

Un mese di Covid-19 in Italia: una guida alla lettura dei dati per bloccare la disinformazione.

Mariangela Sciandra, Alice Blandino

I. INTRODUZIONE

A un mese dal primo paziente risultato positivo a Codogno e dal primo deceduto a Padova, siamo nel bel mezzo di un sovraccarico di informazioni sul nuovo Coronavirus, con aggiornamenti continui dei casi e flussi infiniti di informazioni. E se tante sono le informazioni che ci arrivano, altrettante sono le domande nuove che ogni giorno ci poniamo. Tra queste, quella che riempie le nostre lunghe giornate in cui vediamo la nostra libertà sempre più decurtata da normative sempre più stringenti, e relativa alla temporaneità di queste soluzioni: Quanto durerà il “temporaneo”? Ci sono stati proposti i più svariati grafici, alcuni semplici altri più complessi; abbiamo sentito parlare di andamento esponenziale, logaritmico, di picco e talvolta anche di fattore R_0 . Abbiamo associato ad ogni notizia speranza e paure, cercando di discernere il vero dal falso. Ma la domanda è sempre la stessa: Quando torneremo alla normalità? Quando finirà lo stato di emergenza? Per rispondere a questa domanda seguiamo un percorso semplice che cercherà di chiarire cosa sia successo in questo lungo mese.

La prima cosa da fare quando si vuole studiare un fenomeno e si vuole provare a capire cosa succederà, in che termini e in quanto tempo, è **osservare** il fenomeno stesso. E allora ci chiediamo: Cosa è successo in Italia a partire dal 25 febbraio?



Il 24 febbraio 2020 l'Italia si affacciava ad un nuovo scenario inaspettato: i primi contagi da Covid-19, ad oggi ben noto come Coronavirus o Sars-cov-2. Il virus, inizialmente sottovalutato e visto da molti come “una semplice influenza”,

ha portato ad un numero di contagi spaventosamente alto e ad un tasso di letalità che ad oggi si attesta circa all'8.3%. Le regioni che per prime hanno fatto registrare contagi sono state *Lombardia*, *Veneto* ed *Emilia-Romagna*. Dopo un mese, come mostra la mappa, tutte le regioni d'Italia sono ormai interessate dal fenomeno Coronavirus, e *Lombardia*, *Veneto* ed *Emilia-Romagna* si piazzano ancora ai primi posti nella classifica delle regioni con il maggiore numero di contagi.

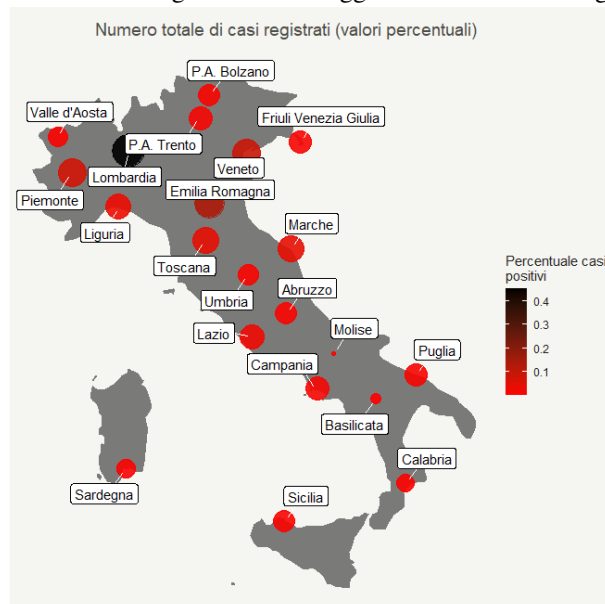


Fig. 1: Numero totale di casi di Covid-19 registrati nelle regioni italiane al 23 marzo 2020.

Andando a guardare nel dettaglio il numero totale dei casi registrati nelle singole regioni (Fig.2), le regioni dell'area Nord sono quelle che presentano il valore più elevato di casi di Coronavirus, e in particolare la *Lombardia* risulta essere la regione più colpita con un totale di 28.761 casi registrati.

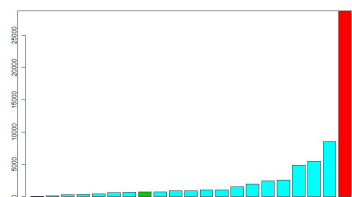


Fig. 2: Numero totale di casi di Covid-19 registrati al 23 marzo 2020 per regione.

La Sicilia (in verde) sembra essere molto lontana dalla situazione lombarda. Ma è davvero così estrema la situazione della Lombardia rispetto a quella delle altre regioni italiane? Quello lombardo può essere considerato un ceppo diverso di Covid-19? Non potrebbe essere semplicemente una distorsione informativa dovuta al fatto che i tamponi effettuati per diagnosticare la presenza del virus non sono eseguiti con la stessa frequenza sul territorio nazionale?

II. IL NUMERO DI TAMPONI ESEGUITI E L'EFFETTO SUI RISULTATI RAPPRESENTATI.

Sui tamponi le Regioni stanno procedendo in ordine sparso attuando un regolamento fai-da-te: forti dell'appello del direttore generale dell'Oms, che invita ad eseguire un numero sempre più alto di test e non essendo arrivata una raccomandazione a effettuare screening di massa, i governatori si stanno muovendo autonomamente in merito alla scelta del numero e delle persone da sottoporre ai tamponi. Si va dal Veneto di Luca Zaia che annuncia "test on the road" (la somministrazione di test a campione per cercare di controllare il più possibile i 5 milioni di veneti) che ne ha già eseguiti 61.115, al Piemonte di Alberto Cirio, che ne ha effettuati appena 13.560. La Regione più colpita d'Italia, la Lombardia, con 28.761 casi totali su 63.927 complessivi e 3.776 vittime su 6.077 nazionali, è quella che appare più in difficoltà anche sul fronte della rilevazione dei contagi (dall'inizio dell'epidemia sono stati eseguiti 73.242 tamponi). Ormai, anche per questioni organizzative, i tamponi vengono effettuati solo sui sintomatici gravi, che richiedono il ricovero in ospedale, e questo, non solo fa impennare la percentuale di letalità, ma impedisce anche di mettere in isolamento gli infetti con sintomi lievi e di conoscere l'effettiva platea degli asintomatici. E la Sicilia? La Sicilia di Nello Musumeci, nella riunione del 17 marzo dei 14 esperti scelti dal governatore, ha deciso di effettuare tamponi a tutti i sintomatici sospetti e a chi è rientrato nelle scorse settimane dalle zone rosse del Nord (si sono registrati in 35mila), estendendo i controlli anche ai loro familiari. E' ancora al vaglio degli esperti, invece, la possibilità di sottoporre al tampone tutti sanitari impegnati nei reparti Covid. In Fig. 3 la distribuzione delle regioni per numero di tamponi somministrati.

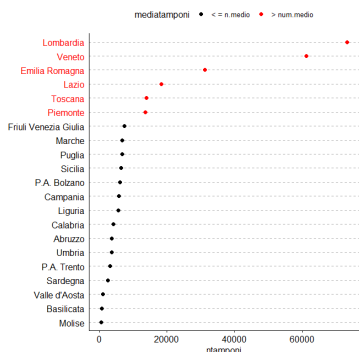


Fig. 3: Numero totale di tamponi eseguiti per regione al 23 marzo 2020.

Alla luce delle considerazioni fatte e ai valori riportati in Fig.3 è chiaro che il numero totale di casi registrati dipenderà dal numero di tamponi eseguiti e dunque, per un corretto confronto tra le regioni, al fine di non creare allarmismi infondati circa un possibile ceppo lombardo, o eccessiva tranquillità nelle regioni che si ritengono lontane dalla situazione lombarda, in Fig.4 si riporta una versione corretta del grafico in Fig.2 in cui il numero totale di casi registrati in ciascuna regione è ponderato per il numero totale di tamponi eseguiti nella medesima regione.

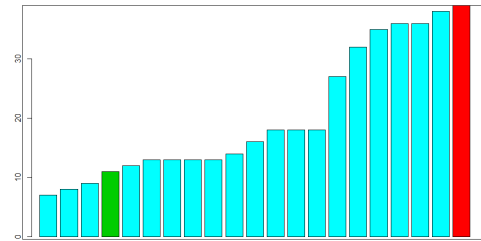


Fig. 4: Numero totale di casi di Covid-19 registrati al 23 marzo 2020 per regione corretto per il numero di tamponi eseguiti.

Come si evince dal confronto tra la Fig 4 e la Fig. 2, correggere per il numero di tamponi effettuati attenua la differenza tra la Lombardia e le altre regioni. Riportando la stessa informazione in un *radarplot* (Fig.5), è ancora più evidente come, relativizzando per il numero di tamponi eseguiti, la distanza tra le percentuali di positivi nelle varie regioni e la Lombardia è meno netta di quella in Fig.2. Va comunque detto, che le analisi qui condotte non tengono in considerazione alcuni problemi legati alla qualità del dato sui tamponi quali: il lag temporale tra il momento in cui si effettua il tampone e l'esito dello stesso e la possibile ripetizione del test su soggetti sintomatici risultati negativi ad una prima rilevazione.

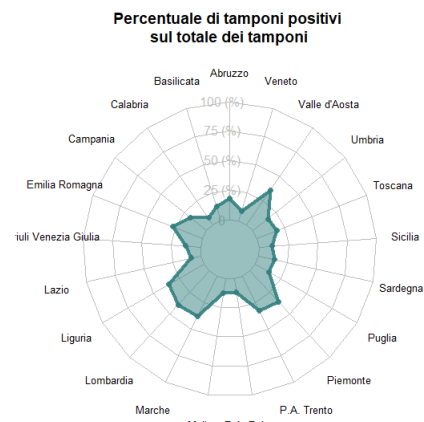


Fig. 5: Radarplot della percentuale di casi positivi sul numero totale di tamponi eseguiti per ciascuna regione al 23 marzo 2020.

Dalla Fig.5 emerge come le regioni in cui si registra la percentuale di tamponi positivi più alta siano *Lombardia* (0.39), *Piemonte* (0.36) e *Marche* (0.38).

III. COME SI DEFINISCE IL TASSO DI LETALITÀ DA CORONAVIRUS?

Iniziamo col dire che spesso i termini “tasso di mortalità” e “tasso di letalità” vengono usati erroneamente come sinonimi. In realtà i due indici danno informazioni diverse. In particolare, il *tasso di letalità* o CFR (*Case Fatality Rate*), indica la proporzione, tipicamente percentuale, di decessi per una determinata malattia sul totale dei soggetti ammalati in un determinato arco temporale. È un fattore di rischio, indicando la probabilità di morire se si è affetti da una specifica malattia e può quindi essere considerato un indice della sua gravità. Tale valore risulta estremamente variabile in relazione alla durata dell’osservazione, per questo va contestualizzato in rapporto a un intervallo di tempo. Il *tasso di mortalità* invece è dato dal rapporto tra il numero delle morti in una comunità o in una popolazione durante un periodo di tempo e la quantità della popolazione media dello stesso periodo (generalmente si fa riferimento ad un anno di calendario). Quello di cui sentiamo parlare in questi giorni è dunque il tasso di letalità del Coronavirus che, per come è definito, dipende (al denominatore) dal numero dei casi identificati e che quindi sarà anche funzione del numero di tamponi effettuati (regioni con un numero maggiore di tamponi effettuati è probabile che facciano rilevare un numero maggiore di casi e quindi, aumentando il denominatore, a parità di numero di decessi, il tasso risulterà più basso). Il report dell’Istituto Superiore della Sanità del 13 marzo 2020 descrive le caratteristiche di 1016 pazienti deceduti e positivi al Coronavirus in Italia. Dal report si apprende che l’età media dei decessi è di 79.4 anni e che nel 71.5% dei casi si tratta di uomini. Sono solo due i pazienti deceduti sotto i 40 anni. Ma cosa vuole dire morire di Coronavirus? Come spiega lo studio dell’Oms (Report of the Who-China Joint Mission on Coronavirus Disease 2019, 28 febbraio 2020) i pazienti che non hanno altre malattie a parte il coronavirus hanno tassi di letalità dell’1.4% e quindi nettamente più bassi della media nazionale (4%). Quelli che invece hanno condizioni sanitarie compromesse da altre malattie oltre al coronavirus hanno percentuali più alte: il 13.2% tra chi ha malattie cardiovascolari, il 9.2 tra i diabetici, l’8.4% tra chi soffre di ipertensione, l’8% tra chi soffre di malattie respiratorie croniche, il 7.6% tra chi è malato di cancro. E se ci condizioniamo, ovvero, se teniamo conto anche del numero di patologie persistenti, è chiaro che si può arrivare a valori nettamente diversi. E in termini di confronto tra regioni? I tassi di letalità più alti, ad oggi, sono stati registrati in Lombardia. Qui la percentuale di pazienti morti tra coloro che sono risultati positivi è pari al 13% (contro il 6% del 25 febbraio) e quindi nettamente superiore alla media nazionale. La tabella accanto e la Fig. 6 mostrano i tassi di letalità per ciascuna regione, rispettivamente al 25 febbraio (Tabella 1) e al 23 marzo 2020 (Fig.6). Va sottolineato che in Tabella 1 non sono riportati i tassi di letalità di alcune

regioni poichè al 25 febbraio non tutte le regioni avevano registrato morti da Corona virus.

	Regione	Tassi di letalità
1	Lombardia	0.06
2	Veneto	0.05
3	Emilia Romagna	0.00
4	Lazio	0.00
5	Liguria	0.00
.	.	.
.	.	.
.	.	.
21	Valle d’Aosta	0.00

TABLE I: Tassi di letalità per ciascuna regione al 25 febbraio

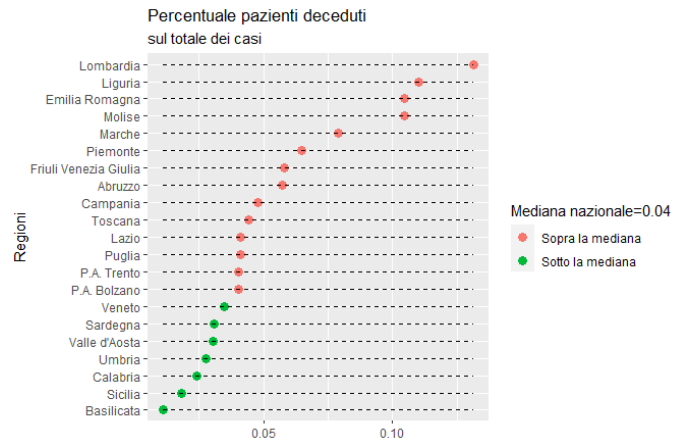


Fig. 6: Tassi di letalità per ciascuna regione al 23 marzo 2020.

IV. C’È UN CEPPLO PIÙ LETALE?

Si legge in molte testate giornalistiche che in Italia si muore molto di più che in altri Paesi per Covid-19. Confrontando i dati nazionali al 23 marzo, nel nostro Paese sono morte 6.077 persone a fronte di 63.927 contagiati: ossia il 9.5%. Una percentuale molto alta soprattutto se confrontata al 4% cinese (3277 deceduti su 81171 contagiati), al 6.5% spagnolo (2.316 deceduti su 35.212 contagiati) e al 4.3% francese (860 morti su 19.856 malati). Ma perchè in Italia si muore di più? Sfatando questa affermazione allarmistica, le ragioni di questo scostamento possono essere molteplici e di varia natura. Tra questi ci possono essere: il metro utilizzato per il conteggio dei deceduti, l’età media della popolazione italiana, il ruolo centrale giocato dagli anziani nella nostra società (e quindi la loro maggiore esposizione al contagio) e una tendenza più accentuata al contatto fisico. Ma vediamo come si è evoluto il tasso di letalità nazionale dal 25 febbraio ad oggi (Fig.7).

Si noti come il tasso di letalità è aumentato considerevolmente nell’arco di un mese passando dal 3% del 24 febbraio al 9.5% del 23 marzo. La Fig. 8 confronta le curve del numero dei decessi dal 25 febbraio a oggi della Lombardia, Veneto, Emilia Romagna e Sicilia.

Va inoltre fatta un’importante considerazione legata all’ospedalizzazione.

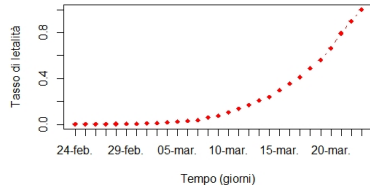


Fig. 7: Tassi di letalità nazionali giornalieri.

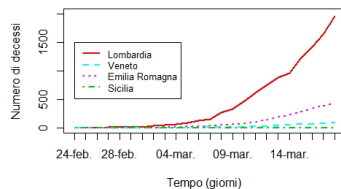


Fig. 8: Confronto tra le curve dei decessi in Sicilia e in altre tre regioni tra le più colpite.

Le regioni con il più alto numero di morti risultano infatti essere tra quelle che presentano una percentuale di pazienti ospedalizzati più alta, come mostra il grafico sottostante.

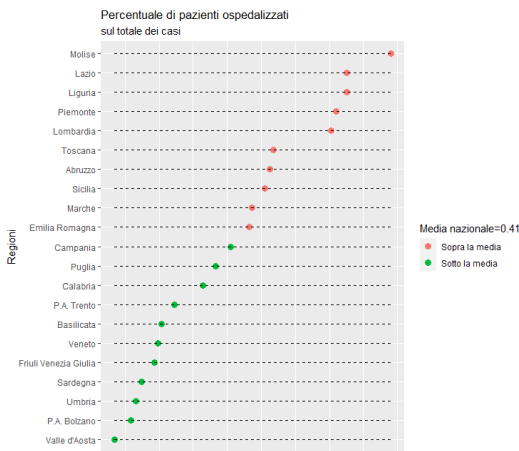


Fig. 9: Tasso di ospedalizzazione nelle varie regioni al 23 marzo.

Diversi studi in letteratura hanno posto l'attenzione sull'insorgenza di infezioni letali derivanti dall'ospedalizzazione dei pazienti. Pertanto, un'ulteriore ipotesi di ricerca sull'elevata letalità lombarda, potrebbe far riferimento alla gestione dei pazienti (sulle politiche decisionali riguardanti l'eventuale ricovero) che presentano sintomatologie da Covid-19.

V. COME SI LEGGONO LE CURVE DI CRESCITA? COSA SI INTENDE PER PICCO?

Quando arriva il picco? O detto in altro modo, quanto ancora dovrò vivere in questa situazione surreale di sacrificio, di ansia, di dubbio? Siamo tutti in attesa del picco. Ma cos'è il picco? Andiamo per punti e cerchiamo di capire anche questo aspetto. Quasi tutti sappiamo cosa significhi "crescita

esponenziale", ma sappiamo fino in fondo interpretarne il significato? Come stabilire l'eventuale punto di arresto dei contagi?

Un andamento esponenziale è caratterizzato dal fatto che il numero di persone infette non aumenta in maniera costante, ma la sua velocità di diffusione cresce sempre di più in modo moltiplicativo. I virus sono un esempio perfetto di crescita esponenziale dal momento che **ciò che causa i nuovi casi sono i casi già esistenti**. Pertanto se indichiamo con N_t il numero di casi osservati in un dato giorno t e con E il numero medio di persone "Esposte" ad un possibile contatto con ciascuno degli N_t individui infetti e p la probabilità per ciascuno di questi individui di diventare infetto, l'incremento giornaliero del numero di casi sarà pari a

$$\Delta N_t = E * p * N_t$$

e di conseguenza i casi al giorno $t + 1$ saranno pari a

$$N_{t+1} = N_t + E * p * N_t = (1 + E * p) * N_t.$$

In parole più semplici, ogni giorno il numero di casi è moltiplicato per una costante maggiore di 1. Pertanto, se questo processo, per esempio, parte con N_1 soggetti infetti e prosegue per 10 giorni, all'undicesimo giorno si avrà

$$N_{11} = (1 + E * p)^{10} * N_1.$$

Come possiamo arrestare questo andamento? O meglio, come possiamo che la diffusione della pandemia rallenti in modo da non stressare oltremisura il sistema sanitario? L'unica arma in nostro possesso è intervenire sui valori assunti da E e p e fare in modo che questi decrescano nel più breve arco di tempo possibile. Ecco il fondamento alla base dei decreti ministeriali e delle relative misure di contenimento sui nostri spostamenti sul territorio nazionale. Da questo momento in poi abbiamo dunque assunto la consapevolezza su come il nostro comportamento, responsabile o meno, possa influire sulla riduzione del fattore $(1 + E * p)$ e quindi sul rallentamento della diffusione esponenziale del Covid-19. E se la crescita non fosse esponenziale? Quando la curva assume una forma a "s"? Se, a partire dal modello esponenziale, correggiamo la probabilità p che un individuo esposto diventi un infetto tenendo conto del numero di individui nella popolazione che sono già infetti (e quindi con una probabilità nulla di essere infettati nuovamente), allora

$$p = 1 - \frac{N_t}{\text{Num. popolazione}}$$

e la curva di crescita seguirà una curva logistica. Non è facile distinguere un accrescimento esponenziale da uno logistico nel primo tratto di crescita (Fig. 10). Da qui l'incertezza da parte di coloro i quali, in questo mese, hanno condotto le analisi per cercare di capire per tempo il tipo di diffusione del Covid-19 e le misure da prendere per limitarne la diffusione. Il punto in cui le due curve cominciano a differenziarsi è il famoso "picco", o meglio noto come *inflection point*, che identifica il punto in cui la curva passa da una curvatura positiva (verso l'alto) ad una negativa (verso il basso).

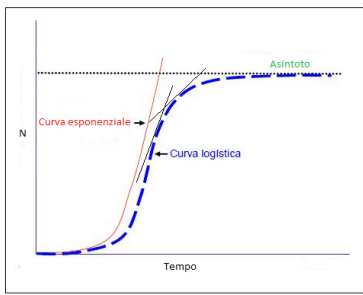


Fig. 10: Andamento di una funzione esponenziale.

Un numero importante da tenere sotto controllo durante un'epidemia è il "fattore di crescita" (*Growth rate (GR)*) definito come il rapporto tra due incrementi successivi, ovvero

$$GR = \frac{\Delta_t}{\Delta_{t-1}}$$

Facciamo un esempio numerico per capire il calcolo di questo fattore: prendiamo gli ultimi tre valori cumulati di casi rilevati in tutta Italia (21 marzo=42.681, 22 marzo=46.638, 23 marzo=50.418), il GR sarà uguale a:

$$GR = \frac{50.418 - 46.638}{46.638 - 42.681} = 0.96$$

Nella fase esponenziale di crescita questo rapporto assumerà sempre valori maggiori di 1; quando il fattore di crescita comincerà ad assumere valori vicini ad 1, vorrà dire che siamo vicini al punto di flesso ("picco"), ovvero si assisterà ad un rallentamento della diffusione della malattia. La curva esponenziale è molto sensibile alle variazioni della costante E^*p : ciò significa che, piccole variazioni nella probabilità di contagio o nel numero medio di persone che possono entrare a contatto con una persona già infetta, possono produrre variazioni consistenti del tasso di crescita. Pertanto, quanto dobbiamo aspettare per raggiungere il "picco epidemico" dipende solo da noi. **Più siamo sensibili al problema, molto più veloce sarà il ritorno alla normalità.** Per approfondimenti in merito a metodologie alternative usate per modellare i dati di diffusione del Coronavirus si consiglia di consultare il gruppo di lavoro StatGroup-19 alla pagina <https://statgroup-19.blogspot.com/>, o il gruppo di lavoro del Dipartimento di Scienze, Economiche, Aziendali e Statistiche dell'Università degli Studi di Palermo al sito <http://unipa.it/covid19>.

VI. QUAL È LA SITUAZIONE ITALIANA AL 23 MARZO?

Cerchiamo di capire adesso qual è la situazione dell'Italia utilizzando i dati forniti dalla Protezione Civile. Nell'ultima giornata di osservazione i casi totali sono balzati a 63.927, con un incremento di ben 4789 nuovi casi registrati in un solo giorno (inferiore però al dato record del 21 marzo che si assestava a 6.557). I casi attualmente positivi sono 46.638, i guariti salgono a 7.432, con un incremento di 952 nell'ultimo giorno di osservazione. Il totale dei decessi è pari a 6077. I pazienti attualmente positivi sono tenuti sotto osservazione secondo la ripartizione in Tab.2: i ricoverati con

sintomi sono in tutto 20.692; in terapia intensiva sono 3.204; in isolamento domiciliare 26.522. I tamponi eseguiti ad oggi sono 275.468. Cosa possiamo dire in merito al fattore di crescita? A livello nazionale i valori del fattore di crescita sono quelli riportati in Fig. 11. Sembra che a livello nazionale ci si stia avvicinando al "picco" e che quindi le misure di contenimento stiano avendo degli effetti tangibili.

Ciò è deducibile anche dalla figura 13 in cui si evidenzia, a livello nazionale, un decremento nel numero dei nuovi casi giornalieri a partire dal 21 marzo.

Nuovi attualmente positivi	Numero di casi
Isolamento domiciliare	26522
Ricoverati con sintomi	20692
Terapia intensiva	3204

TABLE II: Ripartizione dei casi positivi al 23 marzo 2020

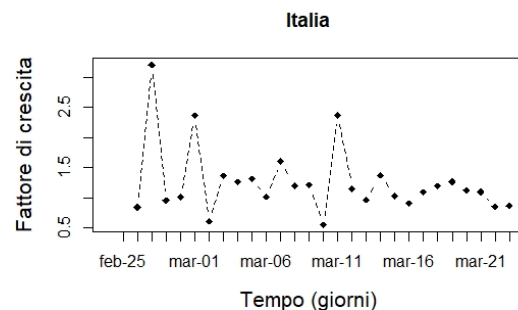


Fig. 11: Andamento del fattore di crescita esponenziale in Italia al 23 marzo 2020.

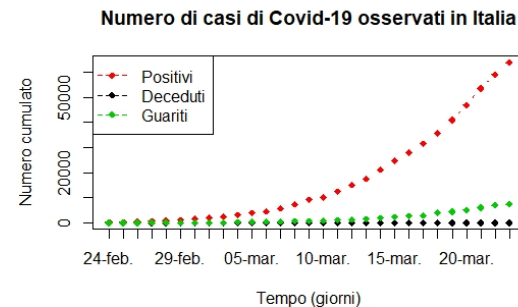


Fig. 12: Numero di casi COVID-19 osservati in Italia al 23 marzo 2020.

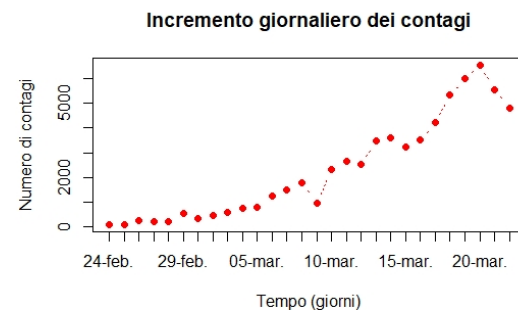


Fig. 13: Incremento giornaliero dei contagi in Italia al 23 marzo 2020.

Occorre a questo punto ricordare che il "picco" non corrisponde ad un solo giorno; il picco è un periodo di stasi

in cui si assiste ad un rallentamento della diffusione, che in genere precede una fase discendente del contagio.

Dal grafico in Fig. 14 emerge uno scostamento della curva esponenziale dalla curva osservata a partire dal 13 marzo, come ad indicare un rallentamento nella diffusione del virus. La crescita è ancora esponenziale ma il fattore di crescita sembra stia iniziando a ridursi.

E' il caso di ribadire che i risultati fin qui osservati non sono usati a fini previsivi poichè il margine di errore legato alla scarsa conoscenza del fenomeno, alla diffusione non omogenea dei contagi a livello nazionale e il possibile impatto, non ancora osservabile, circa l'ondata migratoria di chi è rientrato nelle scorse settimane dalle zone rosse del Nord potrebbe condurre a risultati affetti da ampia distorsione.

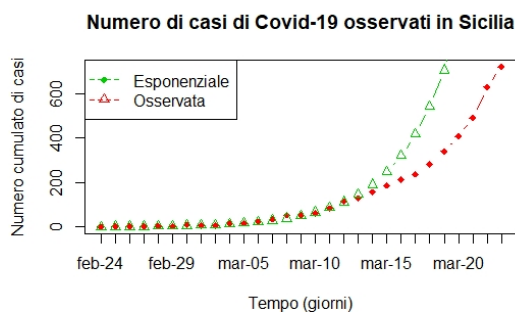


Fig. 14: Numero di casi COVID-19 osservati in Sicilia al 23 marzo 2020.

VII. CONFRONTO DELLE CURVE REGIONALI DI CRESCITA EPIDEMICA

E se volessimo confrontare la curva della Lombardia con quella della Sicilia? Una rappresentazione delle due curve sulla stessa scala di misura non sarebbe informativa dal momento che le due regioni si trovano ad uno stadio di avanzamento dell'epidemia diverso. Da una rappresentazione congiunta delle due, l'unica cosa che ne potremmo derivare è il lag temporale nella diffusione del virus tra noi e i lombardi.

Una rappresentazione alternativa a quella in Fig. 15, in cui le due curve sono rappresentate distintamente, potrebbe consistere nel riallineare le curve rispetto ad un tempo t_0 (che probabilmente non rappresenterà lo stesso giorno di calendario per tutte le regioni) fissando un criterio statistico. Tale rappresentazione rende confrontabile l'andamento della diffusione del virus a parità di stato di avanzamento. Di seguito si è deciso di scegliere t_0 come il momento in cui ciascuna regione ha superato i 100 casi. In Fig.16 si riportano in nero le curve riallineate di tutte le regioni che hanno superato i 100 casi mettendo in evidenza la Sicilia e tre delle regioni tra le più colpite (Lombardia, Emilia Romagna e Veneto). Questa operazione di riallineamento farà sì che adesso sull'asse delle ascisse non si avranno più i giorni di calendario, ma il tempo espresso in "giorni trascorsi dal momento in cui si sono raggiunti i 100 casi".

Dal grafico si evince il ritardo di molte regioni, tra cui la Sicilia, nel processo di espansione epidemica; la Sicilia

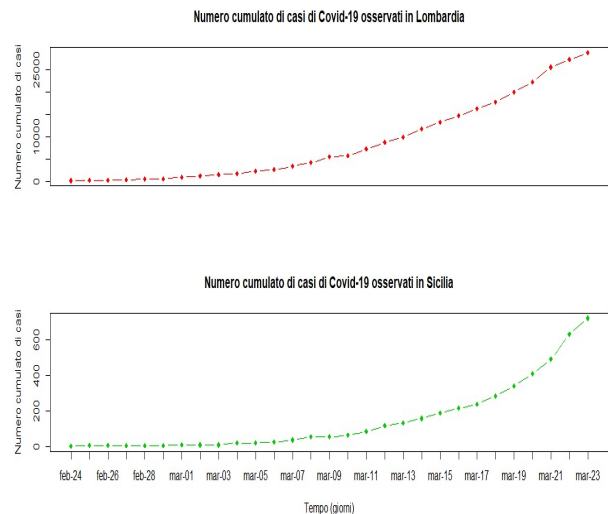


Fig. 15: Numero di casi cumulati per Lombardia e Sicilia al 23 marzo 2020.

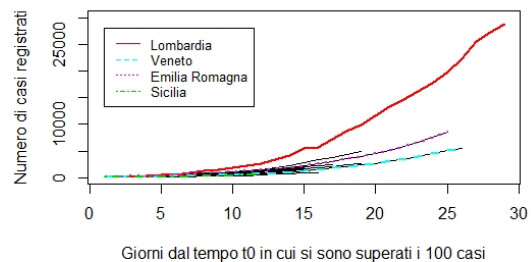


Fig. 16: Incremento giornaliero dei contagi in Italia al 23 marzo 2020.

ha, infatti superato i 100 casi solo da 12 giorni, mostrando pertanto un ritardo rispetto alla Lombardia di circa 17 giorni.

Da questo grafico si evince come l'impatto delle misure di contenimento epidemico dettate dai decreti sia diverso tra le regioni in base al loro stadio di avanzamento epidemico.

Se ci soffermiamo a osservare la curva di crescita giornaliera dei contagi in Sicilia (Fig. 17) a partire dal tempo t_0 in cui si sono registrati un numero di casi maggiore di 100 (giorno 12 marzo, numero di casi 115), possiamo notare come la curva dei contagi osservati (curva verde) sia lontana sia dalla curva che prevede il raddoppio giornaliero dei contagi (curva rossa) che da quella che prevede un incremento giornaliero del 50% dei casi (curva blu). Un risultato importante è invece la quasi perfetta sovrapposizione della curva osservata con la curva esponenziale in cui la costante λ ($\lambda=1+E*p$) è uguale a 1.18 (curva viola), che equivale ad avere un incremento giornaliero di casi pari al 18%.

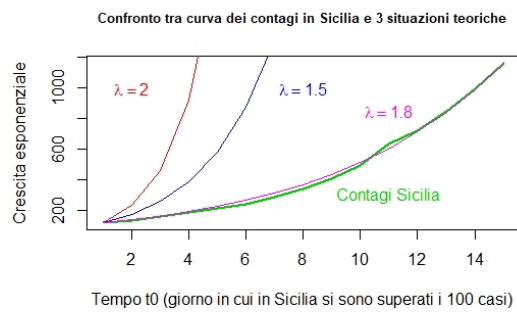


Fig. 17: Confronto della curva dei contagi in Sicilia con tre curve esponenziali con tre diversi valori di λ