

Analisi esplorativa dell'andamento dell'influenza negli U.S.A. dal 2017 al 2022

Gianluca Coletta

Dipartimento di Ingegneria
Università degli Studi di Perugia
Perugia, Italia
gianluca.coletta@studenti.unipg.it

Abstract—In questo documento è descritta un'analisi esplorativa dell'andamento dell'influenza negli Stati Uniti d'America dal 2017 al 2022. L'obiettivo è di verificare se ci sono delle differenze dagli anni precedenti all'inizio della pandemia di COVID-19 (2017-2019), rispetto agli anni successivi (2020-2022).

Index Terms—EDA, FluNet, USA, WHO, COVID-19.

I. INTRODUZIONE

L'influenza è una malattia infettiva respiratoria acuta causata dal virus dell'influenza, un virus a RNA della famiglia degli Orthomyxoviridae [1]. All'inizio del 2020 ha cominciato a diffondersi nel mondo una nuova malattia respiratoria, conosciuta come SARS-CoV-2, che ha stravolto la vita delle persone. Per contrastare questa nuova pandemia sono state adottate varie misure di sicurezza; distanziamento obbligatorio, utilizzo di mascherine e una maggiore attenzione all'igiene. In questo studio andremo ad analizzare i dati relativi all'influenza degli Stati Uniti d'America nel periodo che va dal 2017 al 2022, osservando come le contromisure adottate per il COVID-19 abbiano impattato nel numero di casi positivi di influenza.

II. DESCRIZIONE DEL DATASET

Il dataset utilizzato proviene da **FluNet**, presente nel sito della *World Health Organization*. Questo è uno strumento globale utilizzato per la sorveglianza virologica dell'influenza. I dati sono forniti in remoto dai National Influenza Centers (NICs) del Global Influenza Surveillance and Response System (GISRS) e altri laboratori nazionali di riferimento sull'influenza che collaborano attivamente con il GISRS, oppure sono caricati dai database regionali dell'OMS [2]. Il database si occupa di tenere traccia dei virus influenzali diversificandoli in otto sottotipi; A(H1), A(H1N1/2009), A(H3), A(H5), A(Unsubtyped), B(Yamagata), B(Victoria) and B(Undetermined). I dati vengono raccolti settimanalmente dal 1995 ad oggi, includendo 29 nazioni. In questo studio si è scelto di analizzare una singola nazione come esempio, ovvero gli Stati Uniti D'America, per una questione di numero e completezza di dati. Quindi è stato utilizzato il dataset degli USA da Gennaio 2017 fino a Maggio 2022 [3].

III. MODELLAZIONE DEL DATASET

Prima di iniziare l'analisi esplorativa dei dati, sono state apportate varie modifiche al dataset. Durante una prima osservazione del dataset si può subito notare che, per quanto

riguarda due tipologie di influenza, A(H1) e A(H5), non si sono manifestati casi negli anni presi in considerazione, di conseguenza sono state rimosse le relative colonne. Inoltre sono state eliminate ulteriori variabili inutili, per esempio *country*, *WHORegion*, *FluRegion*. I casi sono riportati settimanalmente, con la relativa data di inizio e di fine settimana, perciò si è scelto di raggrupparli per mese, tenendo conto della data di fine settimana per assegnarli ad un mese specifico. Questo perché, avendo preso in considerazione cinque anni, renderà la visualizzazione e l'analisi dei dati più corretta. Il campo *date*, scritto nel formato **ISO 8601**, è stato convertito dal tipo *character* al tipo *dates*, per permettere di raggruppare i dati per mese. Infine sono state rinominate alcuni variabili in maniera più appropriata, per facilitare la lettura del dataset. Il risultato di questa serie di operazioni è un nuovo dataset composto dalle seguenti variabili:

- **date**: data espressa in formato **ISO 8601**.
- **month**: mese espresso in formato alfanumerico.
- **year**: anno espresso in formato numerico.
- **processed_samples**: campioni esaminati in uno specifico mese e anno.
- **AH1N12009**: casi positivi di influenza A(H1N1/2009) in uno specifico mese e anno.
- **AH3**: casi positivi di influenza A(H3) in uno specifico mese e anno.
- **ANotSubTyped**: casi positivi di influenza A(unsubtyped) in uno specifico mese e anno.
- **total_A**: casi totali positivi di influenza A in uno specifico mese e anno.
- **BYamagata**: casi positivi di influenza B(Yamagata) in uno specifico mese e anno.
- **BVictoria**: casi positivi di influenza B(Victoria) in uno specifico mese e anno.
- **BNotDetermined**: casi positivi di influenza B(Undetermined) in uno specifico mese e anno.
- **total_B**: casi totali positivi di influenza B in uno specifico mese e anno.
- **total_positive**: casi totali positivi di influenza in uno specifico mese e anno.
- **total_negative**: casi totali negativi di influenza in uno specifico mese e anno.

IV. ANALISI ESPLORATIVA

L'analisi esplorativa del dataset è suddivisa in parti distinte:

- Prima osservazione e test di normalità sui campioni esaminati e sul numero di positivi totale nei 5 anni presi in considerazione.
- Divisione del dataset in due parti, anni precedenti (2017, 2018, 2019) e successivi (2020, 2021, 2022) all'avvento del COVID-19, con conseguente analisi dei due dataset.
- Indagine sull'evoluzione dei casi delle diverse tipologie di influenza, prima e dopo l'inizio della pandemia COVID-19.

A. Analisi preliminare

Inizialmente si è osservato l'andamento del numero di casi positivi dal 2017 al 2022, notando da subito il classico andamento dell'influenza, ovvero un'aumento di casi positivi da Novembre a Marzo, per poi diminuire nei mesi successivi. Inoltre si può vedere come, durante i primi due anni di COVID-19 (da Marzo 2020 a Novembre 2021), ci sia stato un netto calo nel numero di positivi per influenza, con numeri molto vicini allo zero. Invece, per quanto riguarda il periodo che va da Novembre 2021 a Maggio 2022 vi è un leggero aumento di casi, probabilmente dovuto alla riduzione delle contromisure adottate per il COVID-19. Queste osservazioni si possono notare dalla Figura 1.

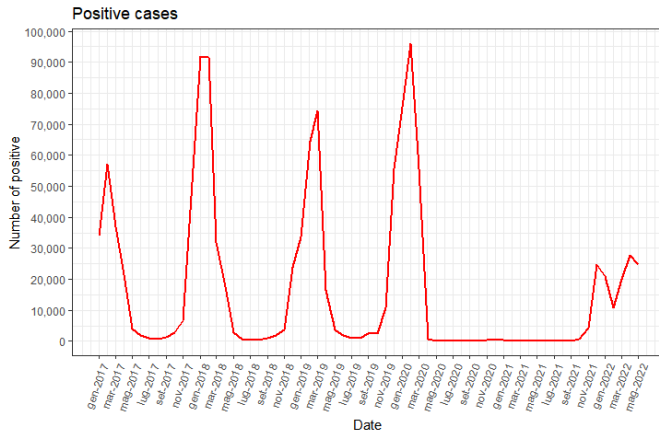


Fig. 1. Andamento casi di influenza dal 2017 al 2022.

Successivamente si è analizzato l'andamento dei campioni esaminati in questi 5 anni, per verificare se il grande calo di numeri positivi sia la conseguenza di un minor numero di campioni processati durante la pandemia. Ma osservando il grafico si evince che non vi è stato un forte calo di campioni analizzati, bensì i numeri rimangono in linea con gli anni precedenti alla comparsa del COVID-19. Inoltre è interessante vedere come il picco si verifica a Gennaio 2022, sfiorando i 900.000 campioni processati. Queste considerazioni si possono notare dalla Figura 2.

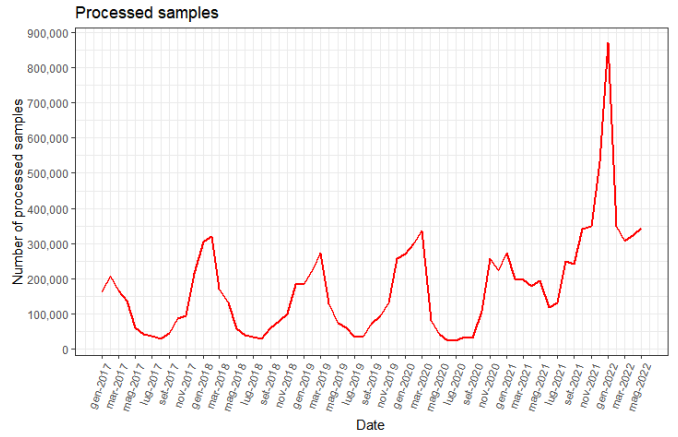


Fig. 2. Andamento campioni esaminati dal 2017 al 2022.

In seguito ho condotto dei test di adattamento per verificare la normalità dei dati. Prima di tutto sono andato ad osservare i due istogrammi relativi ai campioni esaminati e casi positivi di influenza. I risultati sono riportati nella Figura 3.

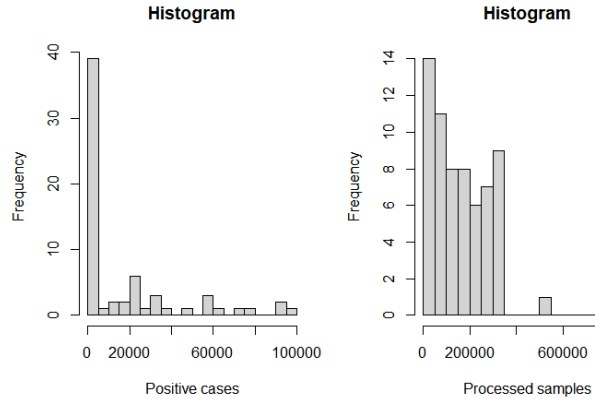


Fig. 3. Istogrammi relativi alle distribuzioni dei casi positivi e dei campioni esaminati.

Dai grafici si può notare come i dati non appartengano ad una distribuzione normale; per confermare questa considerazione è stato effettuato il controllo più potente per la verifica della normalità, il test di Shapiro-Wilk. Nella Tabella I vengono mostrati i risultati.

TABLE I
TEST DI SHAPIRO-WILK

Feature	Statistica W	P-value
Positive cases	0.70038	0.0000000003172
Processed samples	0.82423	0.0000002494

Con i valori ottenuti del p-value si può rifiutare l'ipotesi nulla del test di Shapiro-Wilk (entrambi inferiori a 0.01), ovvero che i dati siano stati generati da un processo normale.

B. Confronto: prima del COVID-19 e durante il COVID-19

Per effettuare un confronto più accurato ho deciso di prendere in considerazione il tasso di positività dell'influenza, ottenuto tramite l'Equazione (1).

$$Tasso_Positivita\acute = 100 \cdot \frac{Casi_Positivi}{Campioni_Esaminati} \quad (1)$$

Per prima cosa possiamo subito osservare, come prima che iniziasse il COVID-19, il tasso di positività non abbia mai superato il 30%, picco verificatosi a Gennaio del 2018. Mentre il picco di positività nel 2017 si è verificato a Febbraio con il 27.5%. Infine, per quanto riguarda il 2019, la percentuale più alta di positivi si è verificata a Febbraio, con un valore vicino al 28%. I risultati sono visibili nella Figura 4.

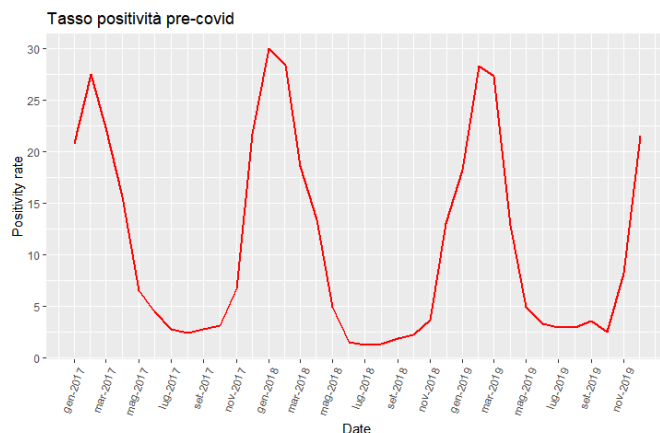


Fig. 4. Andamento tasso di positività prima del COVID-19.

Successivamente è stato analizzato il tasso di positività dopo l'arrivo del COVID-19; è evidente come l'andamento cambia drasticamente. Nei primi mesi del 2020 (Gennaio e Febbraio) in realtà si nota un alto tasso di positività (picco massimo del 32%), dovuto dal fatto che il COVID-19 era appena comparso; si conoscevano pochi dettagli a riguardo e, di conseguenza, non erano ancora state adottate delle forti contromisure. Mentre, nel periodo che va da Aprile 2020 fino a Novembre 2021, il tasso di positività è approssimabile allo 0%. Invece, nel successivo lasso di tempo, fino a Maggio 2022, si può notare una piccola risalita dei casi di influenza, avendo inaspettatamente il picco massimo, vicino all'8%, nel mese di Aprile. I risultati sono visibili nella Figura 5.

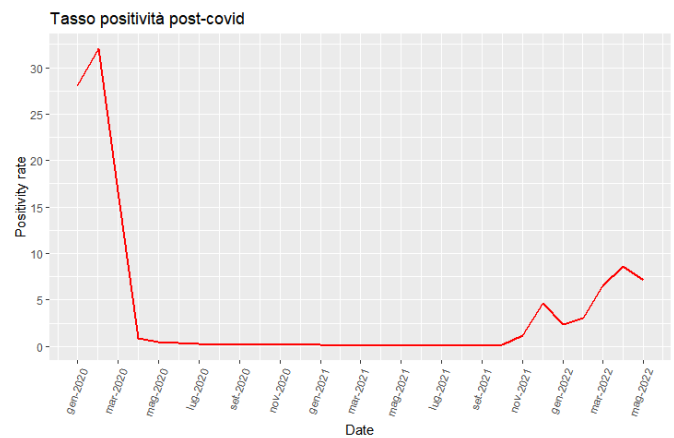


Fig. 5. Andamento tasso di positività dopo l'arrivo del COVID-19.

Questi risultati sono ancor più evidenti confrontando il tasso di positività nei diversi anni per ogni singolo mese; si nota come nel mese di Gennaio e Febbraio del 2021 e 2022 il tasso sia nettamente più basso, vicinissimo allo zero. I risultati sono visibili nella Figura 6.

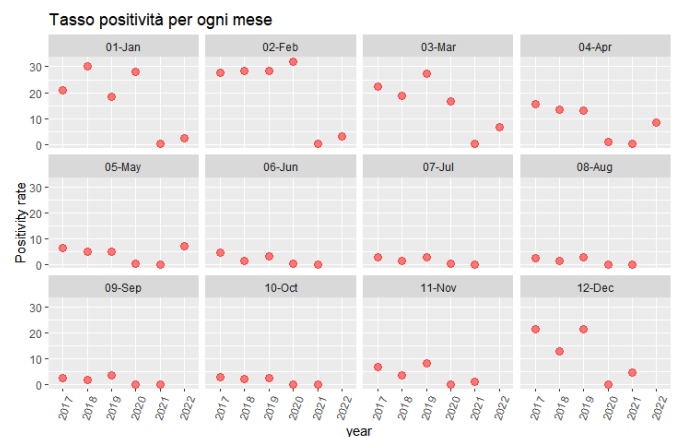


Fig. 6. Tasso positività per ogni singolo mese.

C. Indagine sull'andamento dei casi di diverse tipologie di influenza

In primo luogo si è scelto di confrontare i casi di influenza di *tipo A* con quelli di *tipo B*; solitamente il primo tipo colpisce maggiormente rispetto al secondo, si stima che le infezioni da *influenza A* rappresentino il 75% dei casi di influenza stagionale. Inoltre l'*influenza A* può essere rilevata in molte specie, come gli esseri umani, uccelli e suini, mentre l'*influenza B* può colpire, tipicamente, solo gli esseri umani. In passato, si credeva che l'infezione da *influenza A* fosse più grave dell'infezione da *influenza B*; ma, uno studio svolto nel 2015 sugli adulti ha rilevato che entrambe le influenze provocano tassi simili di malattia e morte [4]. Dal confronto si evince che, anche durante il periodo 2017-2022, l'*influenza di tipo A* ha un'incidenza maggiore rispetto a quella di *tipo B*. I risultati sono mostrati in Figura 7.

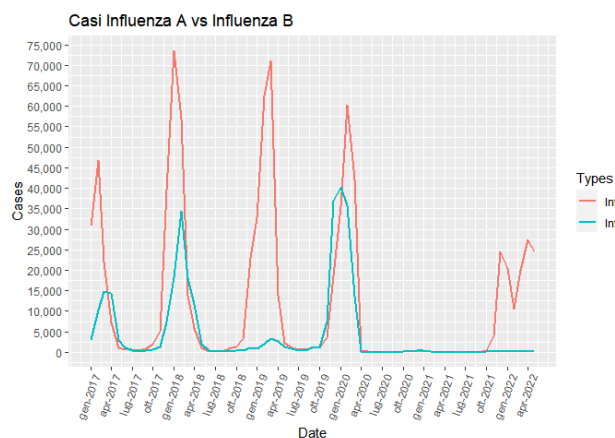


Fig. 7. Confronto tra i casi di influenza tipo A con influenza tipo B.

Successivamente ho analizzato i casi legati alle diverse tipologie di influenza; ho preso in considerazione *A(H1N1/2009)*, *A(H3)*, *B(Yamagata)* e *B(Victoria)*, escludendo *A(Unsubtyped)* e *B(Undetermined)*, essendo casi non appartenenti ad una tipologia conosciuta. Ho suddiviso l'analisi dei diversi casi nel periodo precedente e successivo al COVID-19. Si osserva, prima dell'inizio del COVID-19, che la tipologia con incidenza nettamente maggiore rispetto alle altre è stata la *A(H3)*, con un picco di 12.000 casi nel gennaio 2018. Ma, nel periodo che va da Novembre 2018 a Maggio 2019, i casi di influenza di tipo *A(H1N1/2009)* sono molto simili a quelli di tipo *A(H3)*. Mentre per quanto riguarda l'influenza di tipo *B*, si nota un buono numero di casi, da Gennaio 2017 ad Aprile 2018, di tipo *B(Yamagata)*, con il picco nel Febbraio 2018 di circa 3.500 casi. Mentre, da Aprile 2018 a Novembre 2019, si sono osservati pochissimi casi di influenza tipo *B*, per poi avere un aumento di casi di tipo *B(Victoria)* a Dicembre 2019. I risultati sono visibili in Figura 8.

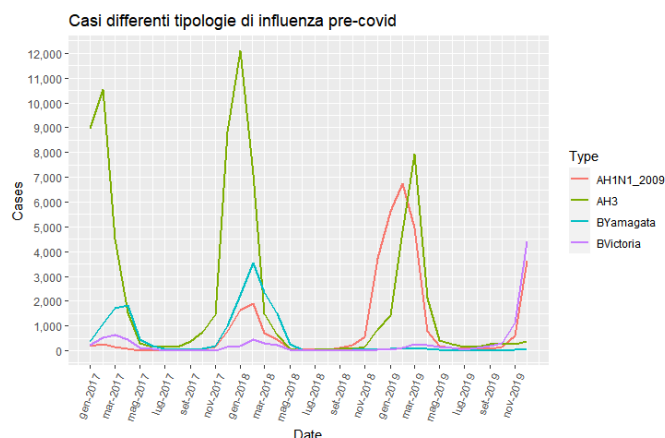


Fig. 8. Confronto tra le varie tipologie di influenza pre-COVID.

Dopo l'avvento del COVID-19, la tipologia che avuto un impatto maggiore rispetto alle altre è stata, di nuovo, la *A(H3)*, toccando il picco di 4000 casi (nettamente inferiore al picco prima dell'arrivo del COVID-19), nel Dicembre 2021.

Inoltre si è verificato un aumento di casi di influenza di tipo *A(H1N1/2009)*, sfiorando i 9.000 casi a Febbraio 2020. Le precedenti considerazioni sono visibili in Figura 9.

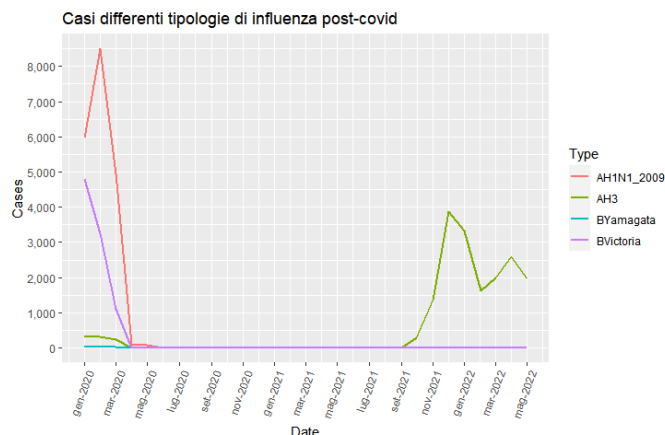


Fig. 9. Confronto tra le varie tipologie di influenza post-COVID.

V. CONCLUSIONI

Le analisi effettuate dimostrano come l'andamento del numero di positivi all'influenza abbia subito un netto calo nel periodo successivo all'arrivo del COVID-19, dovuto, probabilmente, all'utilizzo di misure di contenimento contro la nuova pandemia. Nonostante questo studio si sia basato su dati raccolti negli Stati Uniti d'America, si può supporre che si abbia avuto una situazione simile anche negli altri Stati che abbiano adottato delle contromisure: per esempio, in Italia, il tasso di incidenza dell'influenza è stato il più basso degli ultimi 20 anni [5]. Quindi è importante continuare a mantenere l'attenzione alta durante il periodo dell'influenza stagionale, conservando delle buone abitudini di igiene e utilizzando, quando è necessario, la mascherina.

REFERENCES

- [1] "Influenza" [Online]. Available: <https://it.wikipedia.org/wiki/Influenza> [Accessed: 3 - Jul - 2022].
- [2] "FluNet" [Online]. Available: <https://www.who.int/tools/flunet> [Accessed: 3 - Jul - 2022].
- [3] "Dataset USA from 2017 to 2022" [Online]. Available: <https://apps.who.int/flumart/Default?ReportNo=12> [Accessed: 3 - Jul - 2022].
- [4] "How Are Influenza A and B Different?" [Online]. Available: <https://www.healthline.com/health/cold-flu/influenza-a-vs-b> [Accessed: 4 - Jul - 2022].
- [5] "L'incidenza dell'influenza stagionale è la più bassa degli ultimi 20 anni" [Online]. Available: <https://www.fondoasim.it/incidenza-influenza-stagionale/> [Accessed: 4 - Jul - 2022].