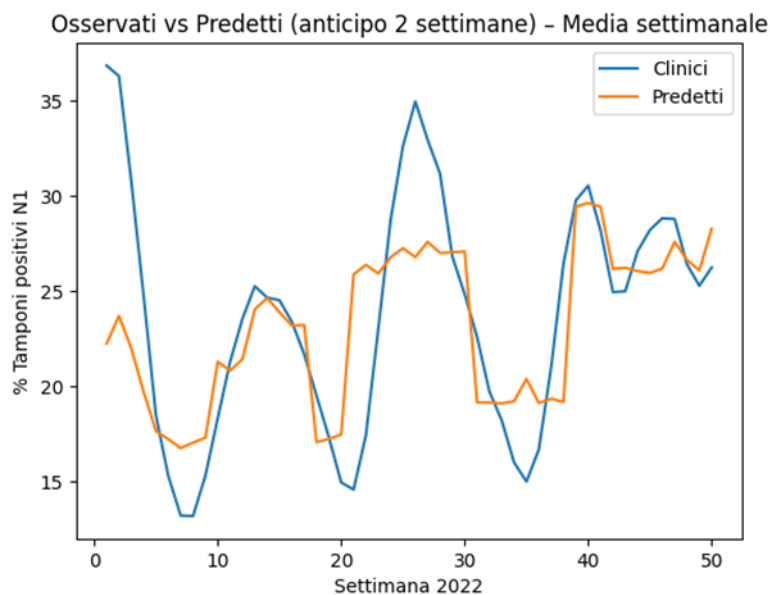


Figura 3.11 – % tamponi positivi N_1 osservati vs predetti-modello di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) - anticipo 2 settimane-media settimanale-anno 2022

Dal grafico si può notare come il modello riproduca bene l'andamento generale della percentuale di tamponi positivi N_1 , mantenendo un anticipo temporale di 14 giorni.

I punti di salita e discesa vengono generalmente intercettati.

La curva del modello è più smussata rispetto a quella osservata verosimilmente a causa dell'effetto di media per impianto e della natura lineare del modello, che tende a catturare l'andamento medio del fenomeno ma a sottostimare i valori estremi, in particolare durante i picchi epidemici.

In *Appendice* viene riportato il grafico senza l'applicazione della media settimanale, per vedere la variabilità spaziale fra i siti.

In *Figura 3.12* viene mostrato il diagramma dei residui rispetto ai previsti:

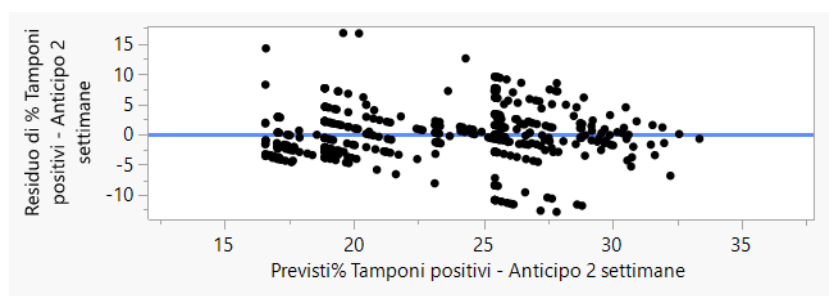


Figura 3.12 – diagramma dei residui rispetto ai previsti-modello di regressione lineare multipla -% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anno 2022

Il grafico mostra che la distribuzione dei residui è abbastanza simmetrica attorno allo zero e non ci sono pattern visibili; i residui più estremi sono coerenti con fasi di rapida crescita o decrescita epidemica.

Sebbene siano presenti alcuni residui di entità elevata, la prossimità tra R^2 e R^2 PRESS suggerisce che tali osservazioni non esercitano un'influenza dominante sulla stima dei coefficienti.

In *Figura 3.13* viene mostrato il Q-Q-plot:

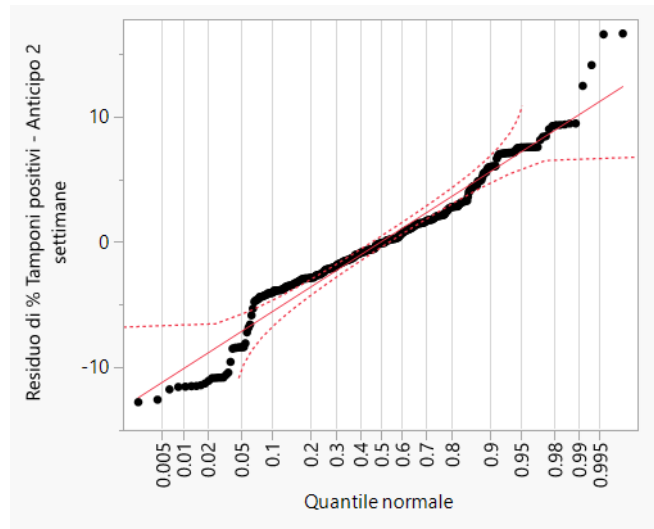


Figura 3.13 – diagramma dei quantili normali residui-modello di regressione lineare multipla -% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione con influenza) -anno 2022

Il grafico mostra una non perfetta normalità dei residui nelle code, tipico nei dati epidemiologici dove i picchi producono residui estremi.

Pur in presenza di lievi violazioni delle ipotesi, il modello risulta adeguato a descrivere la relazione fra prevalenza WBE e percentuale di tamponi positivi.

3.7 Validazione delle metodologie su dati aggiornati e su virus influenza

3.7.1 Descrizione del dataset 24-25

Il dataset 2024–2025, come già accennato nel sottocapitolo *Descrizione del dataset 2022*, presenta una struttura analoga a quella del dataset principale, con l’aggiunta di alcune variabili relative alla sorveglianza dei virus influenzali (Influenza A ed Influenza B).

Questi parametri, riportati come concentrazioni geniche rilevate nei campioni, permettono di estendere le analisi sviluppate per SARS-CoV-2 anche ad altri patogeni di interesse stagionale.

La *Tabella 3.15* riporta una descrizione del dataset 2024-2025:

Nome colonna	Unità di misura	Descrizione
N_1	copie geniche/L	Numero di copie della sequenza genica N_1 del SARS-CoV-2 nel campione di refluo.
Influenza A	copie geniche/L	Numero di copie geniche dell'influenza A nel campione di refluo
Influenza B	copie geniche/L	Numero di copie geniche dell'influenza B nel campione di refluo
Tamponi positivi COVID-19	%	Percentuale di tamponi positivi rispetto ai totali eseguiti
Tamponi positivi influenza	%	Percentuale di tamponi positivi all'influenza generica rispetto ai totali eseguiti

Tabella 3.15— descrizione del dataset 2024-2025

È importante sottolineare che la percentuale di tamponi clinici influenzali non distingue l'influenza A e l'influenza B.

Il valore di shedding rate che viene impiegato per il calcolo del numero di infetti per l'influenza è pari a $2.73 \cdot 10^9$ c.g./L [14].

Il valore di shedding rate per il gene N_1 , invece, rimane invariato; dunque, risulta essere pari a $6 \cdot 10^9$ c.g./L.

Anche in questo caso, per il calcolo della prevalenza WBE in percentuale, vengono calcolati gli abitanti equivalenti AE medi, secondo il metodo di Brescia.

Per semplicità, per ogni impianto, l'azoto viene utilizzato come parametro di riferimento per stimare gli AE.

Di seguito, in *Tabella 3.16* vengono riportate delle tabelle descrittive del dataset e delle variabili:

n osservazioni	807
Numero Impianti	15
Numero settimane	59
Min settimane per impianto	9
Max settimane per impianto	48

Tabella 3.16— tabella descrittiva del dataset 2024-2025

Il periodo considerato si estende dalla settimana 36 del 2024 alla settimana 44 del 2025. Tuttavia, non tutte le settimane intermedie risultano disponibili nei dati; per ciascun impianto sono presenti fino a 48 settimane effettive di osservazione.

La distribuzione delle settimane tra gli impianti è nel complesso equilibrata, sebbene siano presenti alcuni impianti con copertura temporale ridotta.

In *Tabella 3.17* vengono riportate le principali statistiche descrittive del dataset 24-25:

Statistica	% Tamponi positivi N_1	Prevalenza WBE
Media	6.64	6.91
Mediana	3.33	0.85
Deviazione standard	6.08	14.54
Min	0.98	0.00
Max	20.01	176.88
IQR	10.50	8.54

Tabella 3.17— principali statistiche descrittive-dataset 2024-2025

Le variabili hanno distribuzioni asimmetriche positive, coerentemente con la presenza di picchi epidemici.

3.7.2 Modelli di regressione lineare multipla per valutare la prevalenza WBE e il virus influenza

Vengono eseguiti quattro modelli con la seguente forma:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 \dots + \beta_K * X_k + \varepsilon \quad [16]$$

Modello 1-2-3:

- Variabile risposta Y: % di tamponi positivi N_1 al tempo t, % di tamponi positivi N_1 anticipati di una settimana e % di tamponi positivi N_1 anticipati di due settimane;
- Covariate: prevalenza WBE, stagioni, ed interazioni fra stagioni e percentuale di influenza, la cui percentuale è maggiore del 3%.

Modello 4

- Variabile risposta Y: % di tamponi positivi influenza;
- Covariate: prevalenza influenza A e stagioni.

In *Tabella 3.18* vengono riportati i principali parametri dei modelli di regressione lineare multipla riguardanti il gene N_1 :

.Modello	n	R ²	RMSE	Prevalenza WBE (p)	Stagioni (p)	Interazione (p)	DW	Hp violate
Nessun anticipo	807	0.668	3.53	<0.0001	<0.0001	<0.0001	0.15	Autocorr, mancata stima
Anticipo 1 settimana	792	0.707	3.30	<0.0001	<0.0001	<0.0001	0.13	Autocorr, mancata stima
Anticipo 2 settimane	777	0.794	2.76	0.056 (borderline)	<0.0001	<0.0001	0.14	Autocorr, mancata stima

Tabella 3.18— principali parametri dei modelli di regressione lineare multipla riguardanti il gene N_1 -prevalenza WBE (stagione ed interazione influenza) -anno 2024-2025.

In *Tabella 3.19* vengono riportati i principali parametri del modello di regressione lineare multipla dell'influenza.

Modello	n	R ²	RMSE	Prevalenza WBE (p)	Stagioni (p)	DW	Hp violate
Influenza virus	793	0.807	6.90	<0.0001	<0.0001	0.17	autocorrelazione

Tabella 3.19-- principali parametri del modello di regressione lineare multipla dell'influenza-prevalenza WBE (stagione)-anno 2024-2025

Dai risultati dei modelli si può affermare che per quanto riguarda N₁:

- La capacità esplicativa aumenta all'aumentare dell'anticipo temporale;
- Le stagioni sono significative in tutti i modelli e anche l'interazione con l'influenza;
- L'effetto della prevalenza nel modello a due settimane diminuisce;
- Coerentemente con la natura dei dati epidemiologici, le ipotesi di autocorrelazione positiva dei residui (DW=0.15, 0.13 e 0.14) e mancata stima sono violate, probabilmente a causa della struttura temporale non completamente catturata dal modello lineare.

Per quanto riguarda l'influenza:

- Il modello ha una forte capacità esplicativa;
- Le covariate sono significative;
- Anche in questo caso l'ipotesi di autocorrelazione è violata (DW=0.17).

Il modello a due settimane presenta le migliori performance globali, quindi, viene approfondito di seguito, insieme al modello relativo all'influenza.

Gli altri modelli sono riportati in *Appendice*.

Modello di regressione lineare multipla: % tamponi positivi N_1 - anticipo 2 settimane

In *Tabella 3.20* viene riportato il riepilogo della stima:

R-quadro	0.793884
R-quadro corretto	0.791465
Scarto quadratico medio (RMSE)	2.758105
Osservazioni	777

Tabella 3.20-- riepilogo della stima-modello di regressione lineare multipla-%tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane-anno 2024-2025

In *Tabella 3.21* viene mostrata la stima dei parametri del modello:

Termine	Stima	Errore std	Statistica T	p-value
Intercetta	6.1432	0.1161	52.93	<0.0001
Prevalenza	0.1346	0.0704	1.91	0.0565
Stagioni[inverno]	-4.3410	0.2210	-19.64	<0.0001
Stagioni[primavera]	-4.8169	0.2321	-20.75	<0.0001
Stagioni[estate]	-3.0032	0.3196	-9.40	<0.0001
Stagioni[estate/autunno]	10.4937	0.2921	35.93	<0.0001
Stagioni[inverno]*influenza >3%	-0.0617	0.4660	-0.13	0.8947
Stagioni[primavera]*influenza >3%	-0.4020	0.9179	-0.44	0.6615
Stagioni[estate] *influenza >3%	3.9135	2.3385	1.67	0.0946
Stagioni[estate/autunno] *influenza >3%	1.2228	1.9666	0.62	0.5343

Tabella 3.21-- tabella stima dei parametri -modello di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane-anno 2024-2025

Nessuna covariata non significativa viene eliminata poiché globalmente sono significative nel modello. L'effetto della prevalenza WBE tende ad attenuarsi nel medio termine, suggerendo che la dinamica temporale e la struttura stagionale contribuiscano in misura rilevante alla spiegazione della variabilità osservata.

Viene riportato di seguito, in *Figura 3.14*, il grafico dei valori osservati rispetto a quelli predetti dal modello.

Confronto fra valori osservati e stimati dal modello per il gene N1 (anticipo 2 settimane)

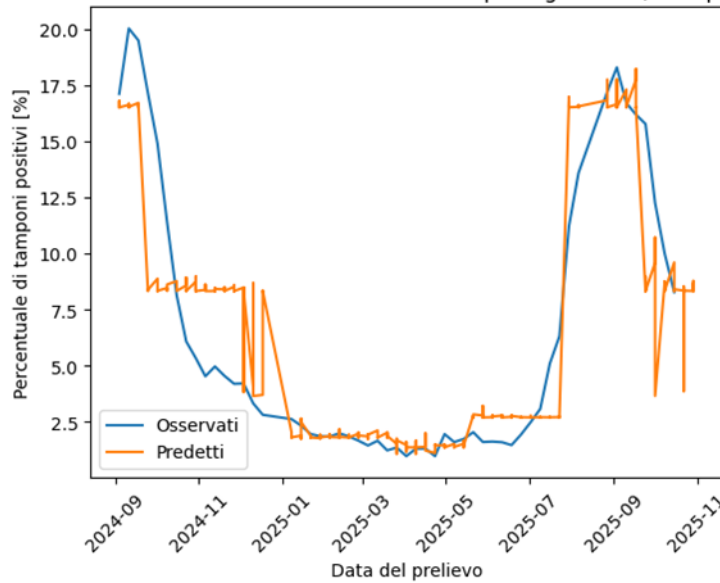


Figura 3.14 – confronto fra valori osservati e stimati dal modello di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) - anticipo 2 settimane-anno 2024-2025

Il modello riproduce bene l'andamento generale, intercettando le fasi di crescita e decrescita. In corrispondenza delle transizioni rapide ci sono maggiori discostamenti tra le due curve: questo è coerente con autocorrelazione residua, presenza di picchi e struttura temporale non completamente catturata.

In *Figura 3.15* viene riportato il diagramma dei residui rispetto ai previsti:

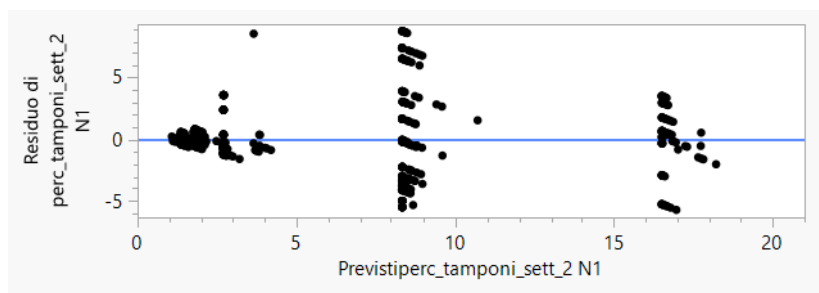


Figura 3.15 –diagramma dei residui rispetto ai previsti-modello di regressione lineare multipla -% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) - anticipo 2 settimane-anno 2024-2025

I residui non si distribuiscono come una nuvola casuale di punti intorno allo zero ma ci sono dei cluster verticali; la linearità generale è accettabile. La struttura a blocchi riflette l'effetto delle stagioni ed è presente una leggera eteroschedasticità.

In *Figura 3.16* viene mostrato il Q-Q plot:

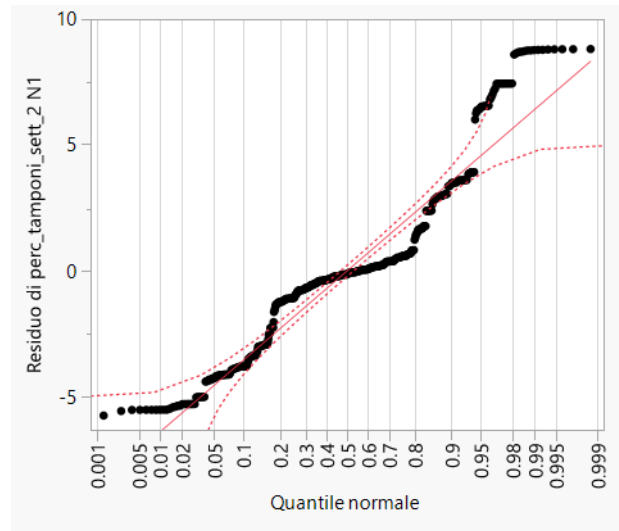


Figura 3.16 – diagramma dei quantili normali residui-modello di regressione lineare multipla -% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane-anno 2024-2025

I residui non presentano perfettamente una distribuzione normale, infatti, nella parte centrale si nota una normalità approssimata e deviazione nelle code a causa della distribuzione asimmetrica.

Anche in questo caso, la similarità fra R^2 della statistica PRESS e quello ordinario, suggerisce che il modello non è influenzato da nessun punto.

Modello di regressione lineare multipla: influenza

Nella *Tabella 3.22* viene mostrato il riepilogo della stima del modello:

R-quadro	0.806978
R-quadro corretto	0.805751
Scarto quadratico medio (RMSE)	6.902408
Osservazioni	793

Tabella 3.22-- riepilogo della stima-modello di regressione lineare multipla-influenza-prevalenza WBE (stagione)-anno 2024-2025

Nella *Tabella 3.23* viene riportata la stima dei parametri del modello:

Termine	Stima	Errore std	Statistica T	p-value
Intercetta	9.5072	0.2654	35.82	<0.0001
Stagioni[inverno]	27.5748	0.5201	53.02	<0.0001
Stagioni[primavera]	-6.5327	0.5685	-11.49	<0.0001
Stagioni[estate]	-9.2111	0.5240	-17.58	<0.0001
Stagioni[estate/autunno]	-6.6231	0.5616	-11.79	<0.0001
Prevalenza influenza	20.0589	3.3221	6.04	<0.0001

Tabella 3.23-- tabella stima dei parametri-modello di regressione lineare multipla-influenza-prevalenza WBE (stagione)-anno 2024-2025

La stagionalità nel caso dell'influenza è molto significativa, la prevalenza è anch'essa significativa ma l'effetto è meno marcato rispetto a quello delle stagioni, che sono il fattore dominante nel modello. L'influenza è infatti, un virus fortemente stagionale.

Di seguito, in *Figura 3.17*, viene riportato il grafico dei valori osservati rispetto a quelli predetti dal modello.

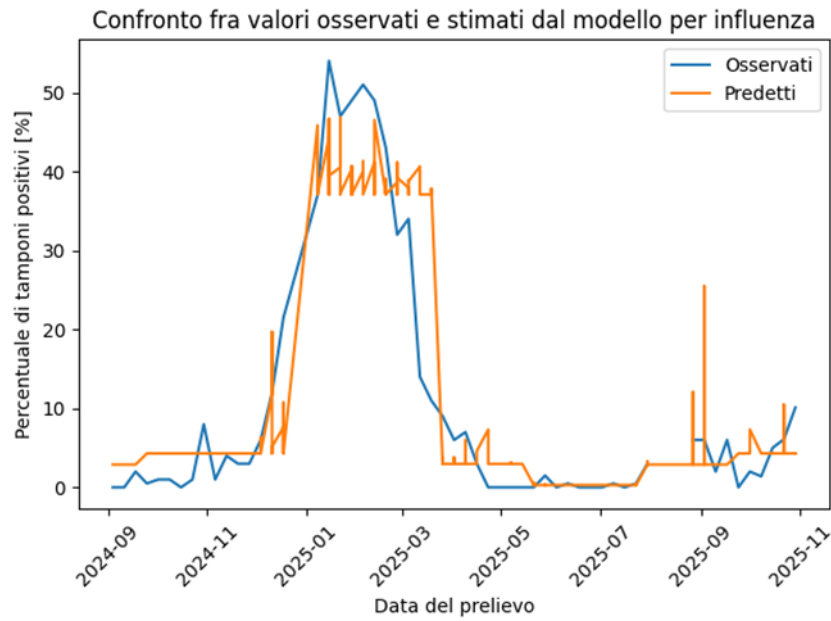


Figura 3.17 – confronto fra valori osservati e stimati dal modello di regressione lineare multipla-influenza-prevalenza WBE (stagione)-anno 2024-2025

L'andamento dell'influenza, come già accennato, è fortemente stagionale e presenta un picco epidemico invernale che viene leggermente sottostimato dal modello.

In generale il modello cattura bene le fasi di crescita e decrescita del virus, ma i valori estremi sono più complicati da catturare con un modello lineare.

In *Figura 3.18* viene riportato il diagramma dei residui rispetto ai previsti:

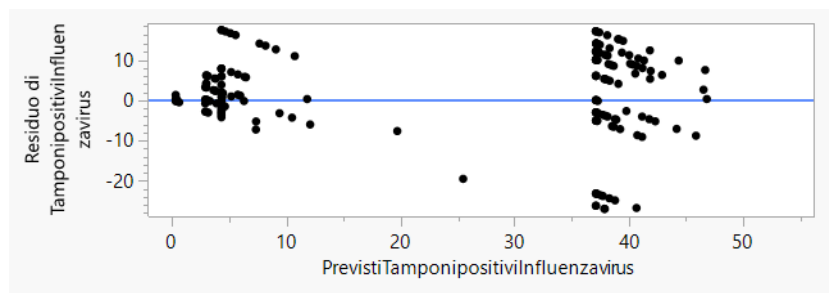


Figura 3.18 – diagramma dei residui rispetto ai previsti-modello di regressione lineare multipla-influenza-prevalenza WBE (stagione)-anno 2024-2025

Si osserva una struttura a gruppi, intorno allo zero, dovuta alla presenza di dummy stagionali; per valori bassi predetti si hanno residui più concentrati e viceversa, e questo suggerisce che sia presente il fenomeno di eteroschedasticità moderata.

Il modello presenta dei residui alti ma l' R^2 della statistica PRESS indica comunque che questi punti non influenzano il modello.

In *Figura 3.19* viene mostrato il Q-Q plot:

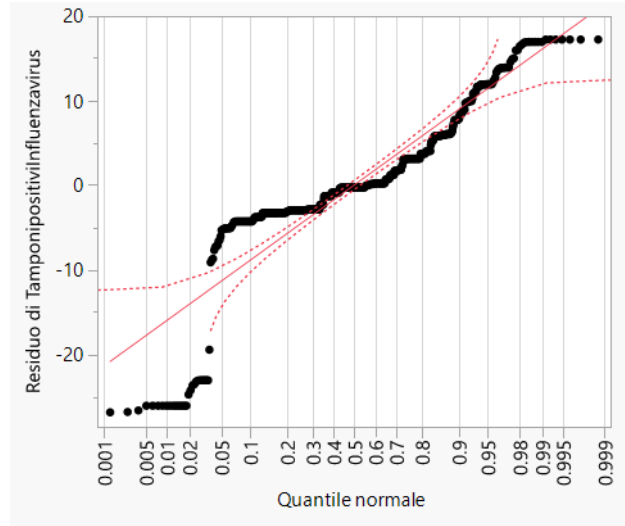


Figura 3.19 – diagramma dei quantili normali residui-modello di regressione lineare multipla-influenza-prevalenza WBE (stagione)-anno 2024-2025

Nella parte centrale i residui seguono la distribuzione normale, nella coda sinistra sono presenti residui negativi, a causa probabilmente della presenza di picchi epidemici, eteroschedasticità e forte stagionalità

4 Conclusioni

4.1.1 Scelta del gene Migliore N_1

Sin dalle prime analisi di correlazione si può affermare che i marcatori fecali ORF ed N_3 mostrino una correlazione debole o quasi assente con i dati epidemiologici, ovvero i nuovi casi clinici, mentre il gene N_1 mostra la correlazione più robusta.

Per tutti e tre i geni, la normalizzazione per l'azoto ammoniacale non determina miglioramenti rispetto ai valori quantitativi non normalizzati perché la variabilità di diluizione legata all'azoto non rappresenta il principale fattore di variabilità nei dati analizzati.

Allo stesso modo, la normalizzazione per gli abitanti equivalenti AE determina una riduzione della correlazione per tutti e tre i marcatori fecali.

Questo risultato suggerisce che la normalizzazione tramite gli AE, frutto di un'interpolazione media mobile delle concentrazioni di azoto totale, non catturi efficacemente la componente epidemica associata alla carica virale.

Quindi, si può affermare che nel periodo e nei siti analizzati, il segnale virale grezzo risulta già rappresentativo dell'andamento epidemico, mentre le normalizzazioni introducono una variabilità aggiuntiva non direttamente correlata ai nuovi casi e limitano il dataset alle sole osservazioni corredate dalle misurazioni dei parametri operativi.

Anche i modelli di regressione lineare semplice (Y = nuovi casi clinici, X = carica virale nel refluo) confermano quanto affermato in fase di correlazione: la relazione fra N_1 e i dati epidemiologici è la più robusta e il modello ha un R^2 pari a 0.3048, ed entrambe le normalizzazioni non migliorano il modello.

Per migliorare la capacità di spiegare la variabilità osservata nei casi clinici, sono stati applicati modelli di regressione lineare multipla includendo, oltre alla carica virale nel refluo, due predittori categoriali rappresentativi del contesto epidemiologico e territoriale: il sito di prelievo e la variante dominante.

Il modello che utilizza il gene N_1 quantitativo in relazione ai nuovi casi clinici risulta il più robusto, spiegando circa il 78% della variabilità osservata ($R^2 = 0.78$). Il gene N_1 mantiene un effetto altamente significativo anche in presenza delle altre covariate, confermandone il ruolo di indicatore affidabile dell'andamento epidemico. La significatività dei predittori categoriali (sito e variante) indica che le differenze tra i bacini serviti dagli impianti e le diverse fasi epidemiche, caratterizzate da specifiche

varianti, contribuiscono in modo rilevante alla spiegazione dei casi clinici, migliorando la capacità esplicativa complessiva del modello.

4.2.1 Stima della prevalenza

Per confrontare la percentuale di tamponi positivi clinici con la percentuale di positivi derivanti dal refluo, si trasformano il numero di infetti giornalieri per impianto, tramite shedding rate, in termini di prevalenza, ovvero in percentuale di popolazione che risulta infetta in quel dato giorno.

I fattori significativi nei modelli di regressione lineare multipla impiegati sono la prevalenza stimata tramite WBE, la stagione di campionamento e l'interazione fra la stagione e la circolazione influenzale, maggiore del 3%.

La prevalenza WBE del gene N_1 ha una relazione positiva con la percentuale di tamponi positivi clinici. Tuttavia, nel modello con anticipo di due settimane l'effetto della prevalenza WBE rimane positivo e significativo, ma la quota di varianza spiegata dalla componente stagionale risulta predominante.

In particolare, a parità di prevalenza WBE, i valori sono più elevati nelle stagioni fredde e meno nei periodi estivi. Questo è coerente con la natura dei virus respiratori, che si trasmettono maggiormente nei mesi invernali.

Ci sono, tuttavia, per tutti i modelli, delle ipotesi della regressione lineare che vengono violate: moderata autocorrelazione positiva dei residui, moderata eteroschedasticità nei picchi epidemici, deviazione dalla normalità nelle code della distribuzione dei residui.

I risultati evidenziano che la prevalenza stimata tramite WBE rappresenta un indicatore significativamente associato all'andamento clinico, pur in presenza di fattori stagionali predominanti e di violazioni delle ipotesi del modello lineare.

4.3.1 Dataset 24-25 ed estensione al virus influenza

L'analisi dei dati relativi agli anni 2024-2025 conferma la robustezza della sorveglianza ambientale anche su un periodo più recente e sul virus influenza.

Nei modelli di regressione lineare multipla stimati, la stagionalità ha un ruolo predominante nella spiegazione della variabilità osservata.

Nel modello che anticipa di due settimane, la prevalenza WBE ha un'associazione positiva con la variabile risposta, e la sua significatività è borderline. La stagionalità è

la covariata che spiega di più e l'interazione con influenza indica che l'effetto virale vari in funzione del periodo stagionale.

Per quanto riguarda l'influenza A, la variabilità della percentuale di tamponi è dominata fortemente dalla stagionalità, coerentemente con l'andamento dei virus respiratori. Anche la prevalenza WBE rimane significativa ma con un contributo inferiore.

Questo indica che per l'influenza, la sorveglianza ambientale sia efficace nel monitoraggio in tempo reale della circolazione del virus, mentre l'effetto di anticipo è meno marcato rispetto a quello osservato per il SARS-CoV-2.

Il ruolo predominante della stagionalità è coerente con diversi fattori, quali la maggiore permanenza in ambienti chiusi nei mesi freddi, minore ventilazione, condizione più favorevole di permanenza del virus ecc.

In conclusione, la sorveglianza ambientale è uno strumento statisticamente associato all'andamento clinico delle infezioni respiratorie, ed è un supporto complementare alla sorveglianza tradizionale, in considerazione di fattori stagionali e ipotesi violate dei modelli lineari.

Per nuovi patogeni con shedding rate ignoti, il confronto con la sorveglianza tradizionale rimane essenziale per interpretare il segnale WBE. Per virus come SARS-CoV-2 ed influenza, la sorveglianza ambientale può assumere un ruolo sempre più centrale nel monitoraggio della circolazione virale e nell'ottimizzazione delle risorse di sanità pubblica.

Bibliografia

- [1] M. & Eddy, *Wastewater Engineering: Treatment and Resource Recovery*.
- [2] ISO 5667-10:2020.
- [3] ISO 15216-1:2017.
- [4] R. P. A. T. F. M. Arianna Azzellino, «PROTOCOLLO DI INTESA STIPULATO IN DATA 25 NOVEMBRE 2022 TRA REGIONE LOMBARDIA –DIREZIONE GENERALE WELFARE E IL POLITECNICO DI MILANO – DIPARTIMENTO DI INGEGNERIA CIVILE E AMBIENTALE PER LA PARTECIPAZIONE AL PROTOCOLLO NAZIONALE SARI PER LA RICERCA DEL SAR,» 2023.
- [5] Steen Ethelberg et al., «GLOWACON Consolidated Technical Guidance».
- [6] Ministero dell' Ambiente e della Tutela del Territorio e del Mare. (2012). *Guida tecnica per la progettazione e gestione dei sistemi di fitodepurazione per il trattamento delle acque reflue urbane*.
- [7] «Evaluating Population Normalization Methods Using Chemical,» Marco Verani 1,† , Ileana Federigi 1,* ,† , Alessandra Angori 1, Alessandra Pagani 1 , Francesca Marvulli 2,, 4 Maggio 2025.
- [8] P. I. G. B. (. scientifico), «Contratto di ricerca con il Politecnico di Milano - Dipartimento di Ingegneria Civile e Ambientale,» 2024.
- [9] RashaMaal-Bared et al., «DoesnormalizationofSARS-CoV-2 concentrationsbyPepperMildMottle,» 2022.
- [10] Sarah M. Prasek et al., « Population level SARS-CoV-2 fecal shedding rates determined via wastewater-based epidemiology,» 2022.
- [11] Wan Yang¹ et al., «The use of wastewater surveillance to estimate SARS-CoV-2 fecal viral shedding pattern and identify time periods with intensified transmission,» 2025.

- [12] Bradley W. Schmitz et al., «Enumerating asymptomatic COVID-19 cases and estimating SARS-CoV-2».
- [13] P. E. Bottarelli, Quaderno di epidemiologia veterinaria.
- [14] Manon Chassaing et al., «Influenza RNA fluxes monitoring in wastewater as a complementary epidemiological surveillance indicator: A four-year nationwide study in Luxembourg».

A Appendice

Nuovi casi giornalieri vs predetti dal modello-Modelli di regressione lineare multipla-Modello N_1 -nuovi casi (sito e variante)

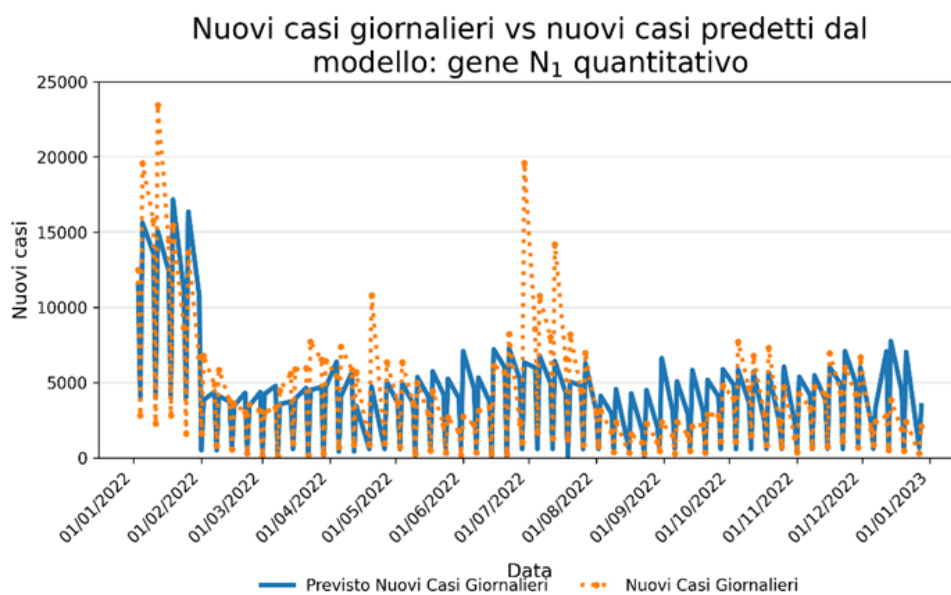


Figura A.1– nuovi casi giornalieri vs predetti dal modello- Modello di regressione lineare multipla- Modello N_1 quantitativo -nuovi casi (sito e variante) -2022

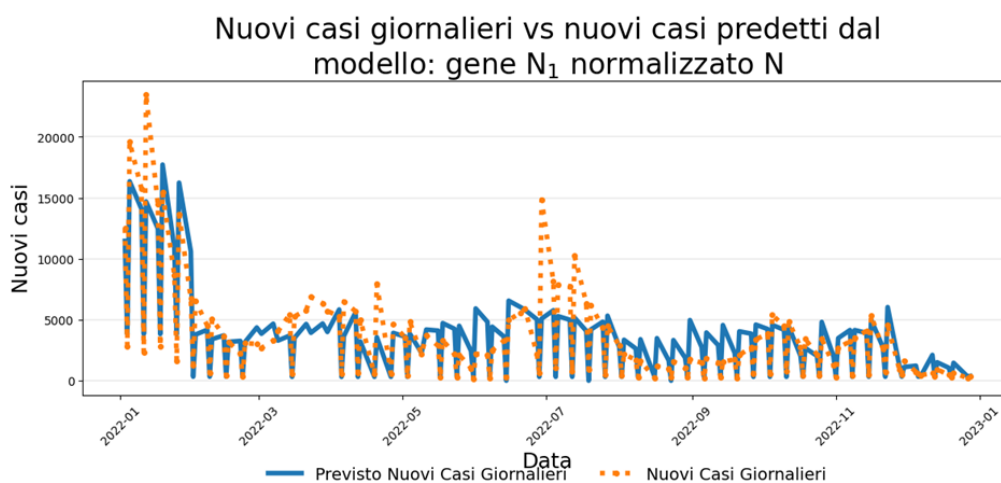


Figura A.2– nuovi casi giornalieri vs predetti dal modello - Modello di regressione lineare multipla- Modello N_1 normalizzato N -nuovi casi (sito e variante) -2022

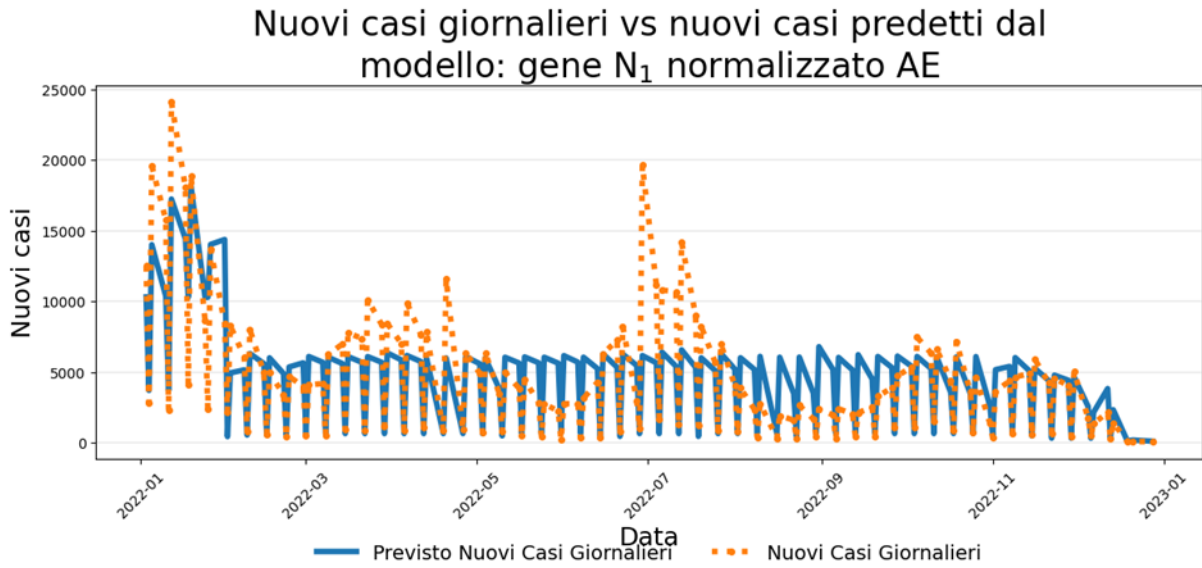


Figura A.3– nuovi casi giornalieri vs predetti dal modello- Modello di regressione lineare multipla- Modello N_1 normalizzato AE -nuovi casi (sito e variante) -2022

Grafici di confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici-2022

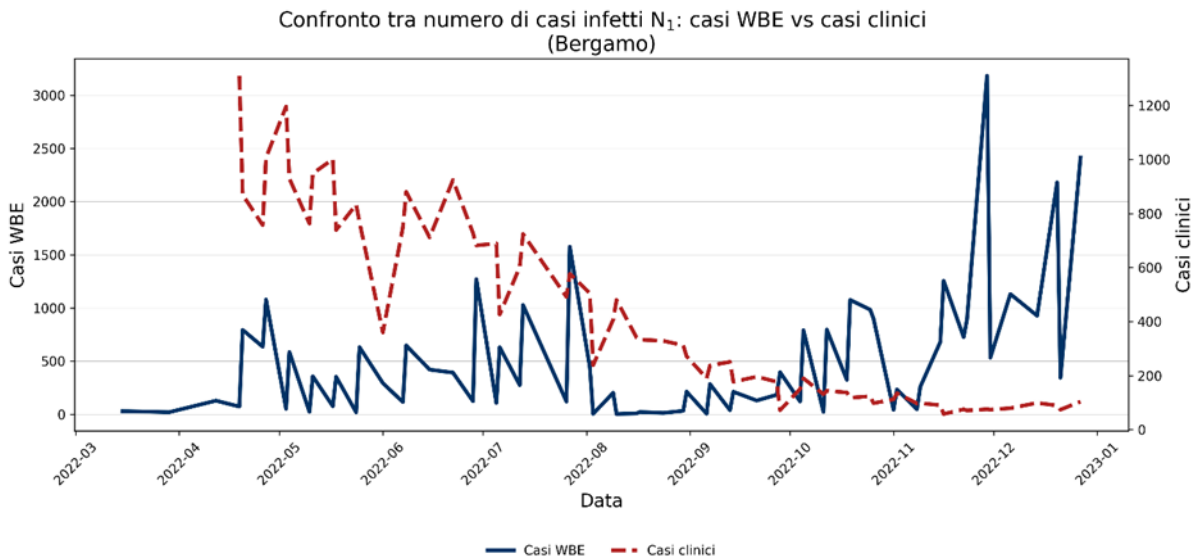


Figura A.4– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Bergamo-2022

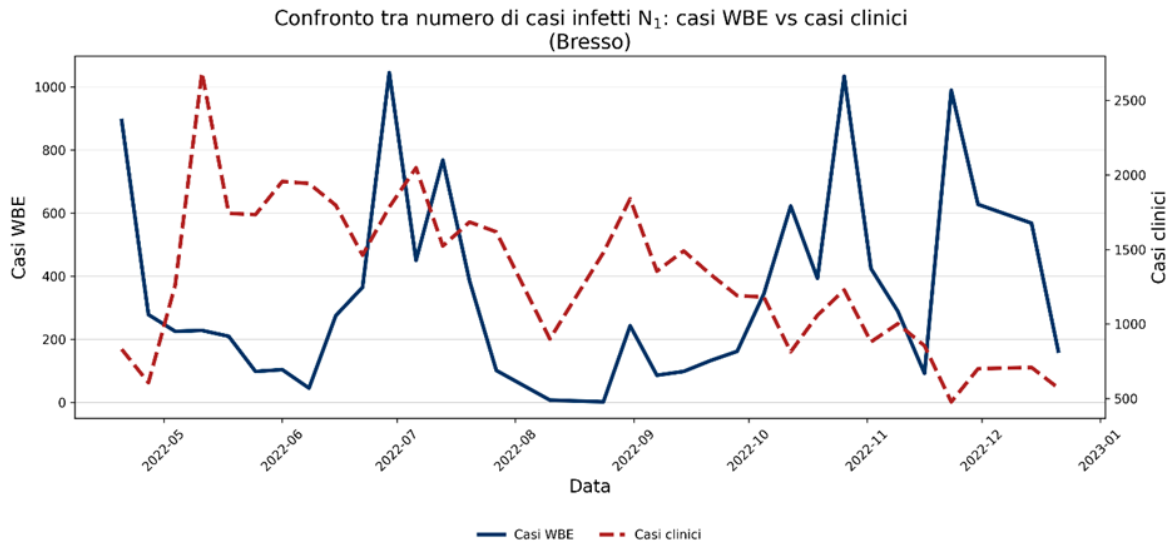


Figura A.5– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Bresso-2022

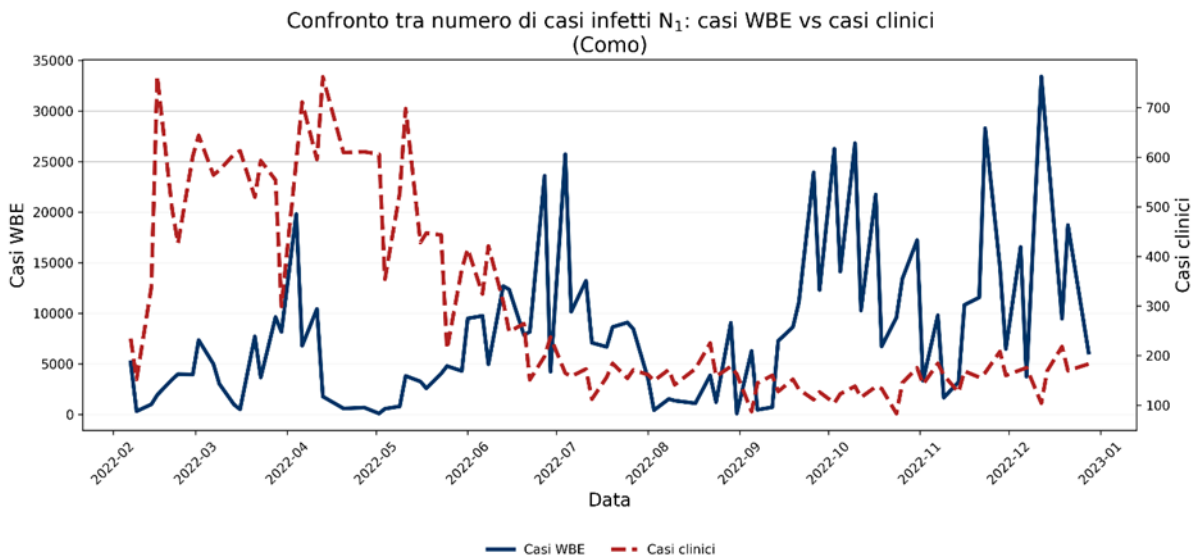


Figura A.6– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Como-2022

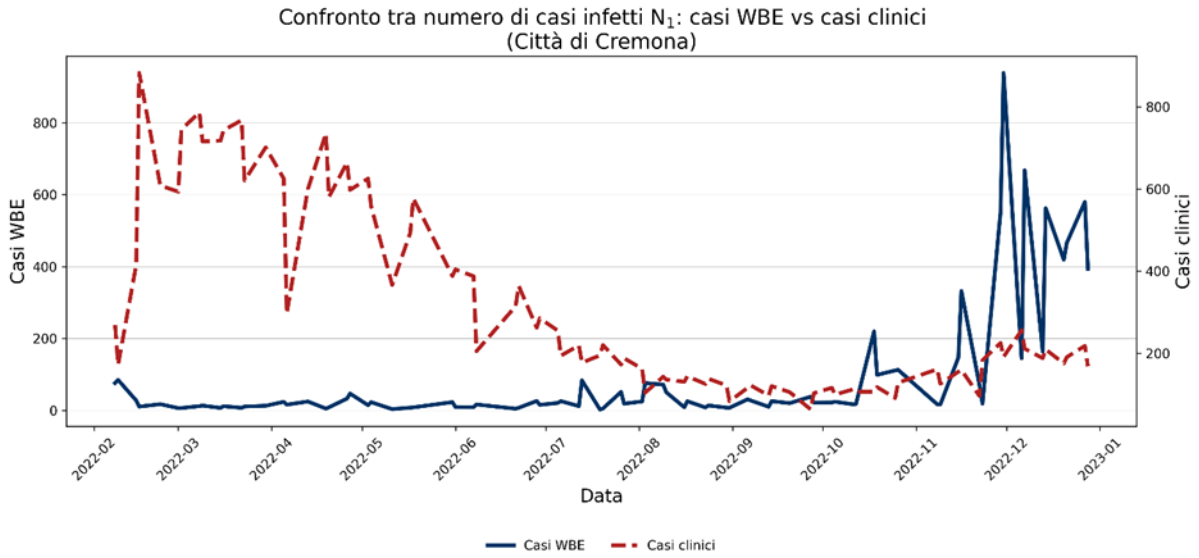


Figura A.7– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Città di Cremona-2022

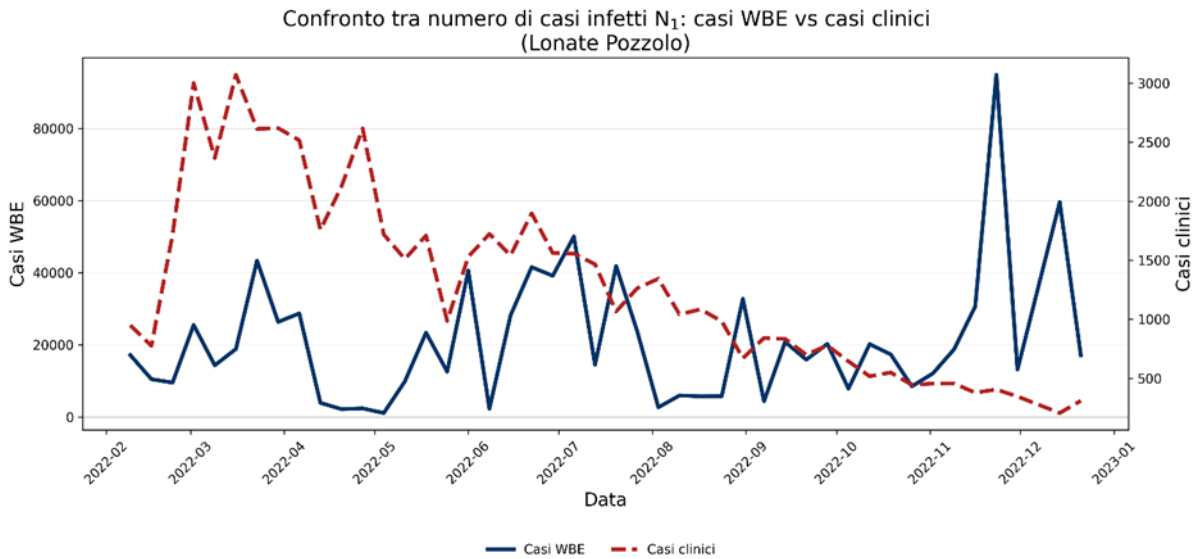


Figura A.8– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Lonate Pozzolo-2022

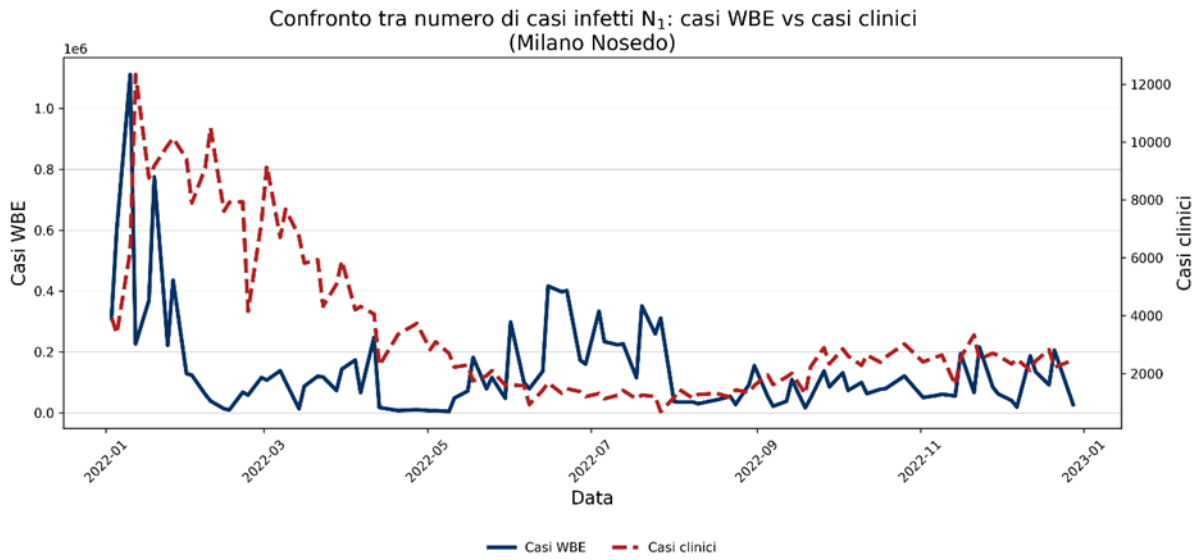


Figura A.9– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Milano Nosedo-2022

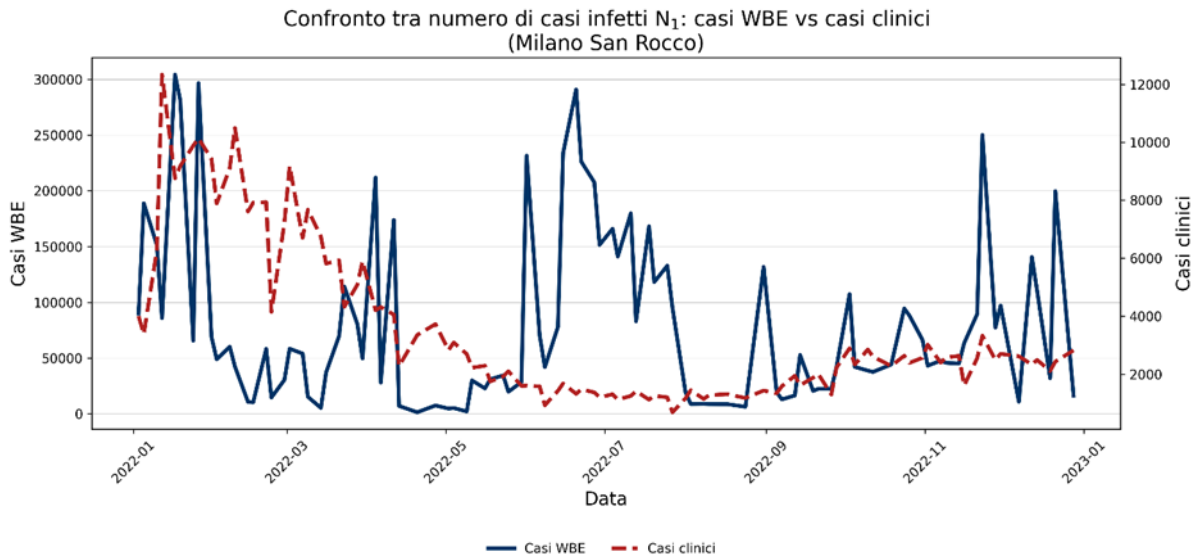


Figura A.10– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Milano San Rocco-2022

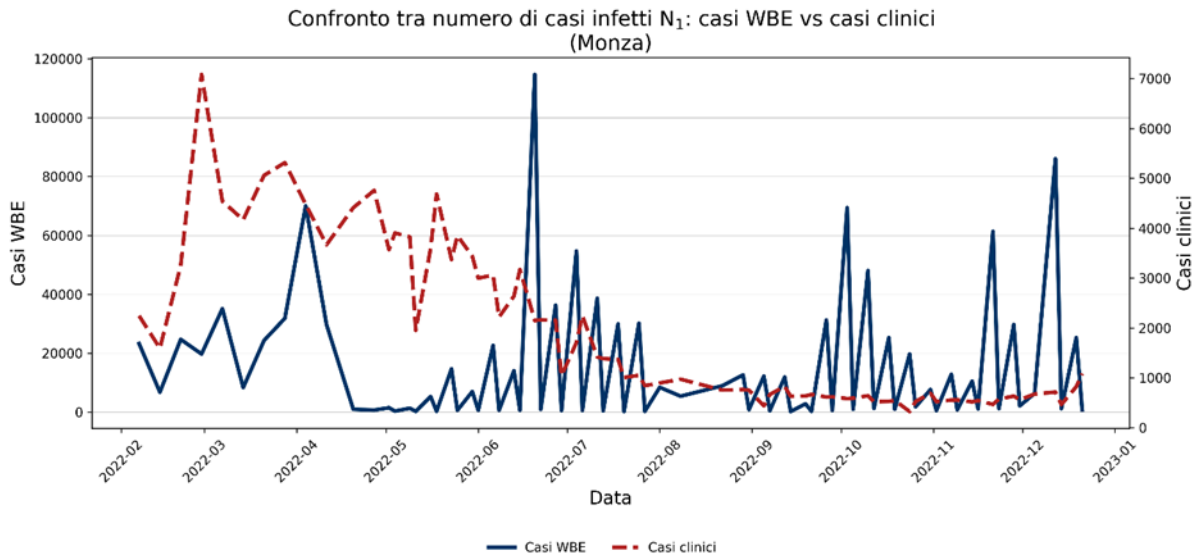


Figura A.11– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Monza-2022

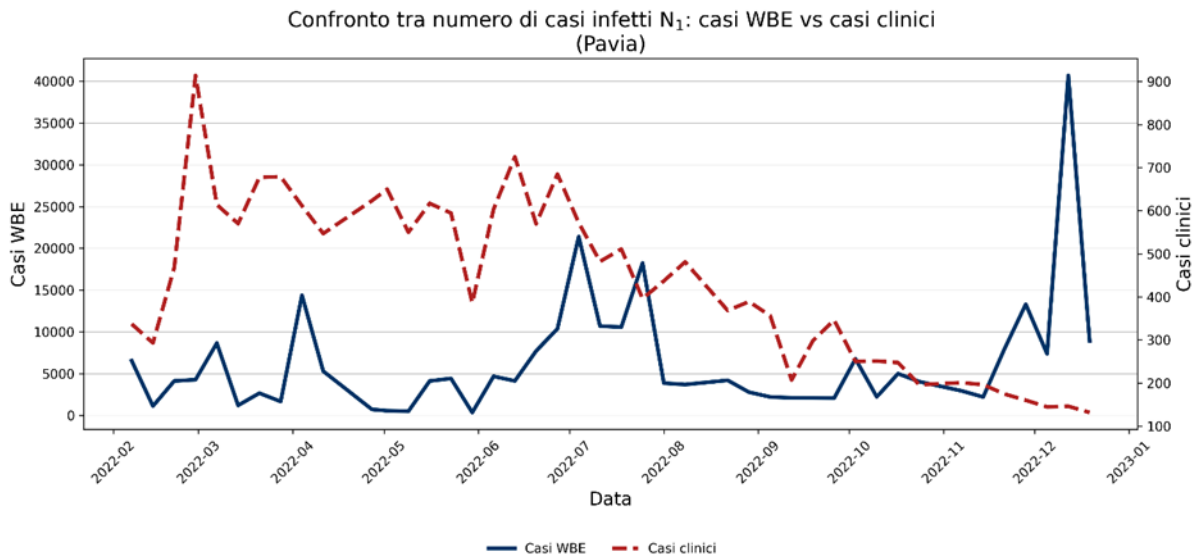


Figura A.12– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Pavia-2022

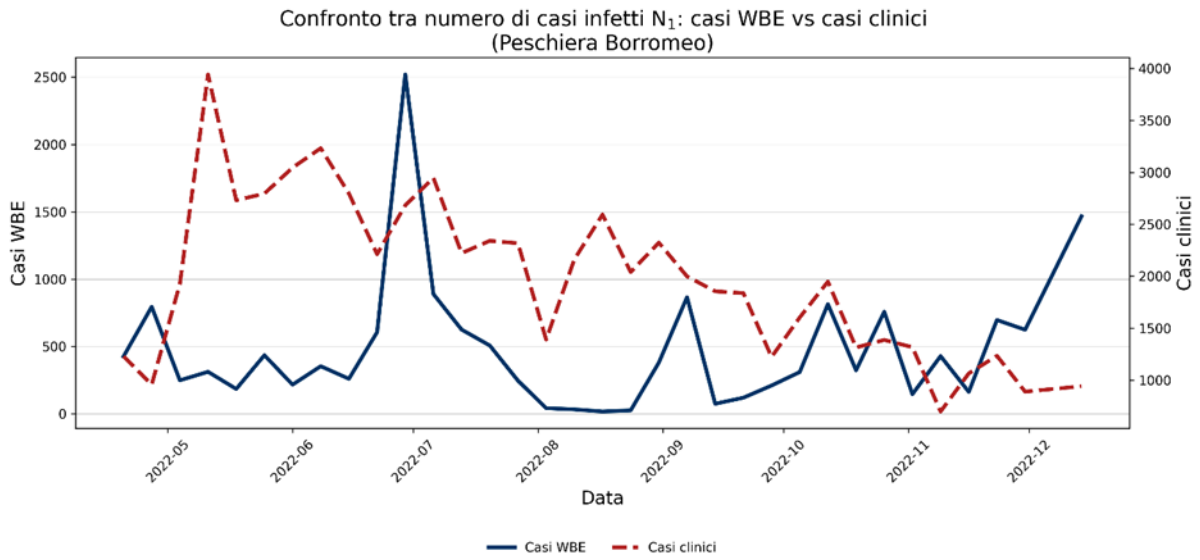


Figura A.13– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Peschiera Borromeo-2022

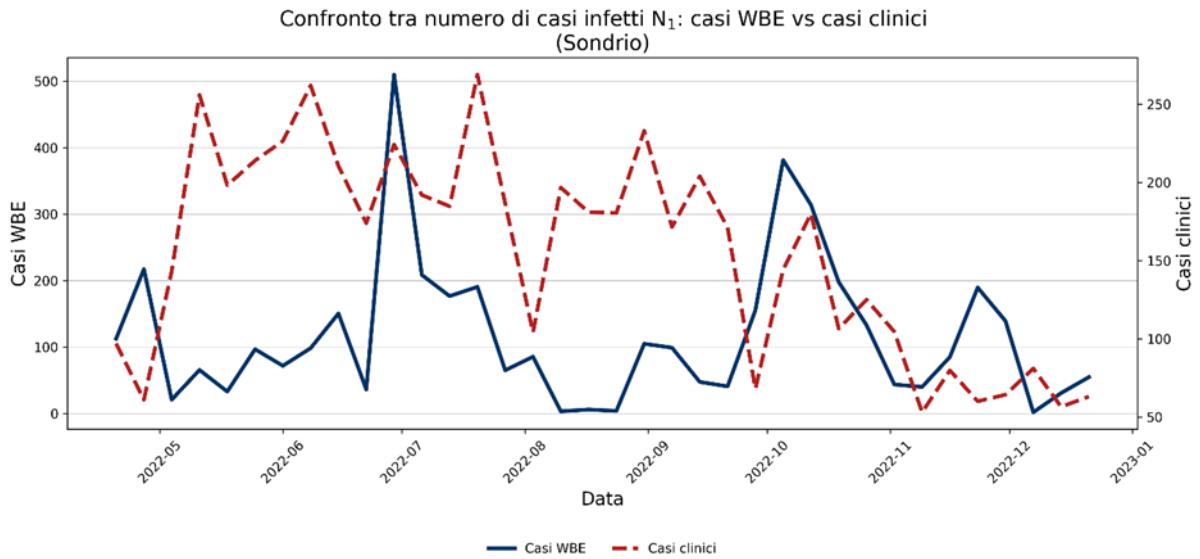


Figura A.14– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Sondrio-2022

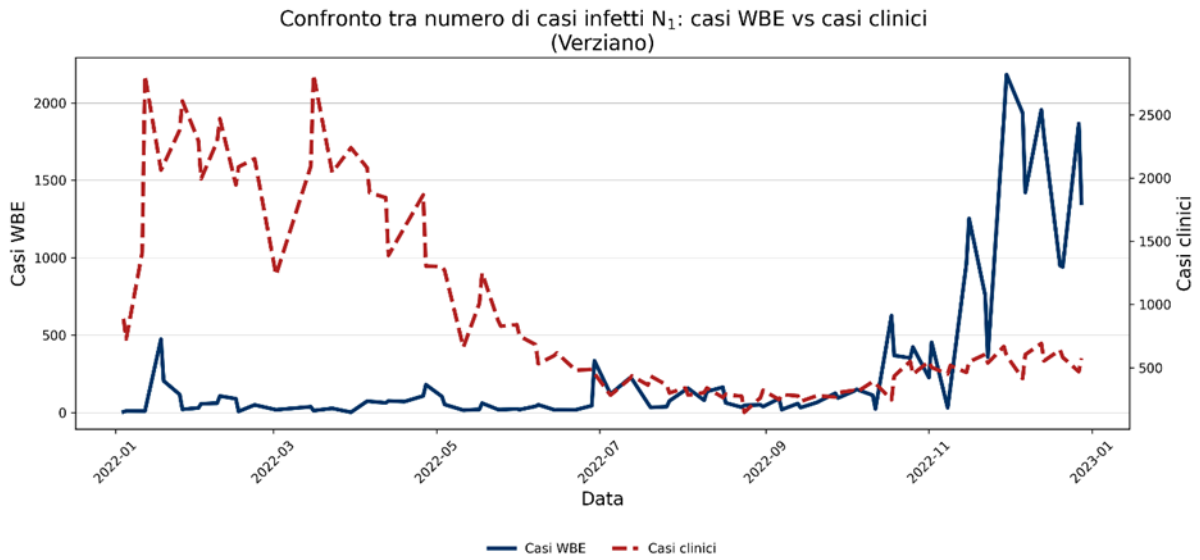


Figura A.15– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Verziano-2022

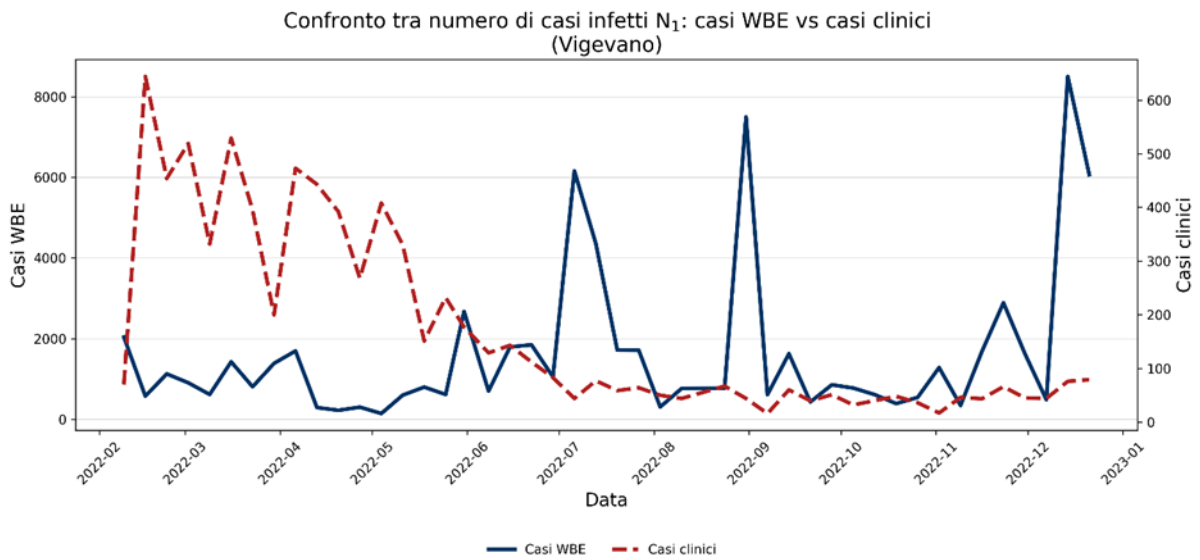


Figura A.16– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Vigevano-2022

*Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE
(stagione ed interazione influenza) -2022*

Risposta % tamponi positivi N₁

Grafico delle risposte osservate rispetto alle risposte previste

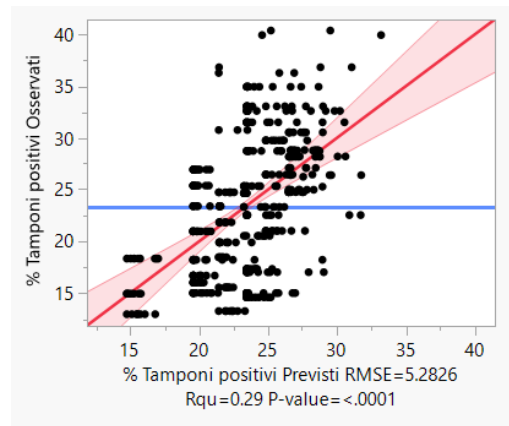


Figura A.17– grafico delle risposte rispetto alle risposte previste- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) - 2022

Riepilogo della stima

R-quadro	0.289991
R-quadro corretto	0.278303
Scarto quadratico medio	5.282561
Media della risposta	23.29984
Osservazioni (o somme pesate)	495

Tabella A.1–riepilogo della stima- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -2022

Analisi della varianza ANOVA

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Modello	8	5539.181	692.398	24.8123
Errore	486	13562.047	27.905	Prob > F
C. totale	494	19101.228		<.0001*

Tabella A.2–analisi della varianza ANOVA- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -2022

Stime dei parametri

Termine		Stima	Errore std	Rapporto T	Prob> t
Intercetta		22.140182	0.356235	62.15	<.0001*
Stagione [Autunno inverno]		3.1836008	0.558895	5.70	<.0001*
Stagione [Estate]	Distorto	2.2726329	0.602262	3.77	0.0002*
Stagione [Fine estate]	Distorto	-2.574369	0.531051	-4.85	<.0001*
Stagione [Inverno]		-2.740306	0.570458	-4.80	<.0001*
Prevalenza WBE		0.2027337	0.032175	6.30	<.0001*
Stagione [Autunno inverno] *(Influenza >3%-0.30303)		1.6737643	1.10448	1.52	0.1303
Stagione [Estate]*(Influenza > 3%-0.30303)	Distorto	3.2101514	2.024454	1.59	0.1135
Stagione [Fine estate] *(Influenza >3%-0.30303)	Azzerato	0	0	.	.
Stagione [Inverno]*(Influenza >3%-0.30303)		-6.668659	1.278129	-5.22	<.0001*

Tabella A.3–stime dei parametri- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -2022

Alcuni termini di interazione risultano “distorti” o “azzerati”. Tale condizione è riconducibile alla distribuzione della variabile “Influenza >3%”, che, come evidenziato nel grafico riportato sotto, si concentra in periodi temporali limitati ed è completamente assente in altre stagioni. Questa forte concentrazione determina una sovrapposizione tra effetto stagionale ed effetto influenzale, introducendo collinearità tra le covariate e rendendo non stimabili o instabili alcuni coefficienti di interazione. La significatività globale del termine indica comunque che l’inclusione dell’interazione migliora la capacità esplicativa del modello, pur in presenza di tali limitazioni strutturali.

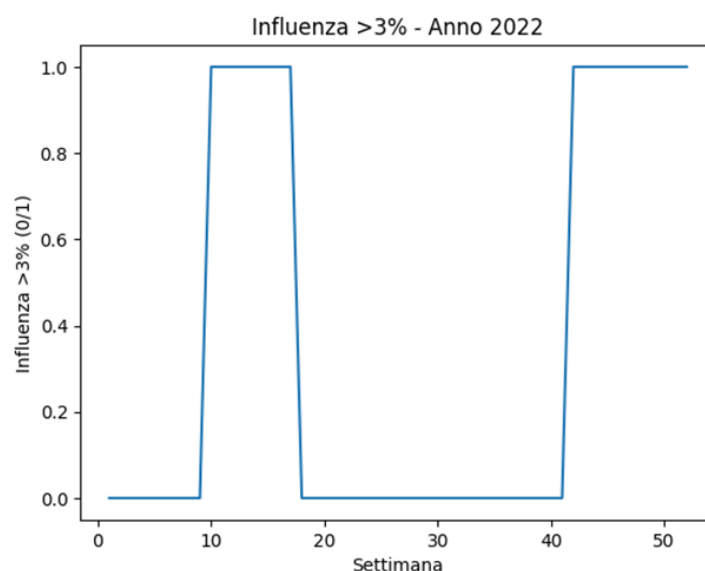


Figura A.18– influenza>3% per stagione --2022

Test degli effetti

Origine	Npar	DF	Somma dei quadrati	Rapporto F	Prob > F
Stagione	4	3	1222.4276	14.6020	<.0001*
Prevalenza WBE	1	1	1107.8867	39.7015	<.0001*
Stagione*Influenza>3%	4	3	890.4627	10.6367	<.0001*

Tabella A.4–test degli effetti- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -2022

Diagramma dei residui rispetto ai previsti

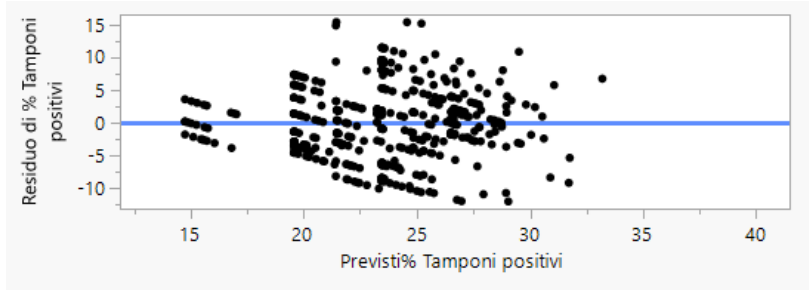


Figura A.19– diagramma dei residui rispetto ai previsti, - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022

Diagramma dei valori residui per riga

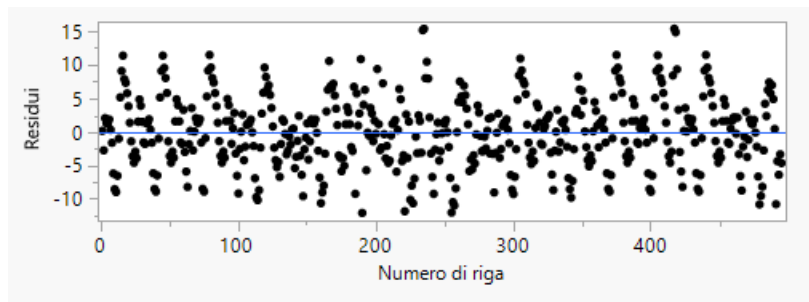


Figura A.20– diagramma dei valori residui per riga - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022

Residui studentizzati esternamente con limiti simultanei al 95% (Bonferroni) in rosso, limiti individuali in verde

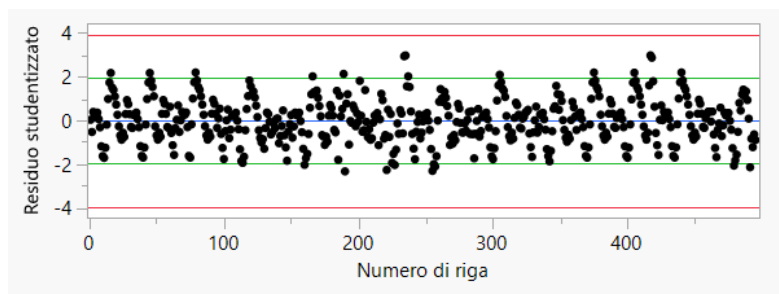


Figura A.21– residui studentizzati - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022

Press

Residui	SSE	RMSE	R-quadro
Press	14016.6766 32	5.32132676	0.2662
Ordinario	13562.0467 51	5.28256055	0.2900

Tabella A.5–statistica PRESS- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022

Durbin-Watson

Durbin-Watson	Numero di osservazioni	Autocorrelazione	Prob<D W
0.5932446	495	0.7026	<.0001*

Tabella A.6–test di Durbin-Watson- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022

Diagramma dei quantili normali residui

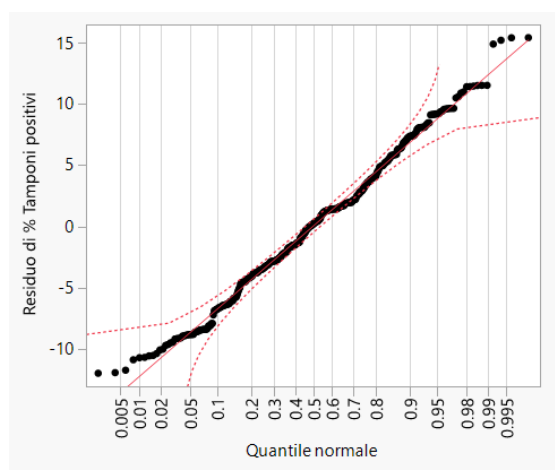


Figura A.22– diagramma dei quantili normali residui - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022

Valori osservati vs previsti

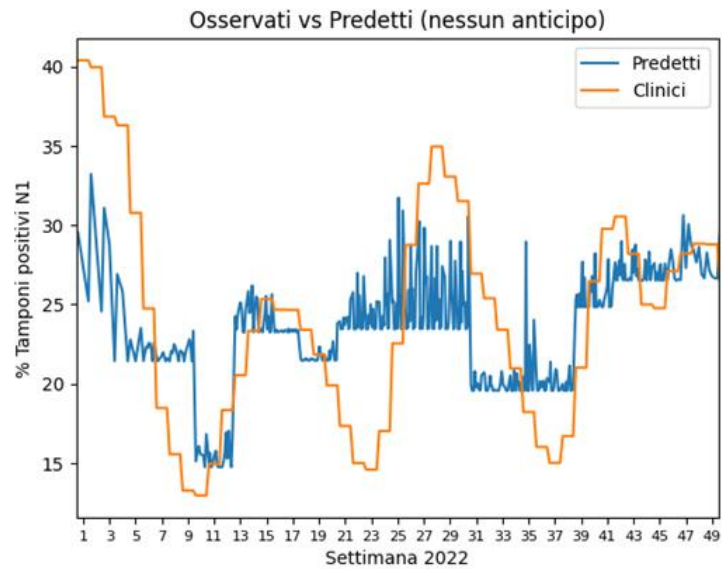


Figura A.23– valori osservati vs previsti - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022

Valori osservati vs previsti-media settimanale

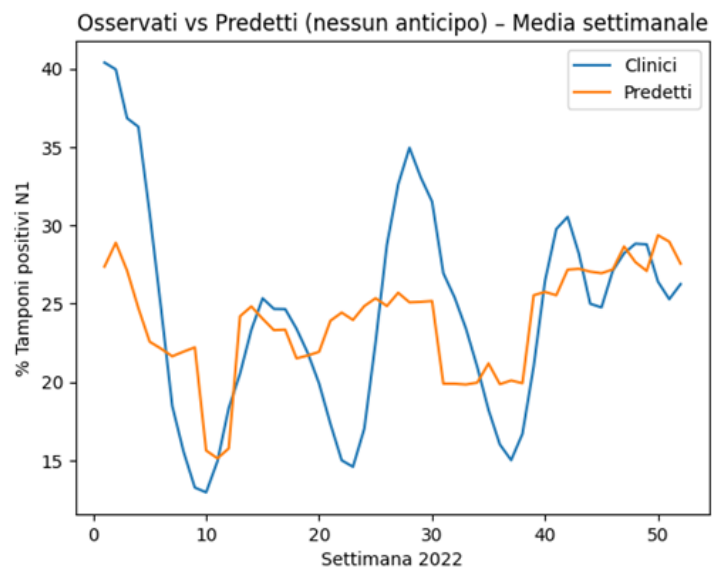


Figura A.24– valori osservati vs previsti-media settimanale - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022

Risposta % tamponi positivi N_1 -anticipo 1 settimana

Grafico delle risposte osservate rispetto alle risposte osservate previste

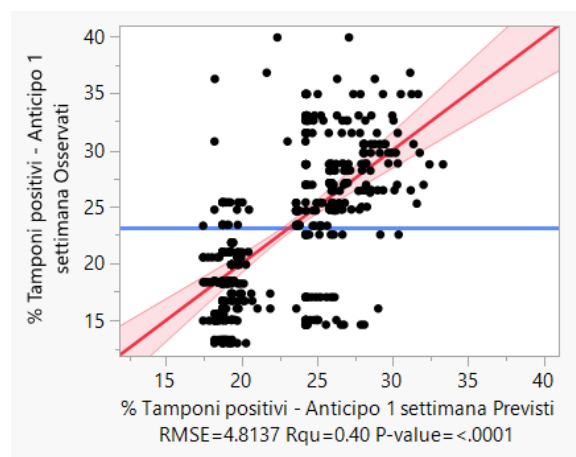


Figura A.25– grafico delle risposte osservate rispetto alle risposte osservate previste- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Riepilogo della stima

R-quadro	0.397497
R-quadro corretto	0.387285
Scarto quadratico medio	4.813696
Media della risposta	23.17218
Osservazioni (o somme pesate)	481

Tabella A.7– riepilogo della stima- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Analisi della varianza ANOVA

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Modello	8	7215.629	901.954	38.9248

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Errore	472	10937.027	23.172	Prob > F
C. totale	480	18152.655		<.0001*

Tabella A.8– analisi della varianza ANOVA- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Stime dei parametri

Termine		Stima	Errore std	Rapporto T	Prob> t
Intercetta		21.692282	0.326602	66.42	<.0001*
Stagione [Autunno inverno]		5.705436	0.526285	10.84	<.0001*
Stagione [Estate]	Distorto	2.1515956	0.537708	4.00	<.0001*
Stagione [Fine estate]	Distorto	-2.98231	0.486454	-6.13	<.0001*
Stagione [Inverno]		-3.721664	0.520702	-7.15	<.0001*
Prevalenza WBE		0.2233832	0.02951	7.57	<.0001*
Stagione [Autunno inverno] *(Influenza >3%-0.28898)		-2.253058	1.046623	-2.15	0.0318*
Stagione [Estate] *(Influenza >3%-0.28898)	Distorto	-1.295987	1.866902	-0.69	0.4879
Stagione [Fine estate] *(Influenza >3%-0.28898)	Azzerato	0	0	.	.
Stagione [Inverno] *(Influenza >3%-0.28898)		-0.748871	1.164773	-0.64	0.5206

Tabella A.9–stime dei parametri- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Test degli effetti

Origine	Npar	DF	Somma dei quadrati	Rapporto F	Prob > F
Stagione	4	3	3121.1706	44.8992	<.0001*
Prevalenza WBE	1	1	1327.7472	57.3005	<.0001*
Settimana (Raccolto)* Influenza > 3%	4	3	524.6059	7.5467	<.0001*

Tabella A.10–test degli effetti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Diagramma dei residui rispetto ai previsti

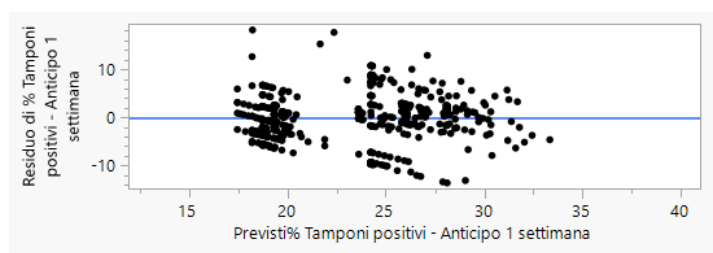


Figura A.26– diagramma dei residui rispetto ai previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Diagramma dei valori residui per riga

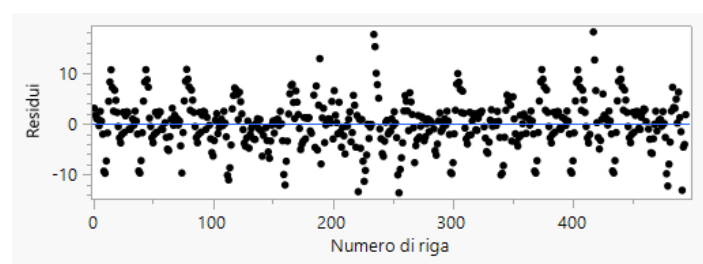


Figura A.27– diagramma dei valori residui per riga- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Residui studentizzati esternamente con limiti simultanei al 95% (Bonferroni) in rosso, limiti individuali in verde

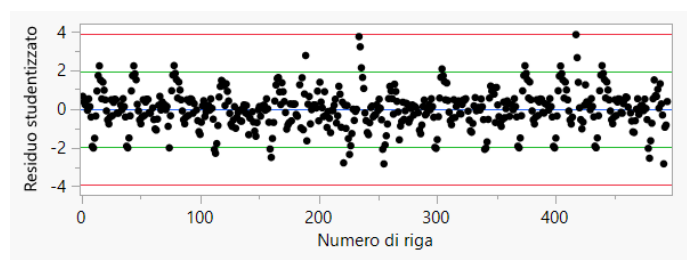


Figura A.28– residui studentizzati- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Press

Residui	SSE	RMSE	R-quadro
Press	11318.24858	4.8508414	0.3765
Ordinario	10937.026535	4.81369571	0.3975

Tabella A.11–statistica PRESS- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Durbin-Watson

Durbin-Watson	Numero di osservazioni	Autocorrelazione	Prob<D W
0.7111563	481	0.6438	<.0001*

Tabella A.12–test di Durbin-Watson- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Diagramma dei quantili normali residui

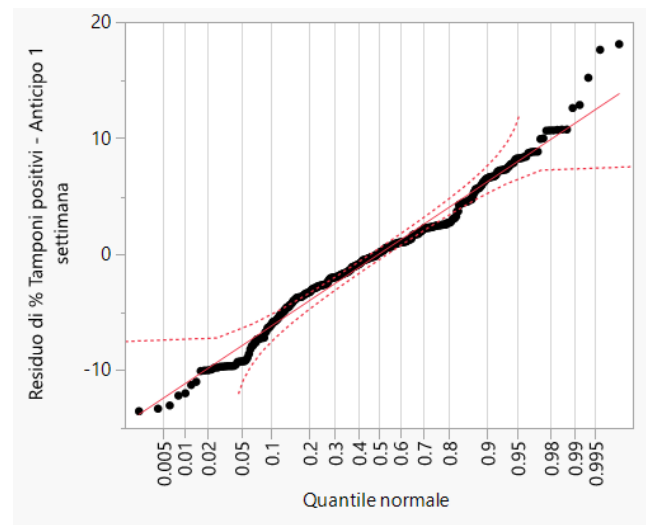


Figura A.29– diagramma dei quantili normali residui- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Valori osservati vs previsti

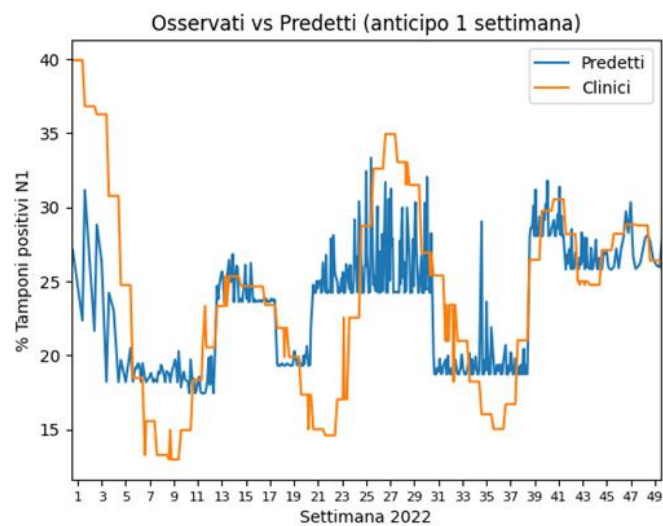


Figura A.30– valori osservati vs previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Valori osservati vs previsti -media settimanale

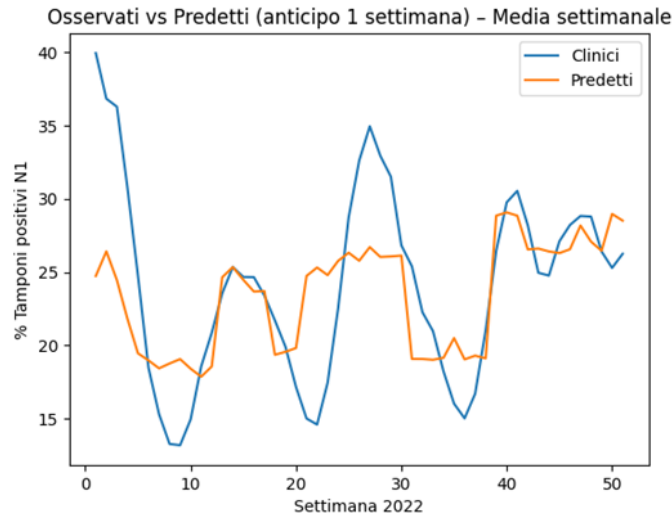


Figura A.31– valori osservati vs previsti-media settimanale- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022

Risposta % tamponi positivi N_1 -anticipo 2 settimane

Grafico delle risposte osservate rispetto alle risposte previste

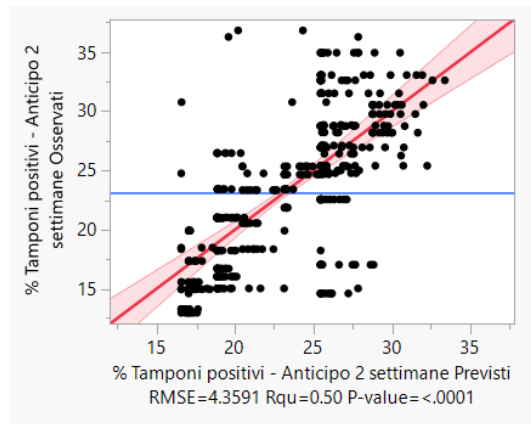


Figura A.32– grafico delle risposte osservate rispetto alle risposte previste- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Riepilogo della stima

R-quadro	0.495019
----------	----------

R-quadro corretto	0.486217
Scarto quadratico medio	4.359087
Media della risposta	23.12894
Osservazioni (o somme pesate)	468

Tabella A.13–riepilogo della stima- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Analisi della varianza ANOVA

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Modello	8	8549.679	1068.71	56.2430
Errore	459	8721.752	19.00	Prob > F
C. totale	467	17271.432		<.0001*

Tabella A.14–analisi della varianza ANOVA- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Stime dei parametri

Termine		Stima	Errore std	Rapporto T	Prob> t
Intercetta		21.324637	0.296299	71.97	<.0001*
Stagione [Autunno inverno]		6.5351744	0.489081	13.36	<.0001*
Stagione [Estate]	Distorto	2.2521322	0.475792	4.73	<.0001*
Stagione [Fine estate]	Distorto	-2.471387	0.443504	-5.57	<.0001*
Stagione [Inverno]		-3.693996	0.47217	-7.82	<.0001*
Prevalenza WBE		0.1940716	0.026822	7.24	<.0001*
Stagione [Autunno inverno] *(Influenza > 3%-0.2735)		-3.213996	0.986324	-3.26	0.0012*
Stagione [Estate]*(Influenza > 3%-0.2735)	Distorto	-6.785842	1.711979	-3.96	<.0001*

Termine		Stima	Errore std	Rapporto T	Prob> t
Stagione [Fine estate] *(Influenza > 3%-0.2735)	Azzerato	0	0	.	.
Stagione [Inverno]*(Influenza>3%-0.2735)		3.9017016	1.054816	3.70	0.0002*

Tabella A.15–stime dei parametri- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Test degli effetti

Origine	Npar	DF	Somma dei quadrati	Rapporto F	Prob > F
Stagione	4	3	3851.8386	67.5703	<.0001*
Prevalenza WBE	1	1	994.7726	52.3519	<.0001*
Stagione*Influenza >3%	4	3	1287.5192	22.5861	<.0001*

Tabella A.16–test degli effetti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Diagramma dei residui rispetto ai previsti

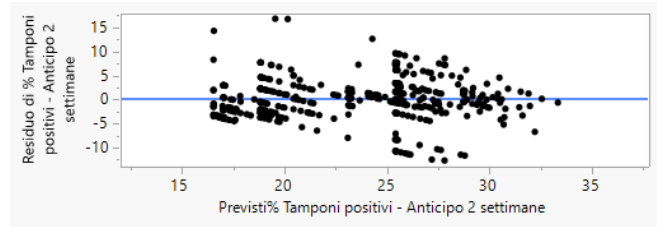


Figura A.33– diagramma dei residui rispetto ai previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Diagramma dei valori residui per riga

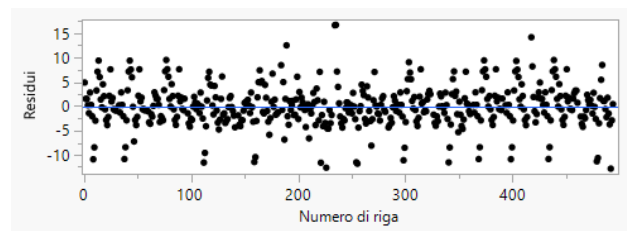


Figura A.34– diagramma dei valori residui per riga- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Residui studentizzati esternamente con limiti simultanei al 95% (Bonferroni) in rosso, limiti individuali in verde

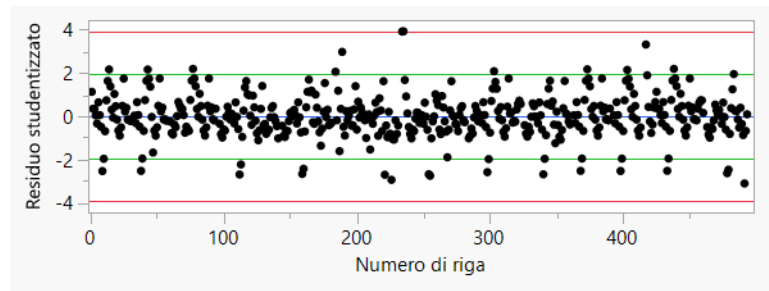


Figura A.35– residui studentizzati- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Press

Residui	SSE	RMSE	R-quadro
Press	9031.5543569	4.39297087	0.4771
Ordinario	8721.7522402	4.35908693	0.4950

Tabella A.17–statistica PRESS- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Durbin-Watson

Durbin-Watson	Numero di osservazioni	Autocorrelazione	Prob< DW
0.8812366	468	0.5580	<.0001 *

Tabella A.18–test di Durbin-Watson- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Diagramma dei quantili normali residui

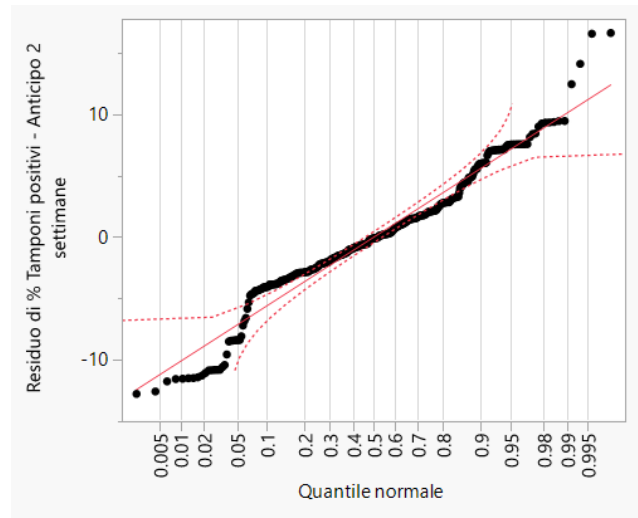


Figura A.36– diagramma dei quantili normali residui- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Valori osservati vs previsti

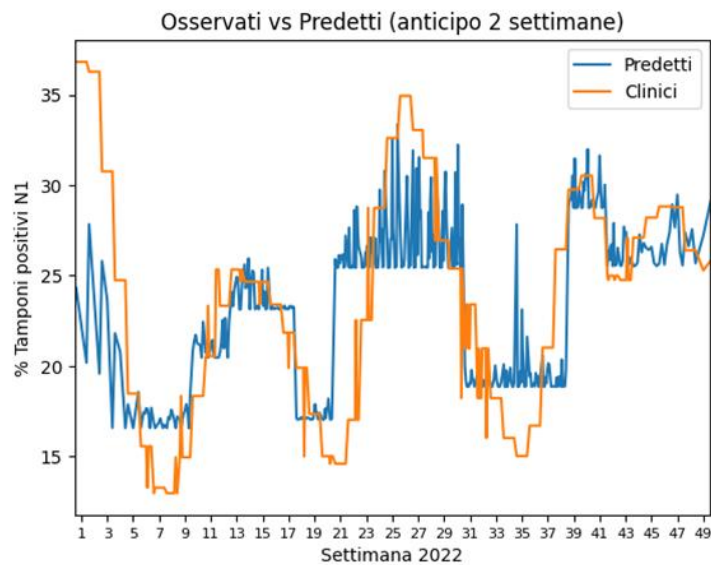


Figura A.37– valori osservati vs previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Valori osservati vs previsti-media settimanale

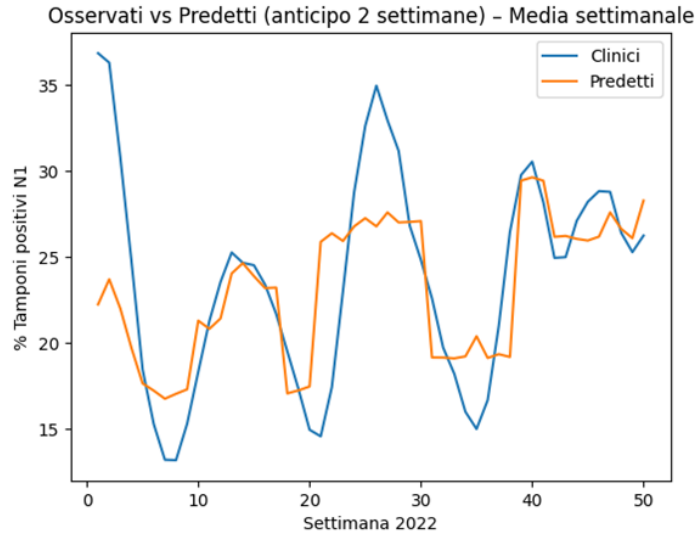


Figura A.38– valori osservati vs previsti-media settimanale- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022

Modelli di regressione lineare multipla-% Tamponi positivi N_1 -prevalenza WBE (stagione ed interazione con influenza) -anno 2024-2025

Modello % tamponi positivi N_1 -prevalenza WBE-nessun anticipo: risposta % tamponi N_1

Grafico delle risposte osservate rispetto alle risposte previste

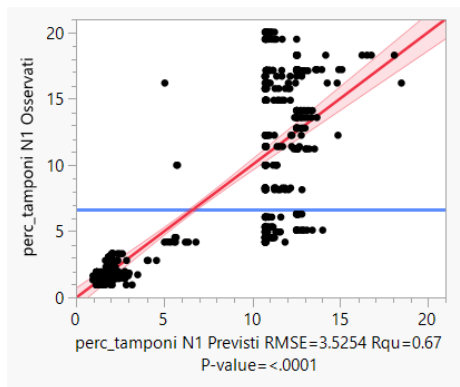


Figura A.39– grafico delle risposte osservate rispetto alle risposte previste- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2024-2025

Riepilogo della stima

R-quadro	0.667964
R-quadro corretto	0.664214
Scarto quadratico medio	3.525412
Media della risposta	6.640956
Osservazioni (o somme pesate)	807

Tabella A.19– riepilogo della stima- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -2024-2025

Analisi della varianza ANOVA

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Modello	9	19927.148	2214.13	178.1488
Errore	797	9905.540	12.43	Prob > F
C. totale	806	29832.688		<.0001*

Tabella A.20– analisi della varianza ANOVA- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -2024-2025

Mancata stima

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Mancata stima	671	8949.1481	13.3370	1.7571
Errore puro	126	956.3920	7.5904	Prob > F
Errore totale	797	9905.5401		<.0001*
				Max R-quadro

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
				0.9679

Tabella A.21– mancata stima- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -2024-2025

Stime dei parametri

Termine	Stima	Errore std	Rapporto T	Prob> t
Intercetta	5.6665065	0.147236	38.49	<.0001*
Stagioni[inverno]	-3.65922	0.282461	-12.95	<.0001*
Stagioni[primavera]	-4.564375	0.296004	-15.42	<.0001*
Stagioni[estate]	-3.068884	0.403277	-7.61	<.0001*
Stagioni[estate/autunno]	6.841746	0.370318	18.48	<.0001*
Stagioni[inverno]*(influenza>3%-0.10905)	-0.410389	0.5954	-0.69	0.4909
Stagioni[primavera]*(influenza>3%-0.10905)	-0.582222	1.173185	-0.50	0.6198
Stagioni[estate]*(influenza>3%-0.10905)	6.8657955	2.972003	2.31	0.0211*
Stagioni[estate/autunno] * (influenza>3%-0.10905)	-0.153914	2.513642	-0.06	0.9512
Prevalenza_WBE	0.4378664	0.089291	4.90	<.0001*

Tabella A.22– stime dei parametri- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -2024-2025

Nel modello 2024–2025 non si osservano problemi di stimabilità dei parametri, nonostante la soglia di influenza >3% rimanga prevalentemente concentrata nei mesi invernali. La maggiore variabilità temporale della circolazione influenzale rispetto al 2022 consente al modello di stimare separatamente gli effetti principali e l'interazione senza fenomeni di singolarità.

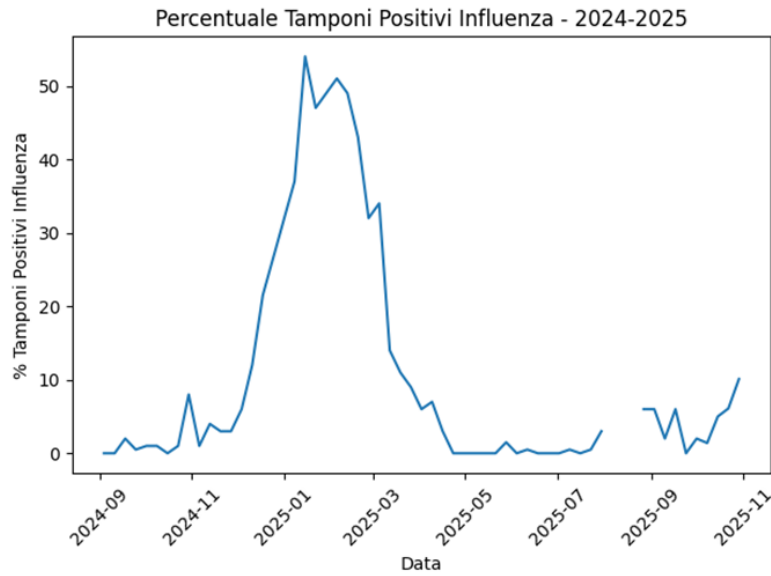


Figura A.40– % tamponi positivi influenza -2024-2025

Test degli effetti

Origine	Npar	DF	Somma dei quadrati	Rapporto F	Prob > F
Stagioni	4	4	11674.092	234.8244	<.0001*
Stagioni*influenza a>3%-	4	4	527.638	10.6134	<.0001*
Prevalenza_WBE	1	1	298.872	24.0473	<.0001*

Tabella A.23– test degli effetti- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -2024-2025

Diagramma dei residui rispetto ai previsti

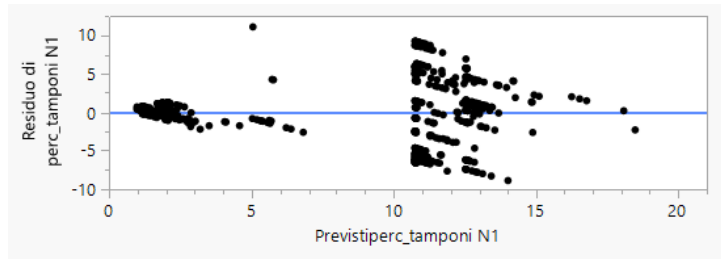


Figura A.41– diagramma dei residui rispetto ai previsti - Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) 2024-2025

Diagramma dei valori residui per riga

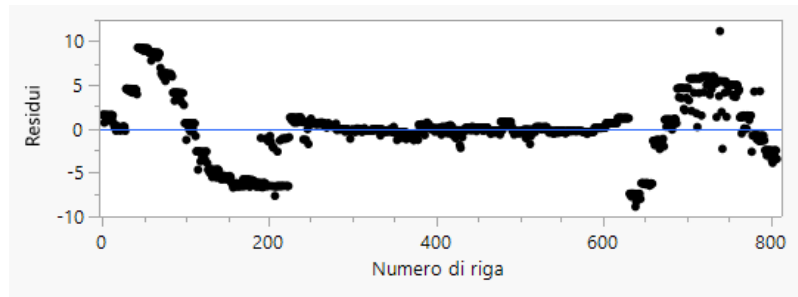


Figura A.42– diagramma dei valori residui per riga - Modelli di regressione lineare multipla-
% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) 2024-2025

Residui studentizzati esternamente con limiti simultanei al 95% (Bonferroni) in rosso,
limiti individuali in verde

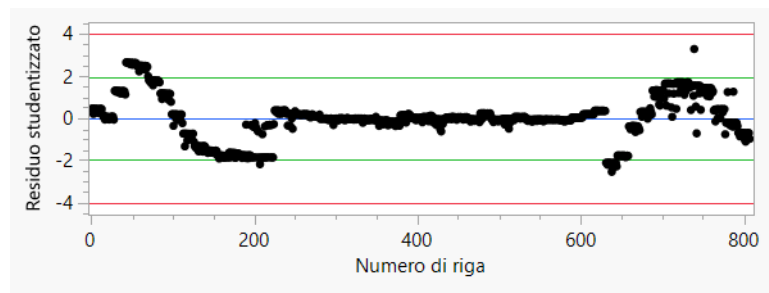


Figura A.43– residui studentizzati - Modelli di regressione lineare multipla-% tamponi
positivi N_1 -prevalenza WBE (stagione ed interazione influenza) 2024-2025

Press

Residui	SSE	RMSE	R-quadro
Press	10052.611249	3.52941459	0.6630
Ordinario	9905.5400529	3.52541233	0.6680

Tabella A.24– statistica PRESS- Modelli di regressione lineare multipla-% tamponi positivi
 N_1 -prevalenza WBE (stagione ed interazione influenza) -2024-2025

Durbin-Watson

Durbin-Watson	Numero di osservazioni	Autocorrelazione	Prob<DW
0.150998	807	0.9238	<.0001*

Tabella A.25– test di Durbin-Watson- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -2024-2025

Diagramma dei quantili normali residui

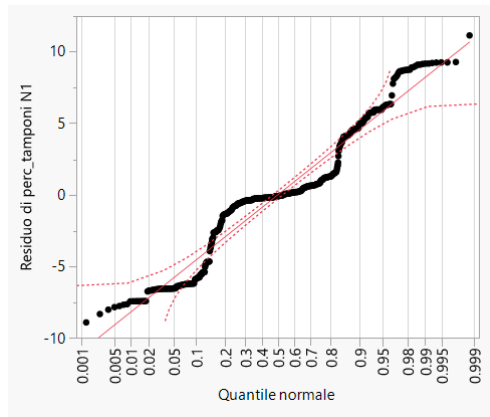


Figura A.44– diagramma dei quantili normali residui- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) 2024-2025

Valori osservati vs previsti

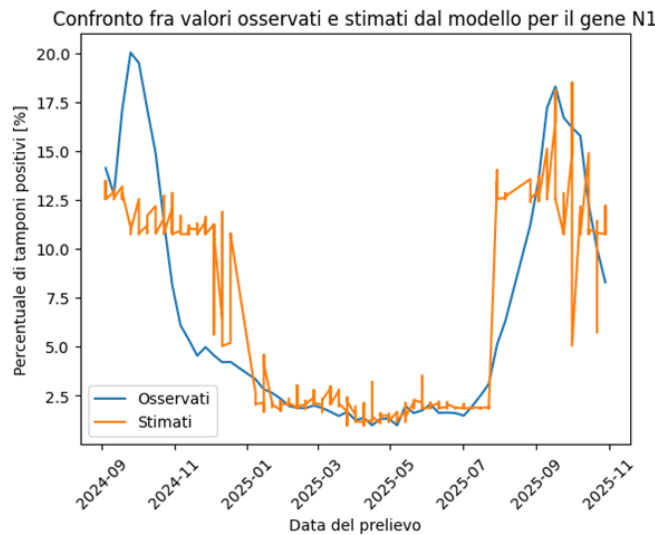


Figura A.45– valori osservati vs previsti- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) 2024-2025

Modello % tamponi positivi N_1 -prevalenza WBE-anticipo 1 settimana: risposta %
tamponi N_1 -anticipo 1 settimana

Grafico delle risposte osservate rispetto alle risposte previste

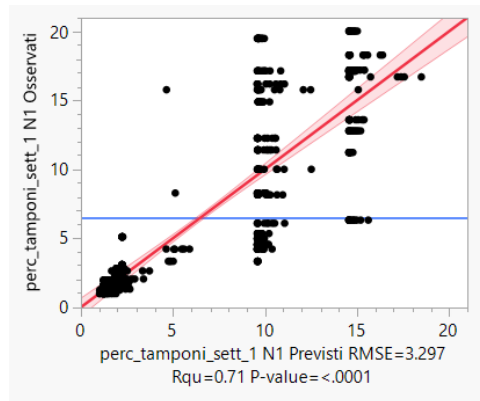


Figura A.46– grafico delle risposte osservate rispetto alle risposte previste- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025

Riepilogo della stima

R-quadro	0.706744
R-quadro corretto	0.703369
Scarto quadratico medio	3.297029
Media della risposta	6.500111
Osservazioni (o somme pesate)	792

Tabella A.26– riepilogo della stima- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025

Analisi della varianza ANOVA

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Modello	9	20486.517	2276.28	209.4017
Errore	782	8500.653	10.87	Prob > F

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
C. totale	791	28987.170		<.0001*

Tabella A.27– analisi della varianza ANOVA- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025

Mancata stima

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Mancata stima	659	7602.6071	11.5366	1.5801
Errore puro	123	898.0455	7.3012	Prob > F
Errore totale	782	8500.6527		0.0010*
				Max R-quadro
				0.9690

Tabella A.28– mancata stima- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025

Stime dei parametri

Termine	Stima	Errore std	Rapporto T	Prob> t
Intercetta	5.9008404	0.138196	42.70	<.0001*
Prevalenza_WBE	0.3101785	0.083992	3.69	0.0002*
Stagioni [inverno]	-3.997979	0.263906	-15.15	<.0001*
Stagioni [primavera]	-4.73554	0.277367	-17.07	<.0001*
Stagioni [estate]	-3.1339	0.381499	-8.21	<.0001*
Stagioni [estate/autunno]	8.7421601	0.349674	25.00	<.0001*

Termine	Stima	Errore std	Rapporto T	Prob> t
Stagioni [inverno]*(influenza>3%- 0.11111)	-0.2981	0.556948	-0.54	0.5926
Stagioni [primavera]*(influenza>3- 0.11111)	-0.483533	1.097238	-0.44	0.6596
Stagioni [estate]*(influenza>3- 0.11111)	4.9212062	2.779981	1.77	0.0771
Stagioni [estate/autunno] *(influenza>3%-0.11111)	0.8124304	2.350817	0.35	0.7297

Tabella A.29– stime dei parametri- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025

Test degli effetti

Origine	Npar	DF	Somma dei quadrati	Rapporto F	Prob > F
Prevalenza_WBE	1	1	148.251	13.6381	0.0002*
Stagioni	4	4	12331.69 3	283.6072	<.0001*
Stagioni*influenza>3%	4	4	393.874	9.0584	<.0001*

Tabella A.30– test degli effetti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025

Diagramma dei residui rispetto ai previsti

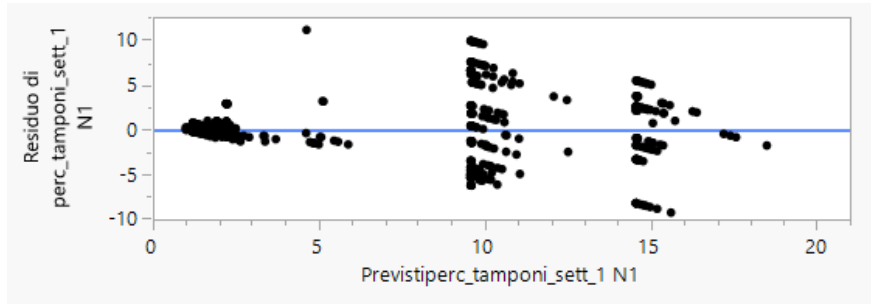


Figura A.47– diagramma dei residui rispetto ai previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) - anticipo 1 settimana- 2024-2025

Diagramma dei valori residui per riga

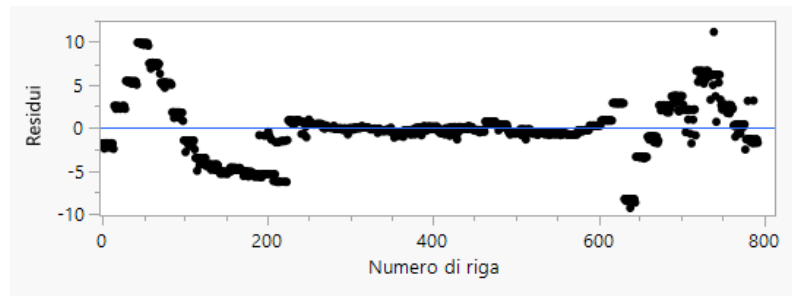


Figura A.48– diagramma dei valori residui per riga- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025

Residui studentizzati esternamente con limiti simultanei al 95% (Bonferroni) in rosso, limiti individuali in verde

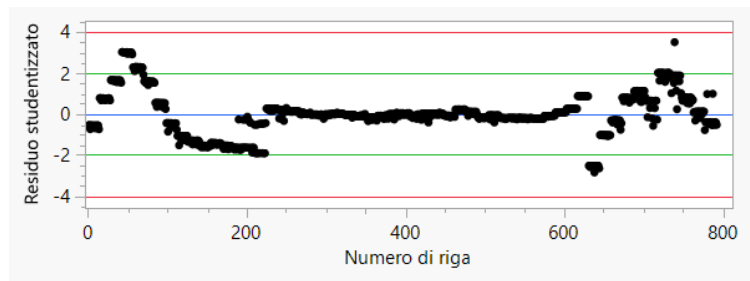


Figura A.49– residui studentizzati- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025

Press

Residui	SSE	RMSE	R-quadro
Press	8643.7467727	3.30360737	0.7018
Ordinario	8500.6526779	3.29702894	0.7067

Tabella A.31– statistica PRSS- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025

Durbin-Watson

Durbin-Watson	Numero di osservazioni	Autocorrelazione	Prob< DW
0.136457	792	0.9314	<.0001 *

Tabella A.32– test di Durbin-Watson- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025

Diagramma dei quantili normali residui

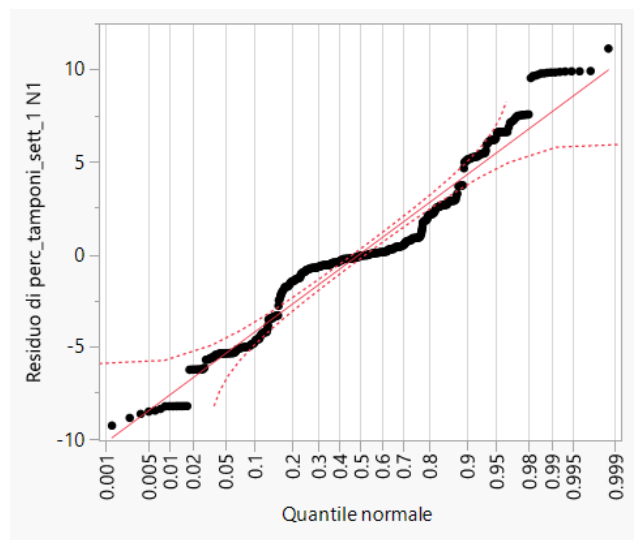


Figura A.50– diagramma dei quantili normali residui- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025

Valori osservati vs previsti

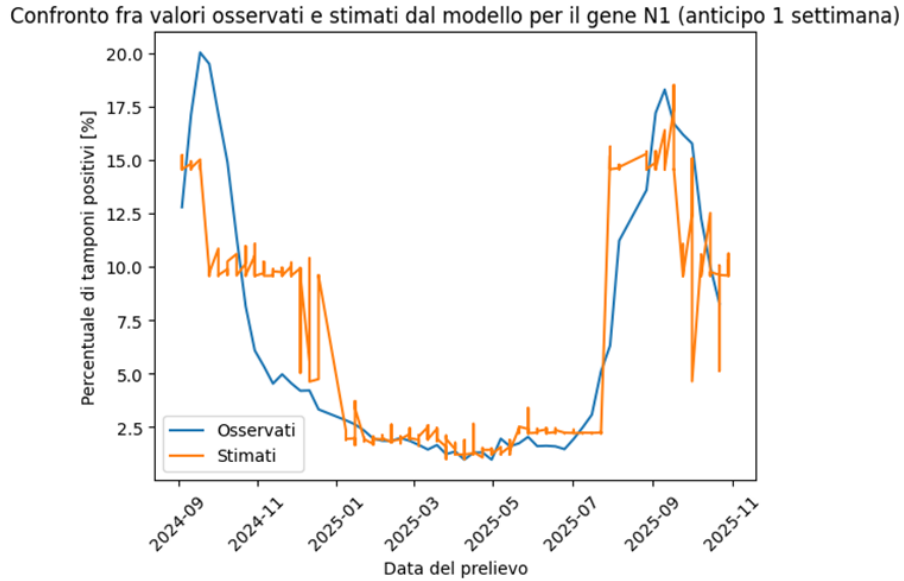


Figura A.51– valori osservati vs previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025

Modello % tamponi positivi N_1 -prevalenza WBE-anticipo 2 settimana: risposta % tamponi N_1 -anticipo 2 settimana

Grafico delle risposte osservate rispetto a quelle previste

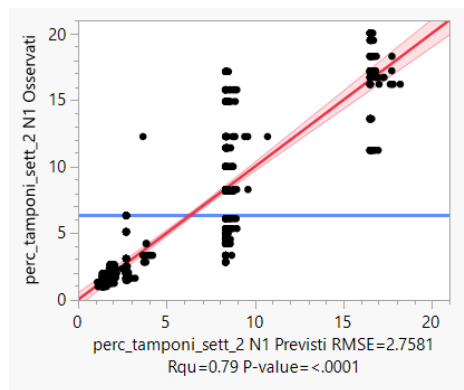


Figura A.52-- Grafico delle risposte osservate rispetto a quelle previste Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025

Riepilogo della stima

R-quadro	0.793884
R-quadro corretto	0.791465
Scarto quadratico medio	2.758105
Media della risposta	6.373144
Osservazioni (o somme pesate)	777

Tabella A.33— riepilogo della stima Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025

Analisi della varianza ANOVA

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Modello	9	22473.016	2497.00	328.2445
Errore	767	5834.677	7.61	Prob > F
C. totale	776	28307.693		<.0001*

Tabella A.34— analisi della varianza ANOVA- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025

Mancata stima

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Mancata stima	645	5122.8211	7.94236	1.3612
Errore puro	122	711.8557	5.83488	Prob > F
Errore totale	767	5834.6768		0.0180*

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
				Max R-quadro
				0.9749

Tabella A.35— mancata stima- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025

Stime dei parametri

Termine	Stima	Errore std	Rapporto T	Prob> t
Intercetta	6.1432381	0.116066	52.93	<.0001*
Prevalenza_WBE	0.1345568	0.070441	1.91	0.0565
Stagioni [inverno]	-4.341024	0.221049	-19.64	<.0001*
Stagioni primavera]	-4.816864	0.232119	-20.75	<.0001*
Stagioni [estate]	-3.003171	0.319563	-9.40	<.0001*
Stagioni [estate/autunno]	10.493725	0.29206	35.93	<.0001*
Stagioni [inverno]*(>3%-0.11068)	-0.061717	0.465954	-0.13	0.8947
Stagioni [primavera]*(influenza>3%-0.11068)	-0.401999	0.917906	-0.44	0.6615
Stagioni [estate]*(influenza>3%-0.11068)	3.9135379	2.338515	1.67	0.0946
Stagioni [estate/autunno]*(influenza>3%-0.11068)	1.2228124	1.966563	0.62	0.5343

Tabella A.36— stime dei parametri- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025

Test degli effetti

Origine	Npar	DF	Somma dei quadrati	Rapporto F	Prob > F
Prevalenza_WBE	1	1	27.758	3.6489	0.0565
Stagioni	4	4	14265.890	468.8322	<.0001*
Stagioni*influenza >3%	4	4	310.910	10.2177	<.0001*

Tabella A.37-- test degli effetti-Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025

Diagramma dei residui rispetto ai previsti

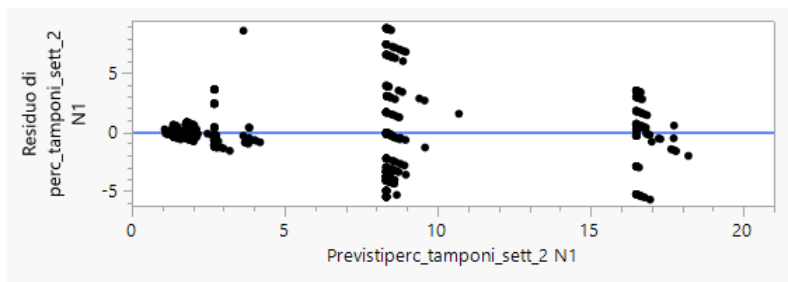


Figura A.53-- diagramma dei residui rispetto ai previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) - anticipo 2 settimane- 2024-2025

Diagramma dei valori residui per riga

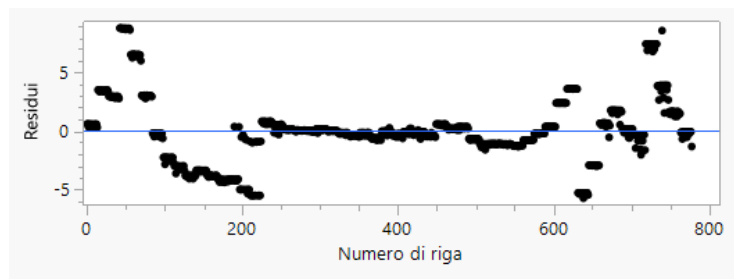


Figura A.54-- diagramma dei valori residui per riga- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025

Residui studentizzati esternamente con limiti simultanei al 95% (Bonferroni) in rosso, limiti individuali in verde

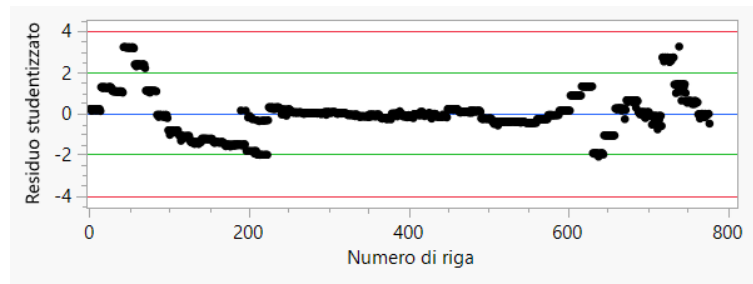


Figura A.55— residui studentizzati- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025

Press

Residui	SSE	RMSE	R-quadro
Press	5921.1524517	2.7605309	0.7908
Ordinario	5834.6767982	2.75810452	0.7939

Tabella A.38— statistica Press-Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025

Durbin-Watson

Durbin-Watson	Numero di osservazioni	Autocorrelazione	Prob<D W
0.1411189	777	0.9293	<.0001*

Tabella A.39— test di Durbin-Watson-Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025

Diagramma dei quantili normali residui

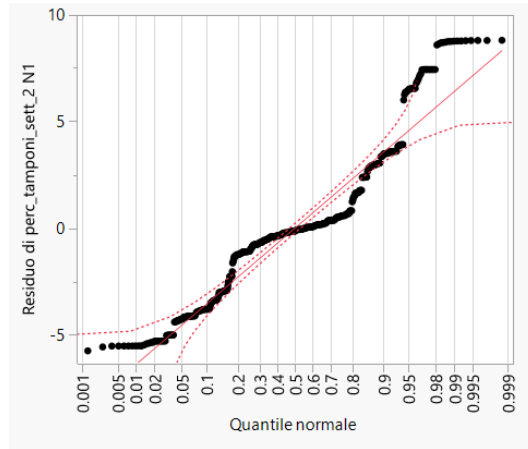


Figura A.56— diagramma dei quantili normali residui- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) - anticipo 2 settimane- 2024-2025

Valori osservati vs previsti

Confronto fra valori osservati e stimati dal modello per il gene N1 (anticipo 2 settimane)

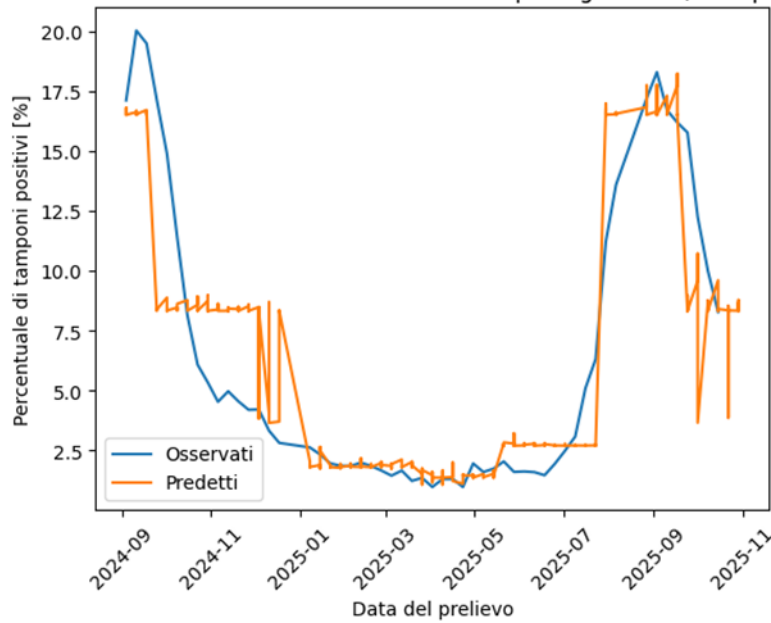


Figura A.57—valori osservati vs previsti- Modelli di regressione lineare multipla-% tamponi positivi N₁-prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025

Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Grafico delle risposte osservate rispetto alle risposte previste

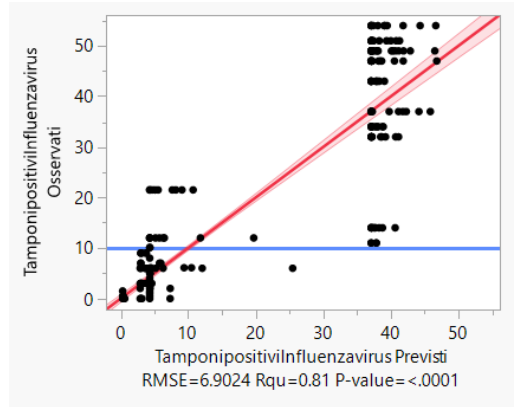


Figura A.58—grafico delle risposte osservate rispetto alle risposte previste- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Riepilogo della stima

R-quadro	0.806978
R-quadro corretto	0.805751
Scarto quadratico medio	6.902408
Media della risposta	9.945776
Osservazioni (o somme pesate)	793

Tabella A.40—riepilogo della stima- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Analisi della varianza ANOVA

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Modello	5	156758.15	31351.6	658.0500

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Errore	787	37495.22	47.6	Prob > F
C. totale	792	194253.37		<.0001*

Tabella A.41—analisi della varianza ANOVA- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Mancata stima

Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Mancata stima	164	20261.803	123.548	4.4663
Errore puro	623	17233.419	27.662	Prob > F
Errore totale	787	37495.222		<.0001*
				Max R-quadro
				0.9113

Tabella A.42—mancata stima- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Stime dei parametri

Termine	Stima	Errore std	Rapporto T	Prob> t
Intercetta	9.5071849	0.2654	35.82	<.0001*
Stagioni[inverno]	27.574846	0.520054	53.02	<.0001*
Stagioni[primavera]	-6.53273	0.568501	-11.49	<.0001*
Stagioni[estate]	-9.211127	0.523967	-17.58	<.0001*
Stagioni	-6.6231	0.561601	-11.79	<.0001*

Termine	Stima	Errore std	Rapporto T	Prob> t
[estate/autunno]				
Prevalenza_influenza	20.05891	3.322148	6.04	<.0001*

Tabella A.43—stime dei parametri- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Test degli effetti

Origine	Npar	DF	Somma dei quadrati	Rapporto F	Prob > F
Stagioni	4	4	136020.38	713.7445	<.0001*
Prevalenza_ influenza	1	1	1736.91	36.4567	<.0001*

Tabella A.44—test degli effetti- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Diagramma dei residui rispetto ai previsti

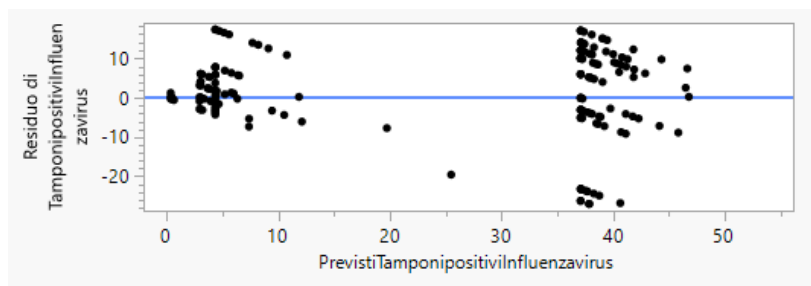


Figura A.59—diagramma dei residui rispetto ai previsti- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Diagramma dei valori residui per riga

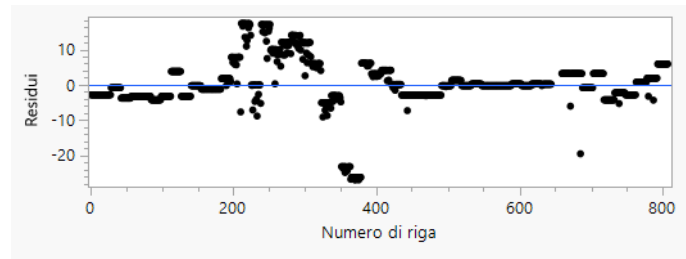


Figura A.60—diagramma dei valori residui per riga- Modelli di regressione lineare multipla-
% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Residui studentizzati esternamente con limiti simultanei al 95% (Bonferroni) in rosso,
limiti individuali in verde

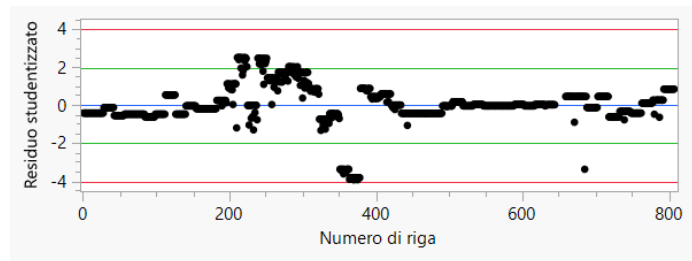


Figura A.61—residui studentizzati- Modelli di regressione lineare multipla-% Tamponi
positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Press

Residui	SSE	RMSE	R-quadro
Press	38446.693291	6.96294401	0.8021
Ordinario	37495.222369	6.90240758	0.8070

Tabella A.45—statistica PRESS- Modelli di regressione lineare multipla-% Tamponi positivi
influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Durbin-Watson

Durbin-Watson	Numero di osservazioni	Autocorrelazione	Prob< DW
0.1698763	793	0.9145	<.0001 *

Tabella A.46—test di Durbin-Watson- Modelli di regressione lineare multipla-% Tamponi
positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Diagramma dei quantili normali residui

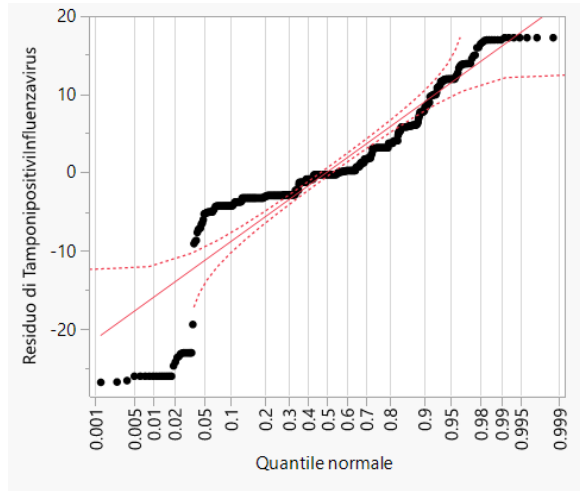


Figura A.62—diagramma dei quantili normali residui- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Valori osservati vs previsti

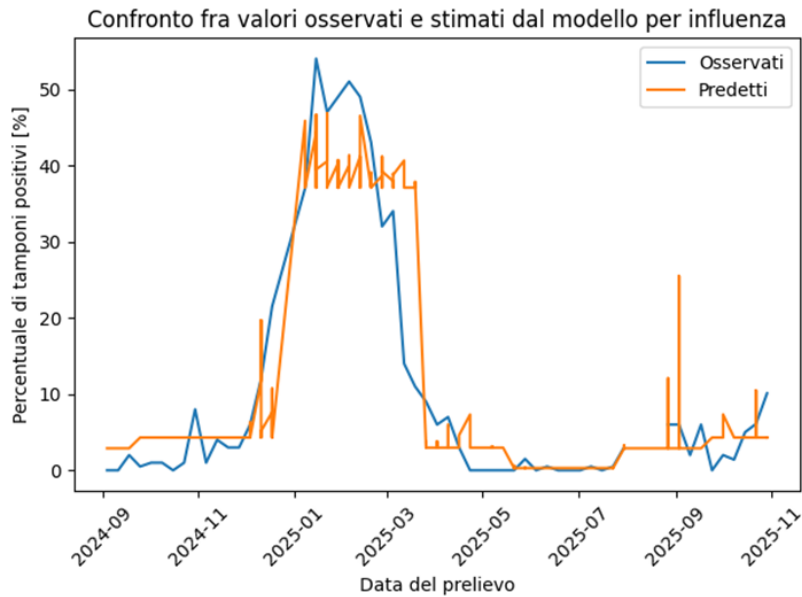


Figura A.63—valori osservati vs previsti- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Lista di figure

Figura 3.1 – boxplot gene ORF per impianto-anno 2022	40
Figura 3.2 – boxplot gene N_1 per impianto-anno 2022	41
Figure 3.3 – boxplot gene N_3 per impianto-anno 2022	41
Figura 3.4 – scatter plot dei nuovi casi vs cariche grezze e normalizzate	46
Figura 3.5 – diagramma di regressione-modello di regressione lineare semplice N_1 quantitativo-nuovi casi-anno 2022	49
Figura 3.6 – diagramma dei quantili normali residui-modello di regressione lineare semplice N_1 quantitativo-nuovi casi-anno 2022	50
Figura 3.7 – diagramma dei residui vs previsti-modello di regressione lineare semplice N_1 quantitativo-nuovi casi-anno 2022	50
Figura 3.8 – nuovi casi giornalieri vs nuovi casi predetti dal modello: gene N_1 quantitativo-nuovi	55
Figura 3.9 – diagramma dei residui rispetto ai previsti-modello di regressione lineare multipla N_1 quantitativo-nuovi casi (sito e variante) -anno 2022	55
Figura 3.10 – diagramma dei quantili normali residui-modello di regressione lineare multipla N_1 quantitativo-nuovi casi (sito e variante) -anno 2022	56
Figura 3.11 – % tamponi positivi N_1 osservati vs predetti-modello di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane-media settimanale-anno 2022	61
Figura 3.12 – diagramma dei residui rispetto ai previsti-modello di regressione lineare multipla -% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anno 2022	62
Figura 3.13 – diagramma dei quantili normali residui-modello di regressione lineare multipla -% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione con influenza) -anno 2022	63

- Figura 3.14 – confronto fra valori osservati e stimati dal modello di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane-anno 2024-2025 69
- Figura 3.15 –diagramma dei residui rispetto ai previsti-modello di regressione lineare multipla -% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza - anticipo 2 settimane-anno 2024-2025 69
- Figura 3.16 – diagramma dei quantili normali residui-modello di regressione lineare multipla -% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane-anno 2024-2025 70
- Figura 3.17 – confronto fra valori osservati e stimati dal modello di regressione lineare multipla-influenza-prevalenza WBE (stagione)-anno 2024-2025 72
- Figura 3.18 – diagramma dei residui rispetto ai previsti-modello di regressione lineare multipla-influenza-prevalenza WBE (stagione)-anno 2024-2025 72
- Figura 3.19 – diagramma dei quantili normali residui-modello di regressione lineare multipla-influenza-prevalenza WBE (stagione)-anno 2024-2025 73
- Figura A.1– nuovi casi giornalieri vs predetti dal modello- Modello di regressione lineare multipla- Modello N_1 quantitativo -nuovi casi (sito e variante) -2022 81
- Figura A.2– nuovi casi giornalieri vs predetti dal modello - Modello di regressione lineare multipla- Modello N_1 normalizzato N -nuovi casi (sito e variante) -2022 81
- Figura A.3– nuovi casi giornalieri vs predetti dal modello- Modello di regressione lineare multipla- Modello N_1 normalizzato AE -nuovi casi (sito e variante) -2022 82
- Figura A.4– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Bergamo-2022 82
- Figura A.5– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Bresso-2022 83
- Figura A.6– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Como-2022 83
- Figura A.7– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Città di Cremona-2022 84

Figura A.8– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Lonate Pozzolo-2022	84
Figura A.9– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Milano Nosedo-2022	85
Figura A.10– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Milano San Rocco-2022	85
Figura A.11– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Monza-2022	86
Figura A.12– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Pavia-2022	86
Figura A.13– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Peschiera Borromeo-2022	87
Figura A.14– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Sondrio-2022	87
Figura A.15– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Verziano-2022	88
Figura A.16– confronto tra numero di casi infetti N_1 : casi WBE vs casi clinici -Vigevano-2022	88
Figura A.17– grafico delle risposte rispetto alle risposte previste- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022	89
Figura A.18– influenza>3% per stagione –2022	91
Figura A.19– diagramma dei residui rispetto ai previsti, - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022	92
Figura A.20– diagramma dei valori residui per riga - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022	92

- Figura A.21– residui studentizzati - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022 92
- Figura A.22– diagramma dei quantili normali residui - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022 93
- Figura A.23– valori osservati vs previsti - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022 94
- Figura A.24– valori osservati vs previsti-media settimanale - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022 94
- Figura A.25– grafico delle risposte osservate rispetto alle risposte osservate previste- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022 95
- Figura A.26– diagramma dei residui rispetto ai previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022 97
- Figura A.27– diagramma dei valori residui per riga- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022 97
- Figura A.28– residui studentizzati- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022 98
- Figura A.29– diagramma dei quantili normali residui- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022 99
- Figura A.30– valori osservati vs previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022 99
- Figura A.31– valori osservati vs previsti-media settimanale- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022 100

Figura A.32– grafico delle risposte osservate rispetto alle risposte previste- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	100
Figura A.33– diagramma dei residui rispetto ai previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	102
Figura A.34– diagramma dei valori residui per riga- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	102
Figura A.35– residui studentizzati- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	103
Figura A.36– diagramma dei quantili normali residui- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	104
Figura A.37– valori osservati vs previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	104
Figura A.38– valori osservati vs previsti-media settimanale- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	105
Figura A.39– grafico delle risposte osservate rispetto alle risposte previste- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2024-2025	105
Figura A.40– % tamponi positivi influenza -2024-2025	108
Figura A.41– diagramma dei residui rispetto ai previsti - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) 2024-2025	108
Figura A.42– diagramma dei valori residui per riga - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) 2024-2025	109

- Figura A.43– residui studentizzati - Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) 2024-2025 109
- Figura A.44– diagramma dei quantili normali residui- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) 2024-2025 110
- Figura A.45– valori osservati vs previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) 2024-2025 110
- Figura A.46– grafico delle risposte osservate rispetto alle risposte previste- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025 111
- Figura A.47– diagramma dei residui rispetto ai previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025 114
- Figura A.48– diagramma dei valori residui per riga- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025 114
- Figura A.49– residui studentizzati- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025 114
- Figura A.50– diagramma dei quantili normali residui- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025 116
- Figura A.51– valori osservati vs previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025 116
- Figura A.52– Grafico delle risposte osservate rispetto a quelle previste Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025 116

- Figura A.53-- diagramma dei residui rispetto ai previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025 119
- Figura A.54-- diagramma dei valori residui per riga- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025 119
- Figura A.55-- residui studentizzati- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025 120
- Figura A.56-- diagramma dei quantili normali residui- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025 121
- Figura A.57--valori osservati vs previsti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025 121
- Figura A.58--grafico delle risposte osservate rispetto alle risposte previste- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025 122
- Figura A.59--diagramma dei residui rispetto ai previsti- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025 124
- Figura A.60--diagramma dei valori residui per riga- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025 125
- Figura A.61--residui studentizzati- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025 125
- Figura A.62--diagramma dei quantili normali residui- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025 126

Figura A.63--valori osservati vs previsti- Modelli di regressione lineare multipla-%
Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025

Lista di tabelle

Tabella 2.1 – tabella shedding rate-studio Wan Yang et al-tutte le ondate di varianti	30
Tabella 2.2 – tabella shedding rate-studio Wan Yang et al-stratificato per ondata/periodo	31
Tabella 3.1 – AE nominali per sito di prelievo	36
Tabella 3.2 – Descrizione del dataset 2022	38
Tabella 3.3 – numerosità dei geni-2022	38
Tabella 3.4 – numerosità dei geni per impianto-2022	39
Tabella 3.5 – statistiche descrittive variabili geni-anno 2022	42
Tabella 3.6 – numerosità dei geni grezzi e normalizzati-anno 2022	44
Tabella 3.7 – correlazione di Pearson fra dati clinici e ambientali-anno 2022	45
Tabella 3.8 – parametri principali dei modelli di regressione lineare semplice-anno 2022	48
Tabella 3.9 – parametri principali dei modelli di regressione lineare multipla (con sito e variante) -anno 2022	52
Tabella 3.10 – riepilogo della stima-modello di regressione lineare multipla N_1 quantitativo-nuovi casi (sito e variante) -anno 2022	53
Tabella 3.11 – tabella stima dei parametri-modello di regressione lineare multipla N_1 quantitativo-nuovi casi (sito e variante) -anno 2022	53
Tabella 3.12— parametri principali dei modelli di regressione lineare multipla-gene N_1 -prevalenza (stagione ed interazione influenza) -anno 2022	59
Tabella 3.13— riepilogo della stima-modello di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane -anno 2022	60
Tabella 3.14—tabella stima dei parametri -modello di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane-anno 2022	60

Tabella 3.15— descrizione del dataset 2024-2025	64
Tabella 3.16— tabella descrittiva del dataset 2024-2025	65
Tabella 3.17— principali statistiche descrittive-dataset 2024-2025	65
Tabella 3.18— principali parametri dei modelli di regressione lineare multipla riguardanti il gene N_1 -prevalenza WBE (stagione ed interazione influenza) -anno 2024-2025.	66
Tabella 3.19— principali parametri del modello di regressione lineare multipla dell'influenza-prevalenza WBE (stagione)-anno 2024-2025	67
Tabella 3.20— riepilogo della stima-modello di regressione lineare multipla-%tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane-anno 2024-2025	68
Tabella 3.21— tabella stima dei parametri -modello di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane-anno 2024-2025	68
Tabella 3.22— riepilogo della stima-modello di regressione lineare multipla-influenza-prevalenza WBE (stagione)-anno 2024-2025	71
Tabella 3.23— tabella stima dei parametri-modello di regressione lineare multipla-influenza-prevalenza WBE (stagione)-anno 2024-2025	71
Tabella A.1—riepilogo della stima- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022	90
Tabella A.2—analisi della varianza ANOVA- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022	90
Tabella A.3—stime dei parametri- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022	90
Tabella A.4—test degli effetti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022	91
Tabella A.5—statistica PRESS- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022	93
Tabella A.6—test di Durbin-Watson- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2022	93

Tabella A.7– riepilogo della stima- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022	95
Tabella A.8– analisi della varianza ANOVA- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022	96
Tabella A.9–stime dei parametri- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022	96
Tabella A.10–test degli effetti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022	97
Tabella A.11–statistica PRESS- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022	98
Tabella A.12–test di Durbin-Watson- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 1 settimana-2022	98
Tabella A.13–riepilogo della stima- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	101
Tabella A.14–analisi della varianza ANOVA- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	101
Tabella A.15–stime dei parametri- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	102
Tabella A.16–test degli effetti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	102

Tabella A.17–statistica PRESS- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	103
Tabella A.18–test di Durbin-Watson- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) anticipo 2 settimane-2022	103
Tabella A.19– riepilogo della stima- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2024-2025	106
Tabella A.20– analisi della varianza ANOVA- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2024-2025	106
Tabella A.21– mancata stima- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2024-2025	107
Tabella A.22– stime dei parametri- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2024-2025	108
Tabella A.23– test degli effetti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2024-2025	108
Tabella A.24– statistica PRESS- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2024-2025	109
Tabella A.25– test di Durbin-Watson- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -2024-2025	110
Tabella A.26– riepilogo della stima- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025	111
Tabella A.27– analisi della varianza ANOVA- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025	112

Tabella A.28– mancata stima- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025	112
Tabella A.29– stime dei parametri- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025	113
Tabella A.30– test degli effetti- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025	113
Tabella A.31– statistica PRSS- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025	115
Tabella A.32– test di Durbin-Watson- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 1 settimana- 2024-2025	115
Tabella A.33– riepilogo della stima Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025	117
Tabella A.34– analisi della varianza ANOVA- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025	117
Tabella A.35– mancata stima- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025	118
Tabella A.36– stime dei parametri- Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025	118
Tabella A.37– test degli effetti-Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025	119

Tabella A.38-- statistica Press-Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025	120
Tabella A.39-- test di Durbin-Watson-Modelli di regressione lineare multipla-% tamponi positivi N_1 -prevalenza WBE (stagione ed interazione influenza) -anticipo 2 settimane- 2024-2025	120
Tabella A.40--riepilogo della stima- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025	122
Tabella A.41--analisi della varianza ANOVA- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025	123
Tabella A.42--mancata stima- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025	123
Tabella A.43--stime dei parametri- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025	123
Tabella A.44--test degli effetti- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025	124
Tabella A.45--statistica PRESS- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025	125
Tabella A.46--test di Durbin-Watson- Modelli di regressione lineare multipla-% Tamponi positivi influenza A-prevalenza WBE influenza (stagione) -anno 2024-2025	125

