

Emergenza COVID - Situazione Italia[†]

Aggiornato al 17 Aprile 2020

Francesco Furno[†]

Contributi Principali

1. Serie storiche regionali e nazionali aggiornate quotidianamente che sono purificate dalla volatilità giornaliera e quindi più facilmente interpretabili.
2. Estrapolazioni statistiche dei trend attuali con diverse metodologie.
3. Stime del tasso di letalità apparente che tengono conto dell'intervallo di tempo tra diagnosi e decesso.
4. Analisi dell'eterogeneità dei tassi di letalità tra regioni.
5. Stime dei casi effettivi e della percentuale di popolazione infettata sia a livello nazionale sia regionale.
6. Previsioni dei decessi (totali e giornalieri) nei prossimi 5 giorni.

Contents

1 Fotografia: Evoluzione dell'Epidemia	2
1.1 Casi, Deceduti, Ricoveri	2
1.2 Tamponi	3
1.3 Pressione su Terapie Intensive	5
2 Stime	6
2.1 Stima del Tasso di Letalità	6
2.2 Stima del numero di giorni tra diagnosi e decesso	7
2.3 Stima del numero di giorni tra diagnosi e dimissione/guarigione	10
2.4 Letalità e Tamponi: Confronto tra Regioni	11
2.5 Stima Casi Effettivi	12
2.6 Previsione dei Decessi	16
2.6.1 Prossimi 5 Giorni	16
2.6.2 Errori Previsionali Percentuali	17

[†]Lo scopo della presente analisi è di fornire una fotografia dello stato *attuale* dell'epidemia COVID in Italia. Non è basata su modelli epidemiologici matematici, ma sull'analisi statistica (per lo piu' descrittiva) dei dati regionali resi disponibili dal Ministero.

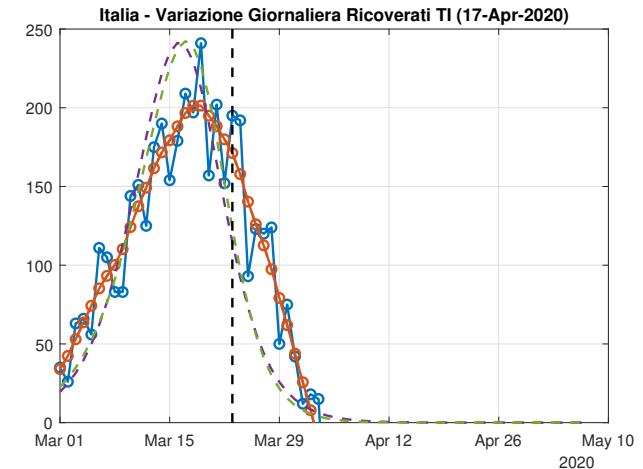
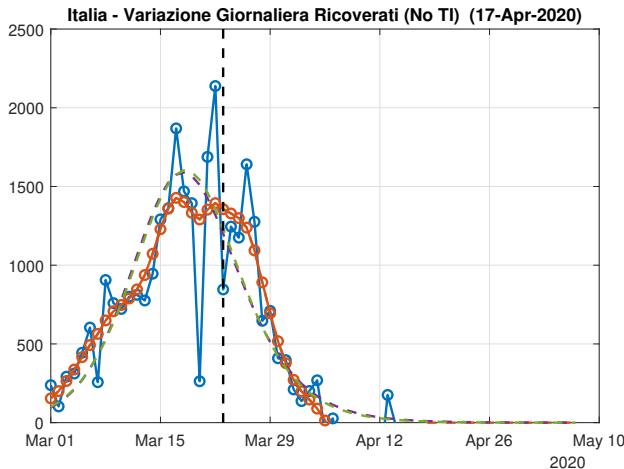
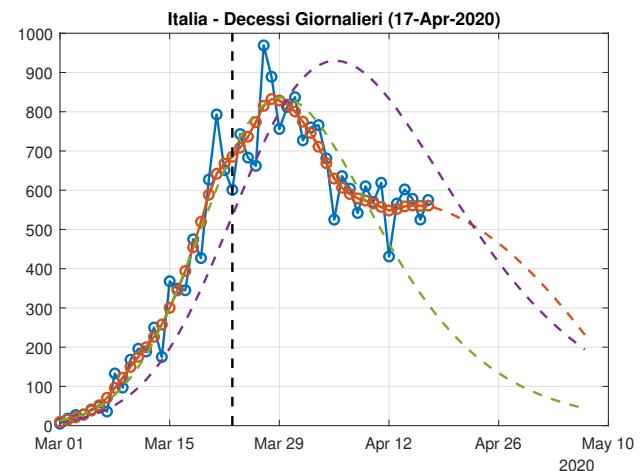
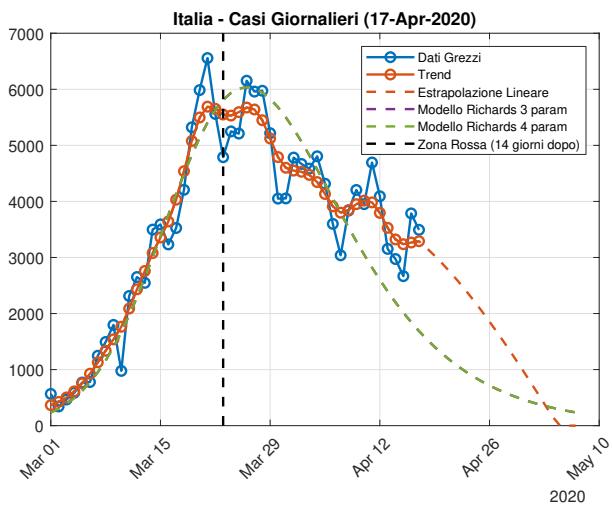
[†]Dipartimento di Economia, New York University. I dati verranno aggiornati giorno per giorno. Grazie a Stefano Olivari per numerosi suggerimenti. Vi prego di scusare eventuali errori e di segnarli a francesco.furno@nyu.edu. Fonte dati: <https://github.com/pcm-dpc/COVID-19/blob/master/dati-regioni/dpc-covid19-ita-regioni.csv>

1 Fotografia: Evoluzione dell'Epidemia

1.1 Casi, Deceduti, Ricoveri

I dati giornalieri sono volatili: utilizzo quindi una tecnica statistica per stimare il trend.¹ Si noti che è fondamentale filtrare la volatilità giornaliera: un filtro rende un dato giornaliero particolarmente negativo meno negativo, e un dato giornaliero particolarmente positivo meno positivo.

Per rendere i numeri più facilmente interpretabili, è utile guardare i dati giornalieri e alla loro extrapolazione statistica effettuata con diversi modelli:²

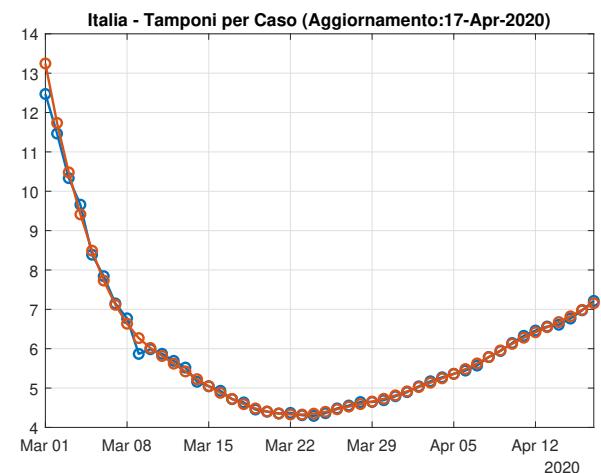
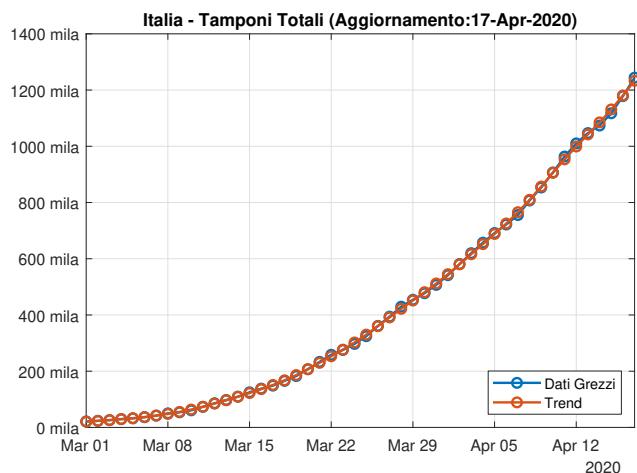
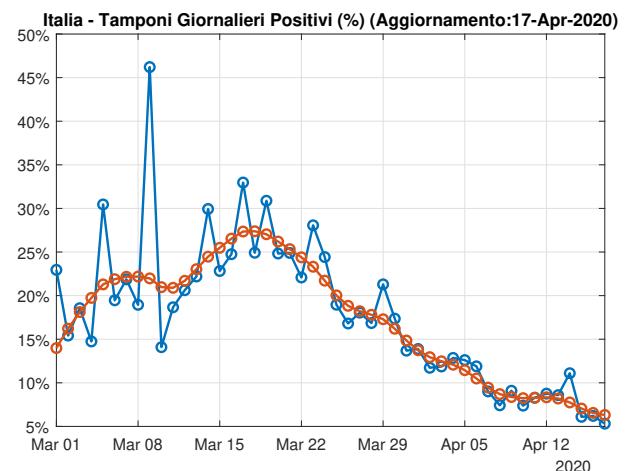
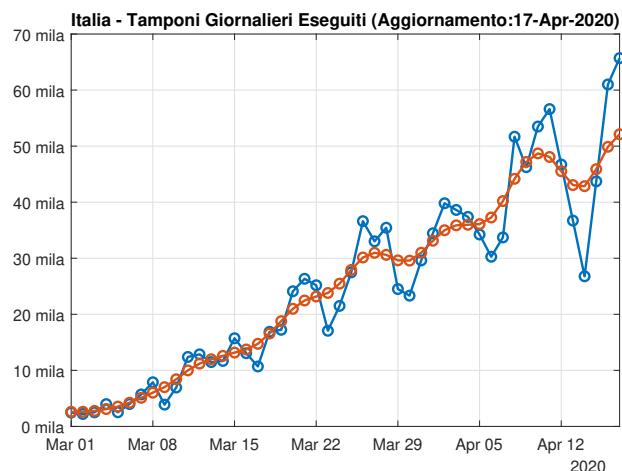


¹Il trend è estratto con un filtro Hodrick–Prescott (HP), parametro di smoothing uguale a 5. Il filtro HP è un filtro comunemente usato in macroeconomia. Lo utilizzo semplicemente per eliminare la volatilità giornaliera dei dati. Si noti che la stima del trend viene aggiornata ogni volta che nuovi dati vengono resi disponibili. In alternativa, è possibile utilizzare filtri più semplici, come una media-mobile.

²La linea tratteggiata arancione è la proiezione lineare del trend; la linea viola rappresenta la proiezione con il modello generalizzato di Richards a 3 parametri; la linea verde rappresenta la proiezione con il modello generalizzato di Richards a 4 parametri. Si veda il paper epidemiologico: <https://doi.org/10.1016/j.idm.2017.08.001>.

Si noti che queste non sono previsioni certe: sono un modo di capire in quale punto della curva ci troviamo attualmente. A 14 giorni dall'istituzione della zona rossa nazionale si vede un appiattimento della curva dei contagi. Dato che i decessi seguono in media 5 giorni dalla rilevazione del caso, i decessi continuano ad aumentare anche dopo che i nuovi casi rallentano.³ Il rallentamento della crescita della terapia intensiva è probabile che sia dovuto a saturazione, anche se si osserva lo stesso comportamento in regioni dove la terapia intensiva non è satura.

1.2 Tamponi

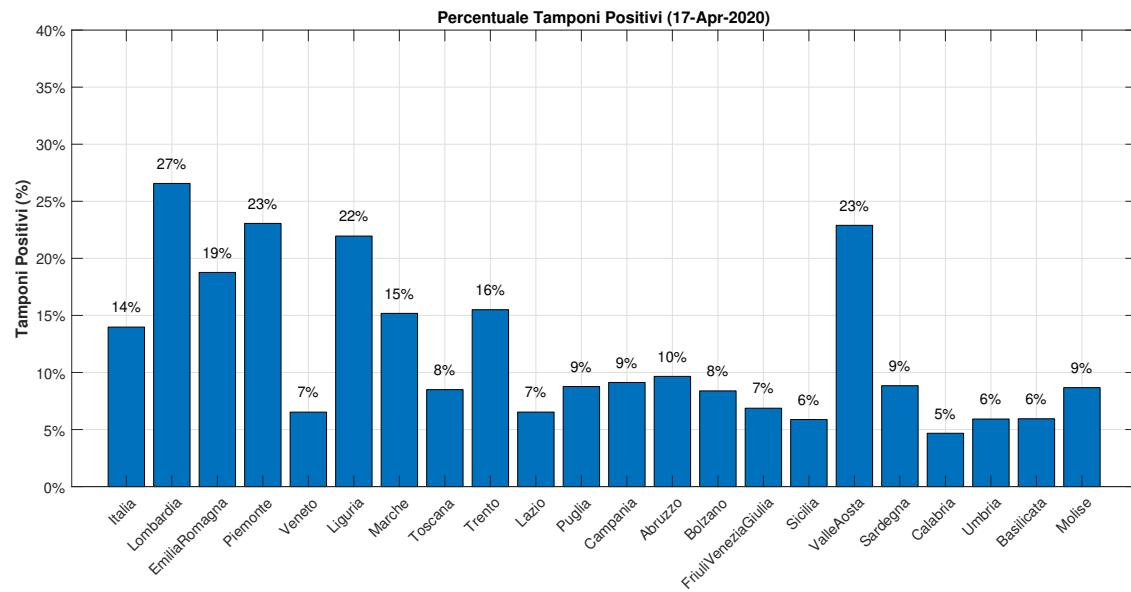


I tamponi totali eseguiti sono in costante crescita.⁴ Diversa è l'evoluzione del numero medio di tamponi effettuati per ogni caso positivo, in stabilizzazione dopo una continua discesa: segnale che la capacita' di testare del sistema sanitario nazionale è al limite e di un cambiamento di strategia.

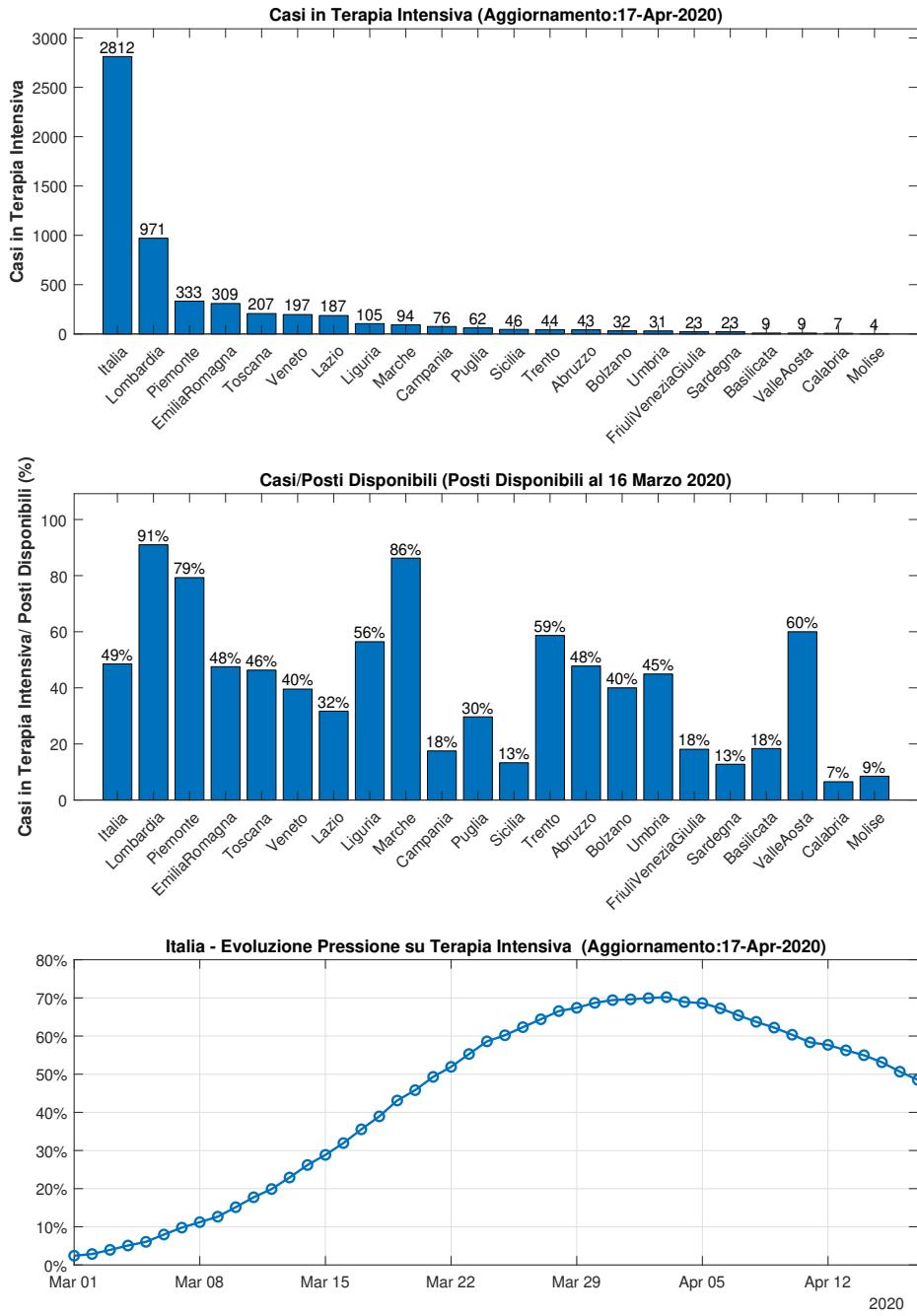
³Si noti che 5 giorni non è l'intervallo temporale tra sintomi e decesso, ma tra rilevazione da parte del sistema sanitario e decesso. Si veda subsection 2.2.

⁴Si noti che i dati ufficiali contengono una misurazione dei test (o tamponi) totali eseguiti. Sembrerebbe che una stessa persona possa essere testata piu' volte.

E' naturale che questo risulti in una *sottostima* del numero effettivo dei positivi. Alta la percentuale di tamponi positivi.



1.3 Pressione su Terapie Intensive



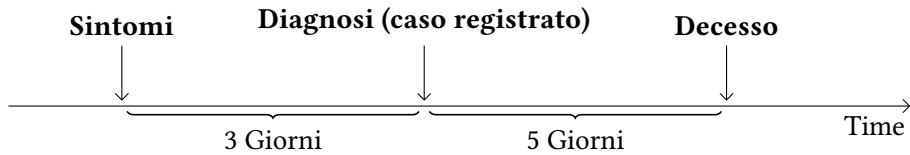
Il pannello superiore della figura mostra il numero di pazienti in terapia intensiva, mentre il pannello centrale divide questo numero per i posti totali di terapia intensiva al 16 marzo, regione per regione. Infine, il pannello inferiore mostra l'evoluzione nel tempo della pressione sulla terapia intensiva nell'aggregato nazionale.⁵

⁵Si noti che il numero di posti totali di terapia intensiva è in costante espansione, e questo spiega come sia possibile avere percentuali superiori al 100% nel pannello inferiore.

2 Stime

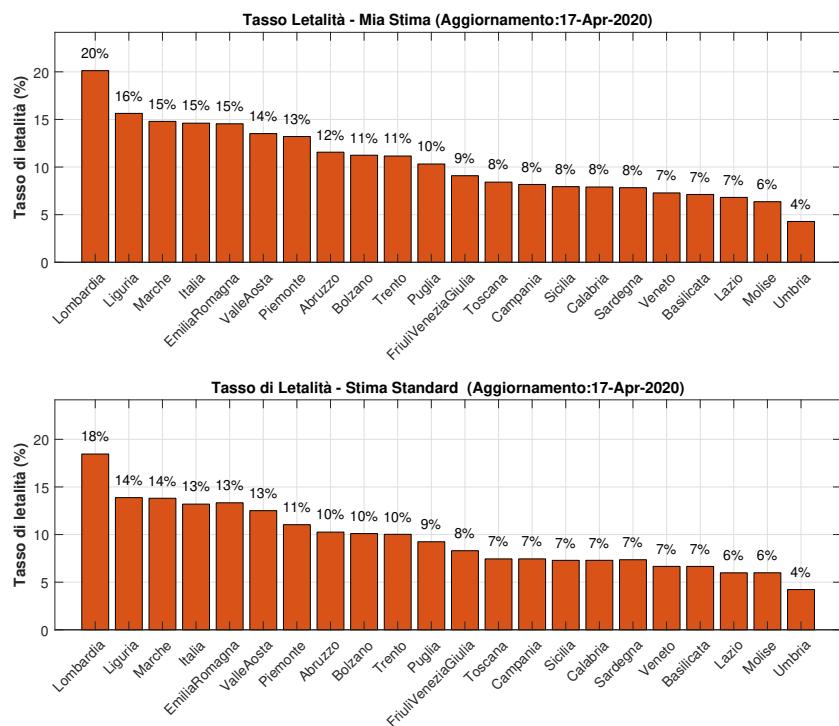
2.1 Stima del Tasso di Letalità

La mia stima del tasso di letalità è basata sull'idea che, in caso di decesso, intercorrono in media 8 giorni tra sintomi e decesso (fonte ISS). E' molto probabile pero' che la diagnosi si presenti con un paio di giorni di ritardo rispetto ai sintomi.



Sulla base di questa ipotesi, i decessi registrati oggi derivano in media dai casi di 5 giorni fa. Propongo quindi di calcolare il tasso di letalità dividendo i decessi ad oggi con il numero di casi k giorni fa, dove scelgo $k = 5$. Il mio modo di calcolare il tasso di letalità si contrappone a una modalità "standard" utilizzata in molti mezzi di informazione.⁶

$$\text{Letalità}_{t-k} = \frac{\text{Deceduti Totali}_t^{\text{trend}}}{\text{Casi Totali}_{t-k}^{\text{trend}}}, \quad \text{Letalità}_t^{\text{standard}} = \frac{\text{Deceduti Totali}_t}{\text{Casi Totali}_t}$$

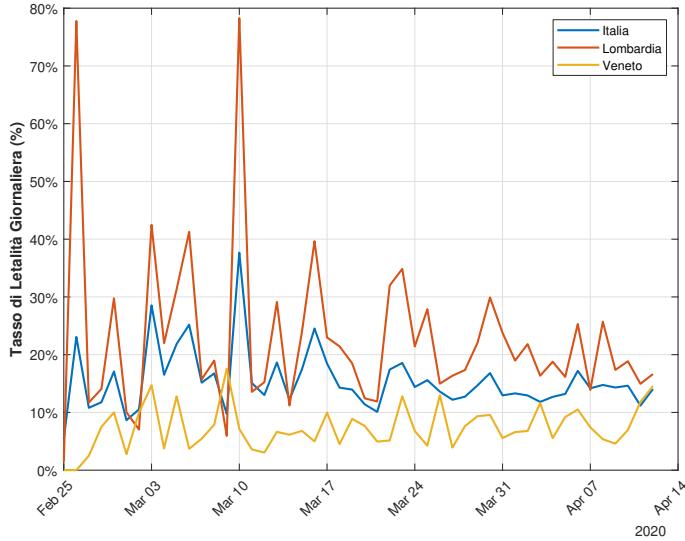


⁶Per essere precisi, quello che stimo è il tasso medio di letalità nei dati italiani dall'inizio della crisi fino a 5 giorni fa. Si noti che maggiore è il valore di k , maggiore è il tasso di letalità stimato, perchè si comparano i decessi totali al tempo t con il numero di casi totali al tempo $t - k$ (che è per definizione inferiore al crescere di k). Il tasso di letalità ottenuto con il metodo standard (cioè con $k = 0$) è quindi sempre il minor tasso di letalità stimabile con i dati disponibili durante un'epidemia.

Costruisco anche una misura del tasso di letalità giornaliero come segue:

$$\text{Letalità Giornaliera}_{t-k} = \frac{\text{Deceduti Giornalieri}_t}{\text{Casi Giornalieri}_{t-k}}$$

e riporto la sua evoluzione temporale per l'Italia, la Lombardia e il Veneto:



2.2 Stima del numero di giorni tra diagnosi e decesso

Ritengo che il modo ‘standard’ di calcolare il tasso di letalità sia ingannevole se si applica *durante* l’epidemia. Questo perchè le fonti ministeriali ci dicono che intercorrono in media 8 giorni tra sintomi e decesso. Calcolare la letalità comparando i deceduti di oggi con i casi di oggi equivale ad assumere che il decesso avviene lo stesso giorno in cui avviene la diagnosi. Il mio ragionamento è supportato dal seguente studio epidemiologico: <https://academic.oup.com/aje/article/162/5/479/82647>.

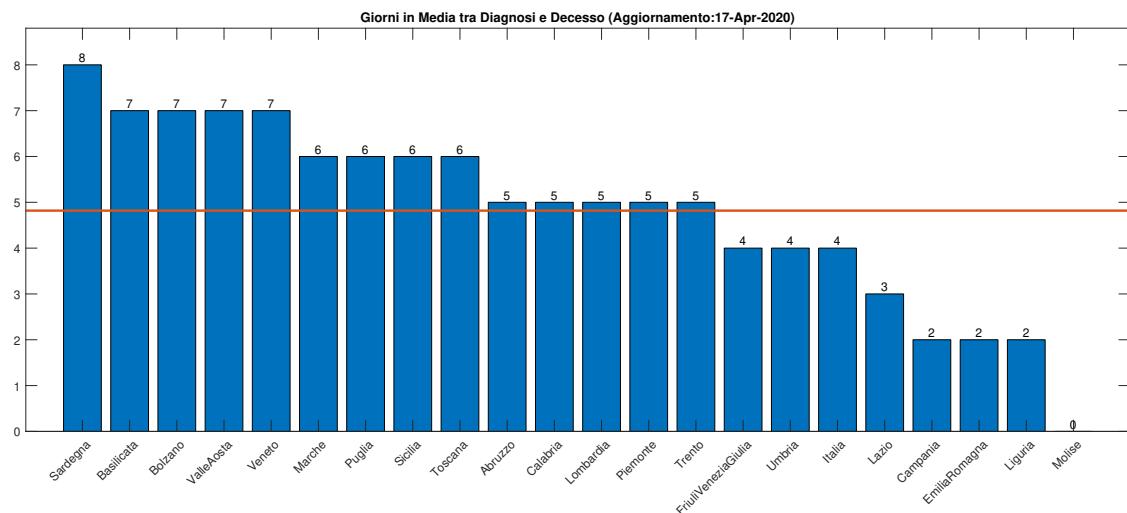
E’ evidente quindi che bisogna comparare i deceduti ad oggi con i casi di k giorni passati, ma non è ovvio scegliere k . Scelgo di adottare $k = 5$ giorni in media tra il momento in cui viene diagnosticato il caso (e quindi venga conteggiato nei casi totali) e il momento in cui viene registrato il decesso, perchè ipotizzo due giorni tra la comparsa dei sintomi e la diagnosi (un giorno per eseguire il test e un giorno per avere i risultati). Ovviamente potrei sbagliarmi, ed è per questo che ricorro a una procedura statistica. Eseguo la seguente regressione lineare per ogni valore del lag k :⁷

$$\text{Deceduti Totali}_t^{\text{trend}} = \alpha + \beta \cdot \text{Casi Totali}_{t-k}^{\text{trend}} + \varepsilon$$

e scelgo il valore di k che massimizza la statistica R^2 , cioè l’abilità’ del regressore di prevedere la variabile dipendente nel campione.⁸ Ripeto l’esercizio regione per regione:

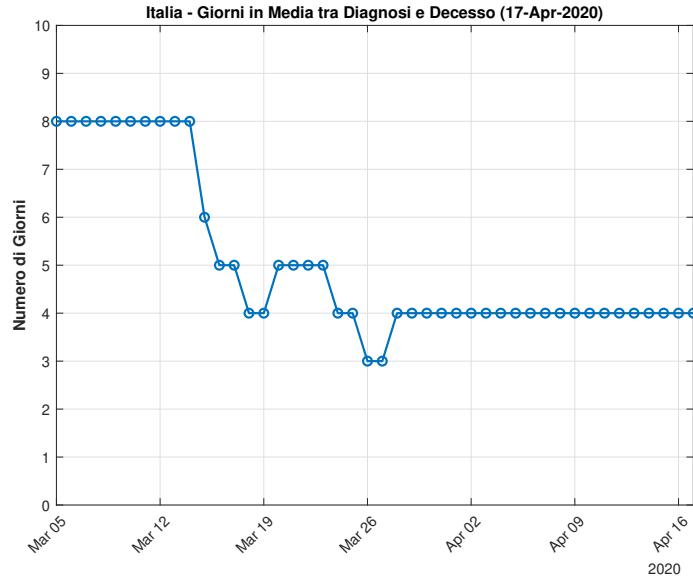
⁷Si ottiene un risultato simile con la seguente regressione: $\text{Deceduti Giornalieri}_t^{\text{trend}} = \alpha + \beta \cdot \text{Casi Giornalieri}_{t-k}^{\text{trend}} + \varepsilon$

⁸Impongo che il valore massimo di k sia 8.

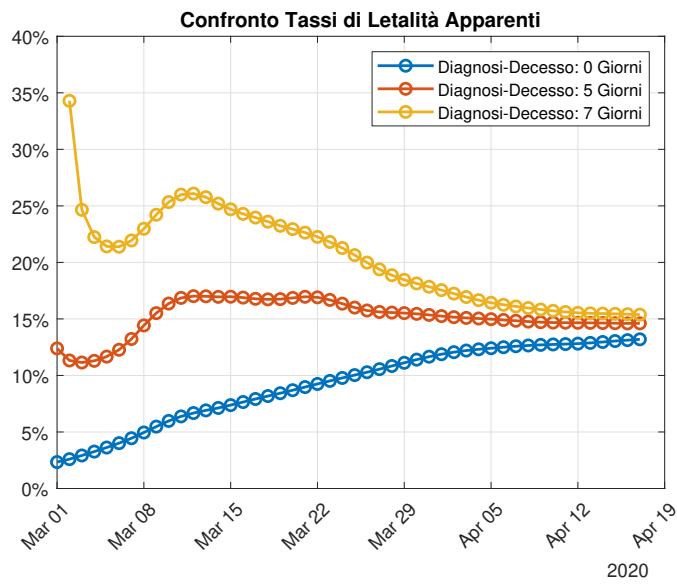


Ci sono significative differenze regionali, ma l'idea di fondo è che in media, il numero di giorni intercorsi tra diagnosi e decesso che massimizza la statistica R^2 oscilla tra 4, 5 e 6, fornendo supporto alla mia ipotesi.

Posso analizzare l'evoluzione temporale stimato tra diagnosi e decesso nei dati. Si noti che questa è un'informazione utile per tenere traccia dell'abilità del sistema sanitario di tenere il passo con l'epidemia. Infatti, un valore più basso indica un minor numero di giorni tra diagnosi e decesso, che è sintomo di un sovraccarico del sistema sanitario.



Infine, confronto l'evoluzione temporale del tasso di letalità calcolato con diverse ipotesi circa il numero di giorni tra diagnosi e decesso.



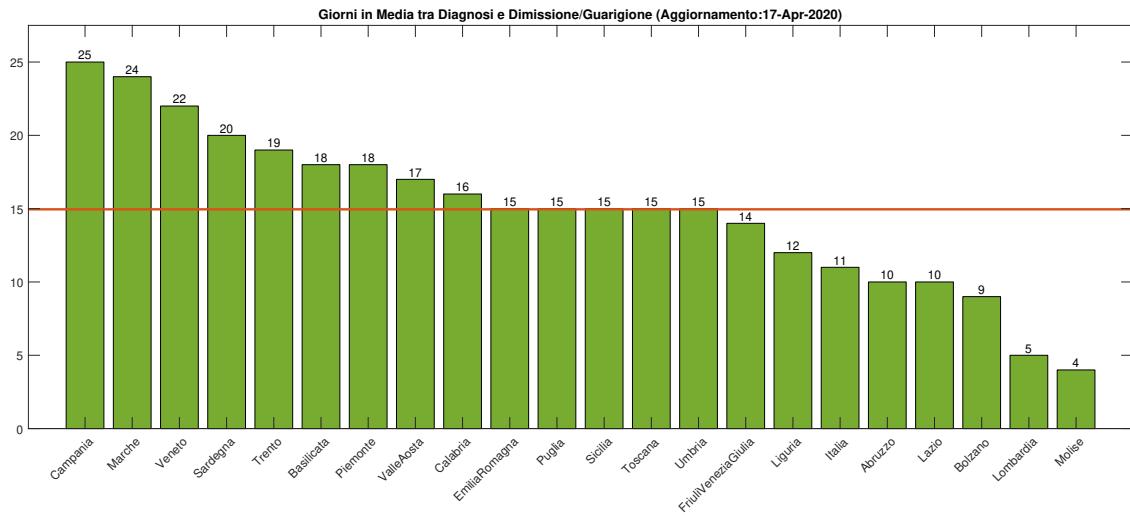
Come spiegato nella letteratura epidemiologica, il tasso di letalità calcolato senza tenere conto dei giorni tra diagnosi e decesso cresce *meccanicamente* nel tempo. Se invece si ipotizza un numero di giorni tra diagnosi e decesso troppo elevato, il tasso di letalità decresce *meccanicamente* nel tempo. La stima più verosimile è quella che mantiene il tasso di letalità costante nel tempo.

2.3 Stima del numero di giorni tra diagnosi e dimissione/guarigione

Utilizzando una procedura molto simile a quella per stimare il lag medio tra diagnosi e decesso, posso stimare il lag medio tra diagnosi e dimissione guarigione.⁹ Eseguo quindi la seguente regressione lineare per ogni valore del lag k :

$$\text{Dimessi_Guariti Totali}_t^{\text{trend}} = \alpha + \beta \cdot \text{Casi Totali}_{t-k}^{\text{trend}} + \varepsilon$$

e scelgo il valore di k che massimizza la statistica R^2 , cioè l'abilità del regressore di prevedere la variabile dipendente nel campione.¹⁰ Ripeto l'esercizio regione per regione:

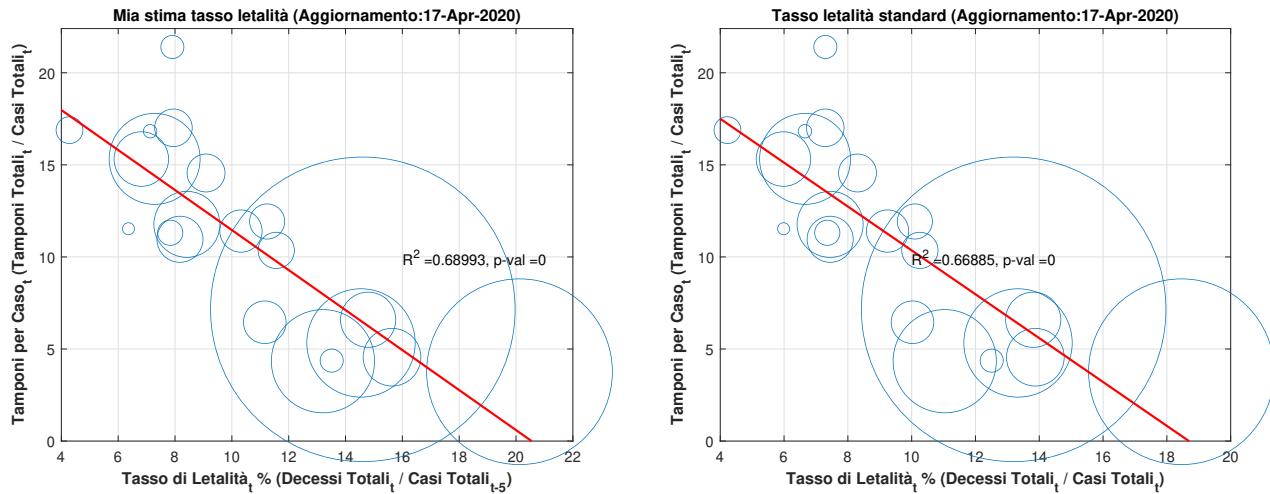


⁹La variabile considerata è “dimessi_guariti” fornita alla Protezione Civile. Idealmente, la variabile misura il numero di individui guariti dal Covid-19.

¹⁰Impongo che il valore massimo di k sia 25.

2.4 Letalità e Tamponi: Confronto tra Regioni

Per cercare di razionalizzare gli alti tassi di letalità italiani, comparo la mia stima del tasso di letalità a livello regionale con la propensione a testare a livello regionale, misurata come ‘tamponi totali’ diviso per ‘casi totali’ regione per regione. Ogni cerchio rappresenta una regione, e la grandezza del cerchio rappresenta il numero totale di casi in quella regione. Per assicurarmi che il risultato non sia indotto dal mio metodo di stima della letalità, ripeto l'esercizio anche con la misura standard di letalità (deceduti oggi diviso per casi oggi).



La figura mostra che il tasso di letalità è più elevato nelle regioni che hanno una propensione a testare minore. L'ipotesi naturale è che quando si testa di meno, si testano casi più gravi e non si testano i casi meno gravi, quindi il tasso di letalità sale. Inoltre, le regioni con più casi (rappresentato da un cerchio più grande), sono quelle che testano di meno (probabilmente in maniera strategica per preservare risorse sanitarie, ma forse anche perché faticano a stare al passo con l'epidemia).

Tuttavia, si noti che la bassa propensione a testare è solamente uno dei fattori che potrebbe spiegare l'elevato tasso di letalità dei dati italiani. Infatti la statistica R^2 ci dice - in termini grezzi - che la differente propensione a testare spiega una frazione (che negli ultimi giorni oscilla tra il 20% e il 30%) della differente letalità tra regioni.

2.5 Stima Casi Effettivi

I tassi di letalità presenti nei dati del ministero sono eccezionalmente più alti dei tassi di letalità suggeriti sia dalla Cina sia dal rapporto dell'Imperial College di Ferguson et al., riportati nella tabella sottostante e applicati alla struttura demografica italiana.

Fascia Età	Letalità % (Dati Cina)	Letalità % (Dati Ferguson et al)	Popolazione Italiana %
0-9	0,00	0,00	8,43
10-19	0,20	0,01	9,56
20-29	0,20	0,03	10,27
30-39	0,20	0,08	11,72
40-49	0,40	0,15	15,31
50-59	1,30	0,60	15,49
60-69	3,60	2,20	12,16
70-79	8,00	5,10	9,88
80+	14,80	9,30	7,17

Data la struttura demografica italiana, i tassi di letalità ‘teorici’ dovrebbero essere per l’Italia: 2.62% (dati cinesi), 1.57% (dati Ferguson et al.).

Ipotizzando una *distribuzione uniforme dei contagi tra fasce d’età* e ipotizzando che la *vera letalità in Italia coincida con questi valori*, posso stimare quindi il numero di casi totali effettivi in Italia e regione per regione che sarebbero necessari a generare il tasso di letalità teorico.¹¹ Infatti:

$$\text{Letalità teorica} = \frac{\text{Deceduti}_t}{\text{Casi Effettivi}_{t-5}}, \quad \text{Letalità dati italiani} = \frac{\text{Deceduti}_t}{\text{Casi Rilevati}_{t-5}}$$

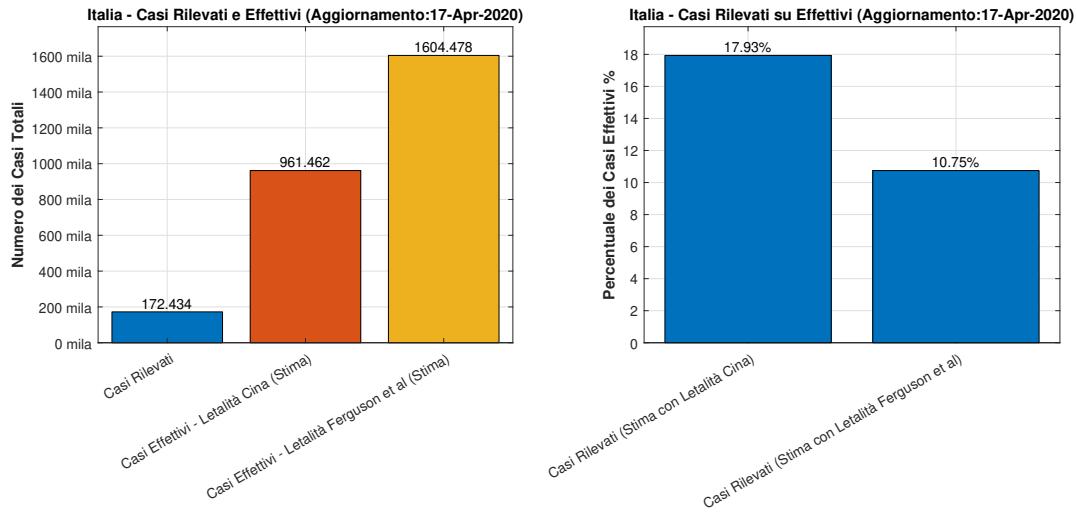
Abbiamo quindi che

$$\frac{\text{Letalità dati italiani}}{\text{Letalità teorica}} = \frac{\text{Casi Effettivi}_t}{\text{Casi Rilevati}_t}$$

per ogni valore di t , e che si possono stimare i casi effettivi come segue:

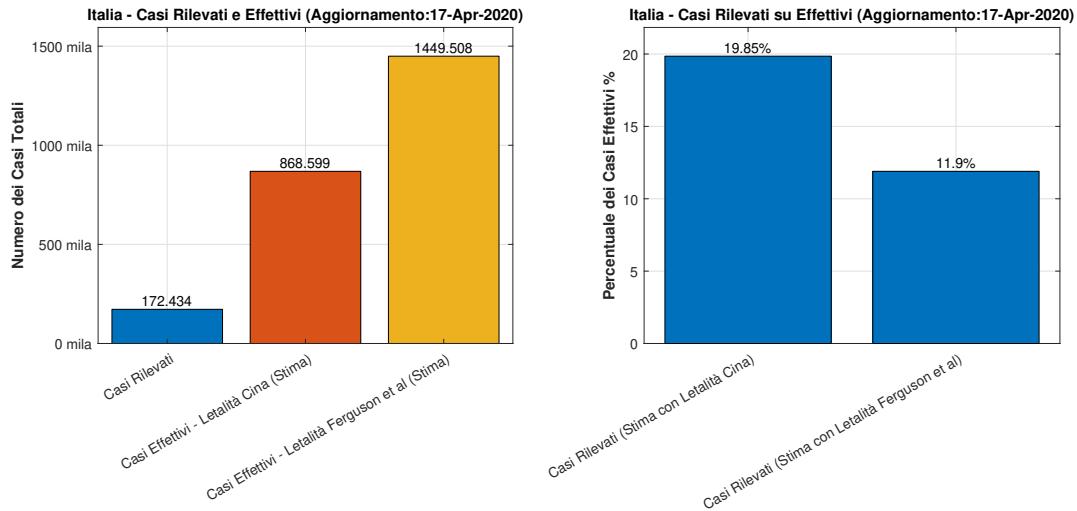
$$\text{Casi Effettivi}_t = \text{Casi Rilevati}_t \times \frac{\text{Letalità dati italiani}}{\text{Letalità teorica}}$$

¹¹Purtroppo, l’ipotesi di contagio uniforme tra fasce d’età non è verificabile. Guardare i casi rilevati dal sistema sanitario per fascia d’età non è informativo al riguardo poiché il testing non è casuale, ma selettivo (si privilegia test in soggetti più gravi, che tendono ad essere più anziani.).



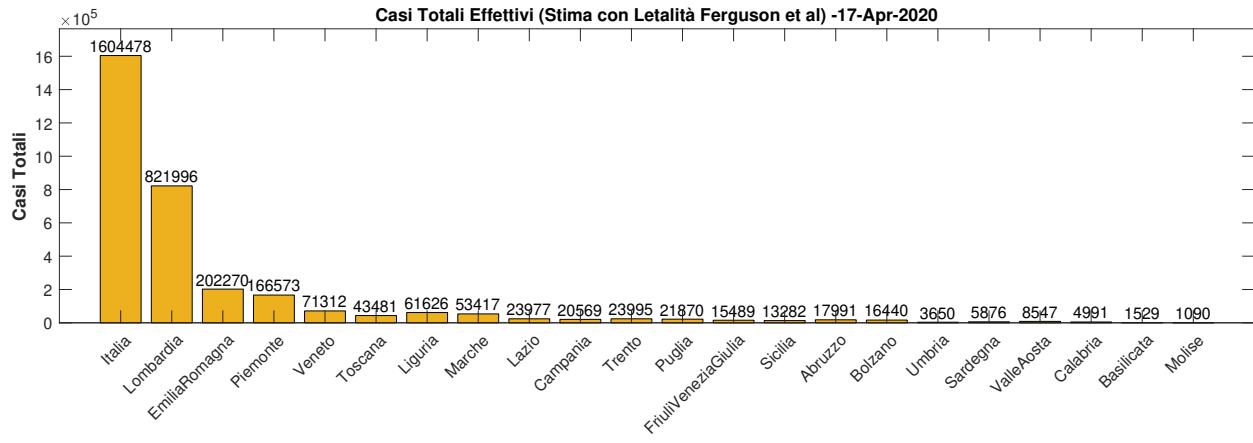
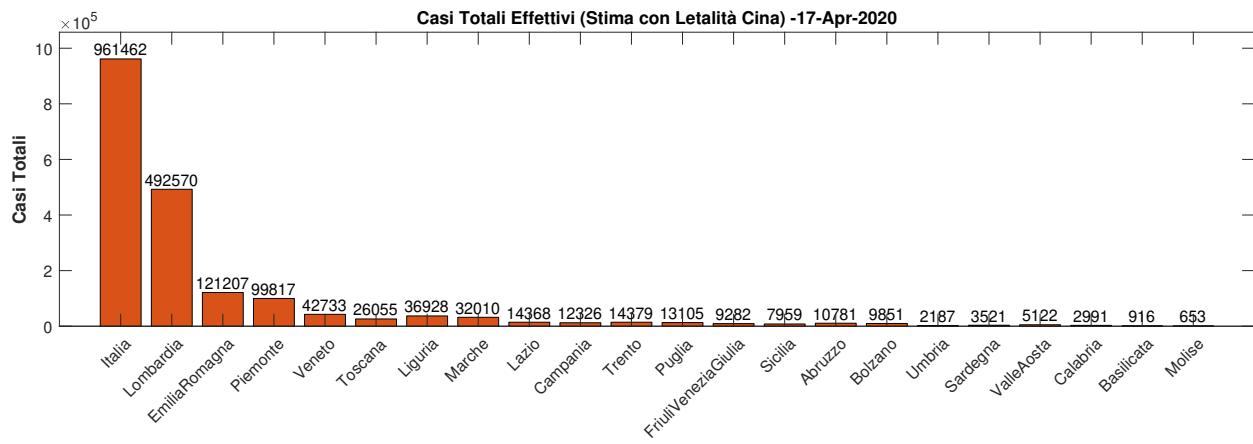
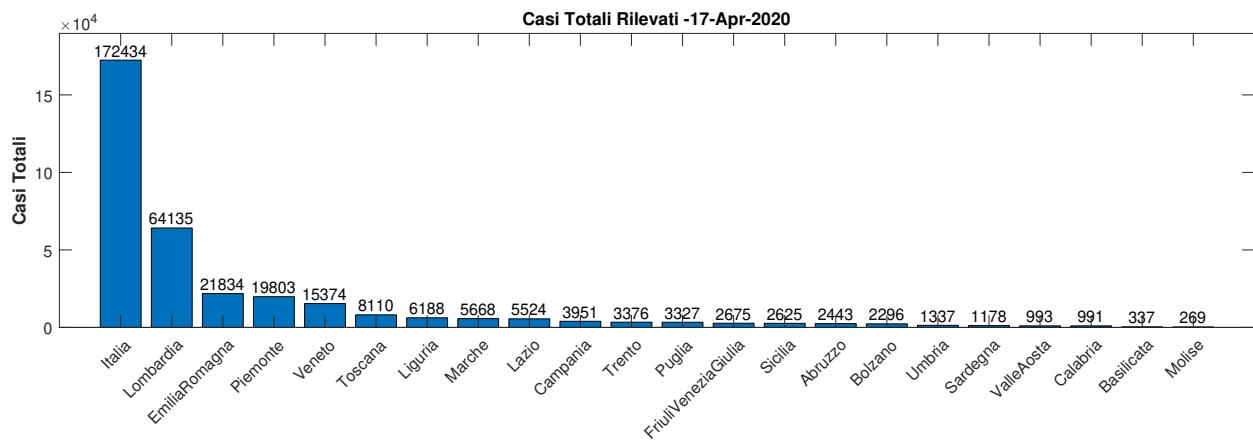
Si noti che, maggiore il tasso di letalità italiano stimato, maggiore il numero dei casi effettivi. Quindi, il numero di casi effettivi varia enormemente se si usa il tasso di letalità standard o quello da me stimato. Difendo la mia scelta di stimare il tasso di letalità comparando i decessi ad oggi con i casi di 5 giorni fa sulla base di: 1) informazioni del Ministero; 2) selezione attraverso procedura statistica.

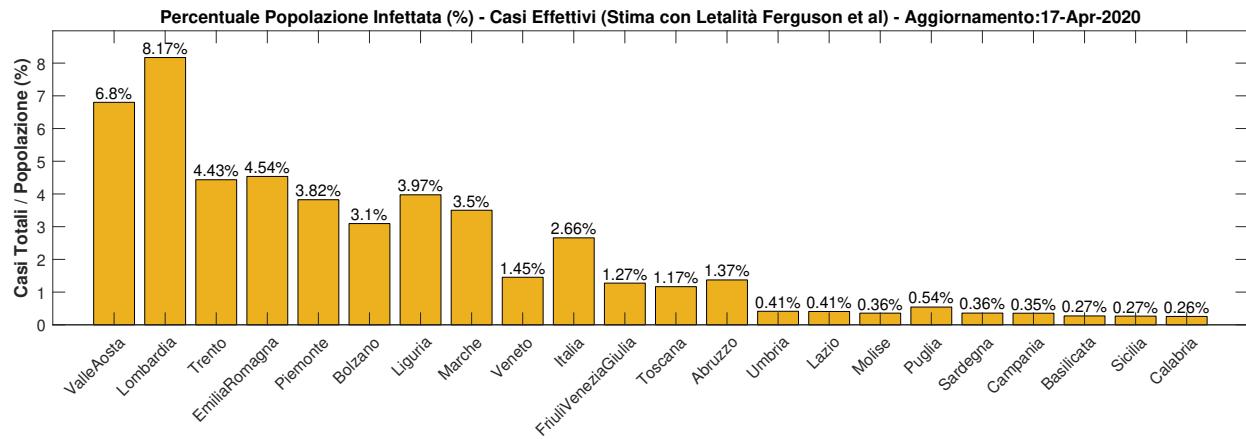
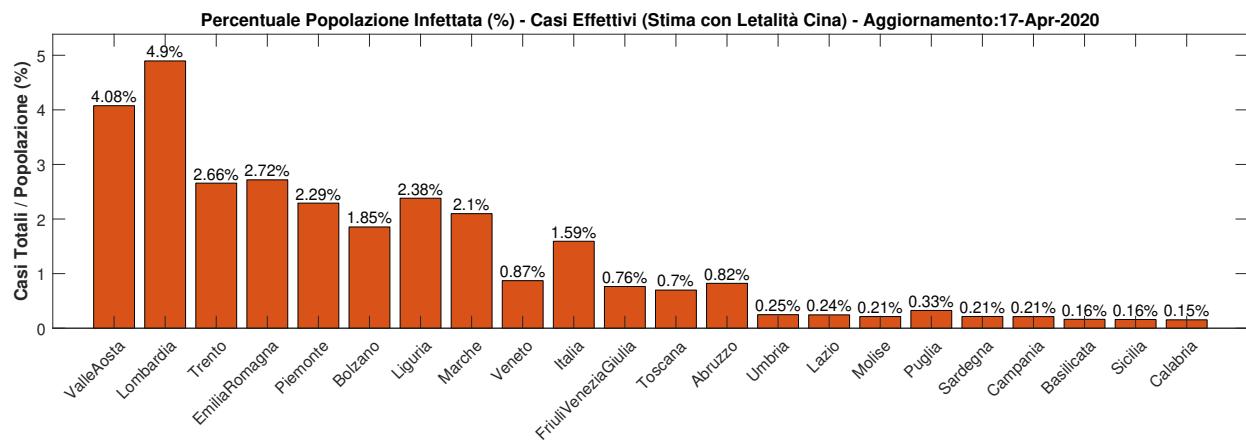
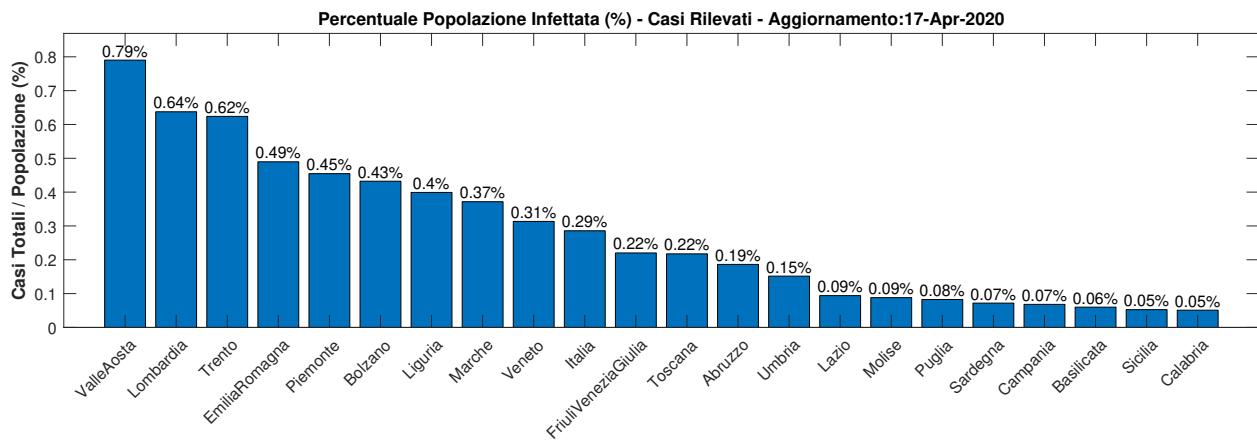
Ripeto pero' per completezza l'esercizio usando i tassi di letalità ottenuti con il metodo standard (deceduti oggi diviso per casi oggi):



Poichè il tasso di letalità calcolato con il metodo ‘standard’ è sempre il minor tasso di letalità stimabile, ne risulta che i casi effettivi stimati con esso siano il *limite inferiore* dei casi effettivi. In altre parole, il numero di casi effettivi è sempre *superiore* alla stima riportata nella figura sopra.

Le stime regione per regione che seguono sono ottenute utilizzando la mia stima del tasso di letalità:





Dividere per la popolazione aiuta a capire quali sono le regioni più colpite, tenendo conto che alcuni regioni sono più grandi (in termini di popolazione residente) di altre. è importante affiancare al numero totale dei casi, il numero totale dei casi diviso per la popolazione della regione.¹²

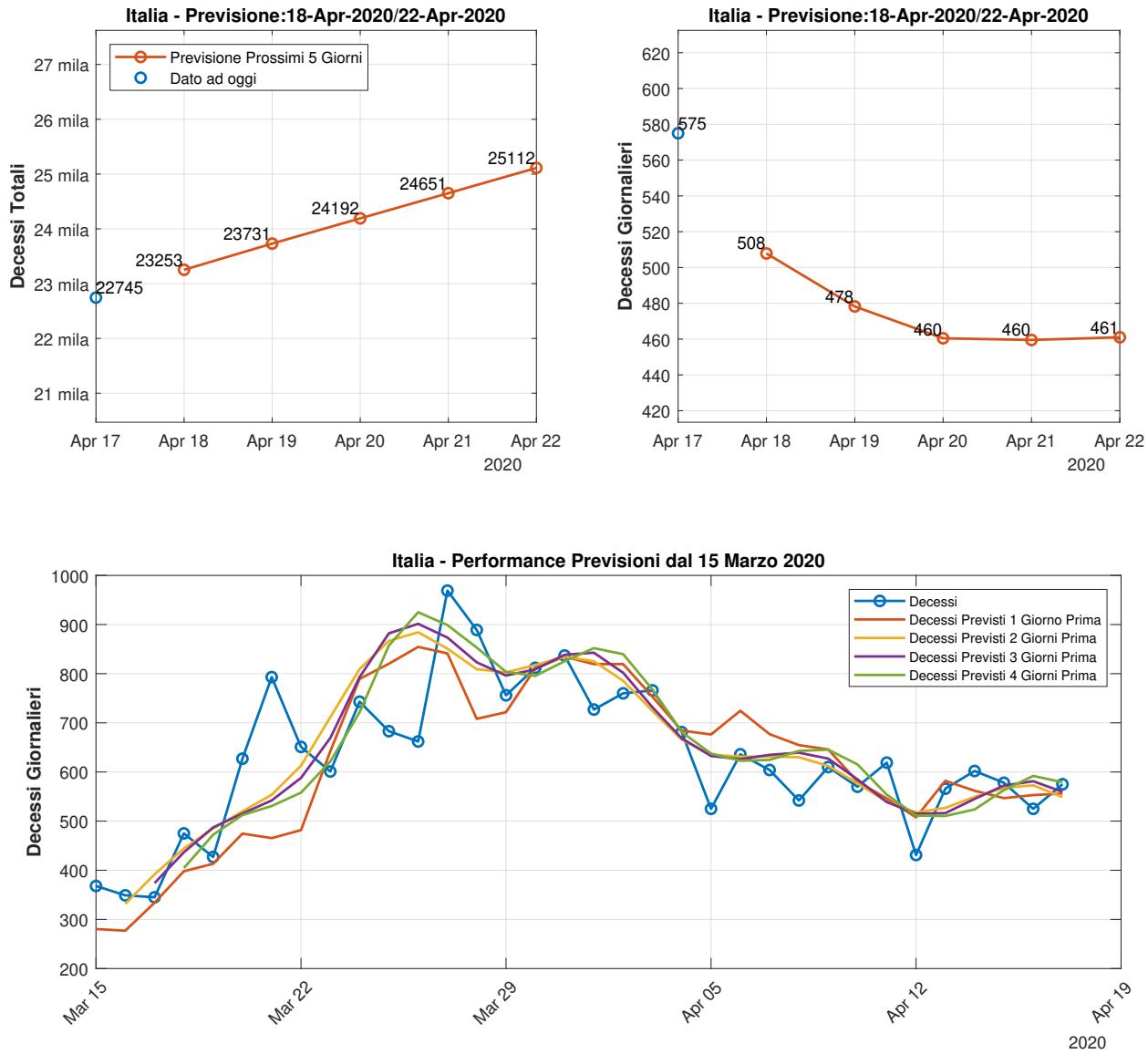
¹²I grafici rappresentano la percentuale di popolazione infettata dall'inizio della crisi ad oggi. Quest'ultima è data da casi totali rilevati divisi per la popolazione residente (blu), casi totali effettivi stimati (con la letalità cinese) divisi per popolazione (arancio), casi totali effettivi stimati (con la letalità di Ferguson et al.) divisi per popolazione (giallo).

2.6 Previsione dei Decessi

2.6.1 Prossimi 5 Giorni

Se la stima dei tassi di letalità è corretta, dovrei essere in grado di fornire una previsione corretta dei decessi a 5 giorni. Infatti:¹³

$$\text{Deceduti}_{t+k|t} = \text{Casi}_t^{\text{trend}} \times \text{Tasso Letalità Stimato}_t$$



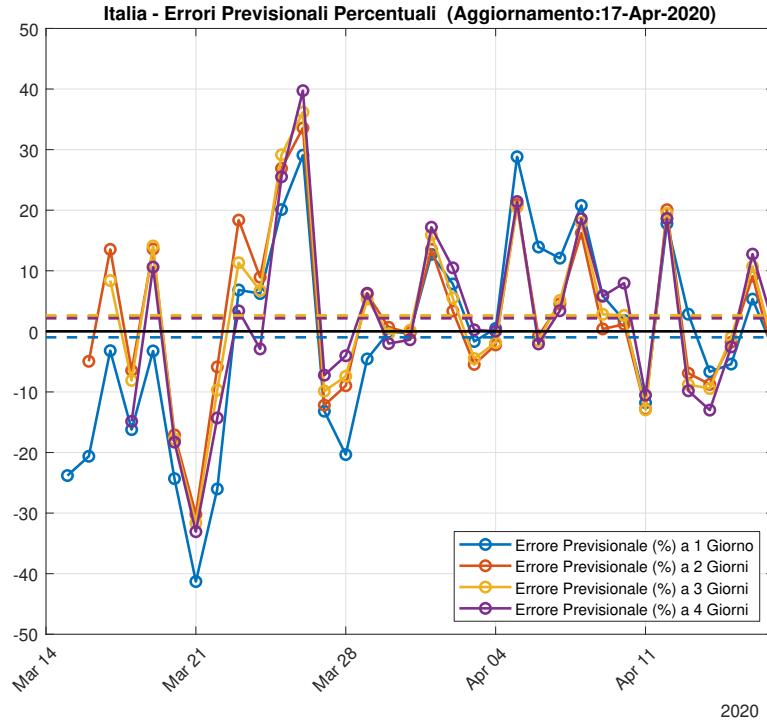
¹³Si noti che questa è più simile a una *nowcast* che a una *forecast*. La mia previsione sovrasta il dato attuale quando la mia stima del tasso di letalità è maggiore di quello effettivo, e sottosta quando è inferiore a quello infettivo.

2.6.2 Errori Previsionali Percentuali

Gli errori previsionali percentuali (EP) sono calcolati come segue:

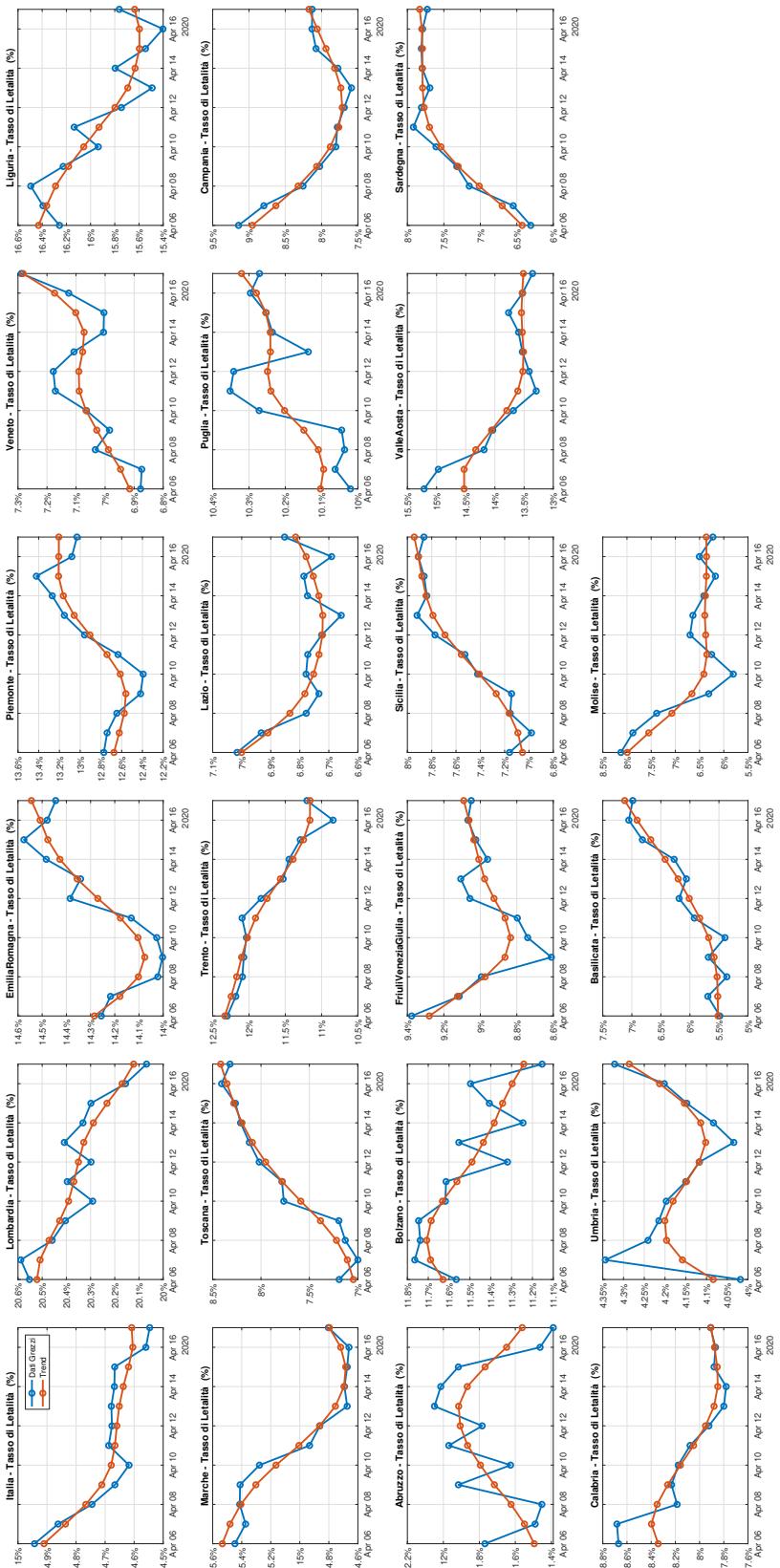
$$EP_t^j = \frac{\text{deceduti giornalieri previsti}_{t|t-j} - \text{deceduti giornalieri}_t}{\text{deceduti giornalieri}_t}$$

dove j è l'orizzonte previsionale. Comparo la previsione del modello a 1, 2, 3 e 4 giorni in avanti con i decessi effettivamente avvenuti. Nel complesso la procedura è molto precisa:

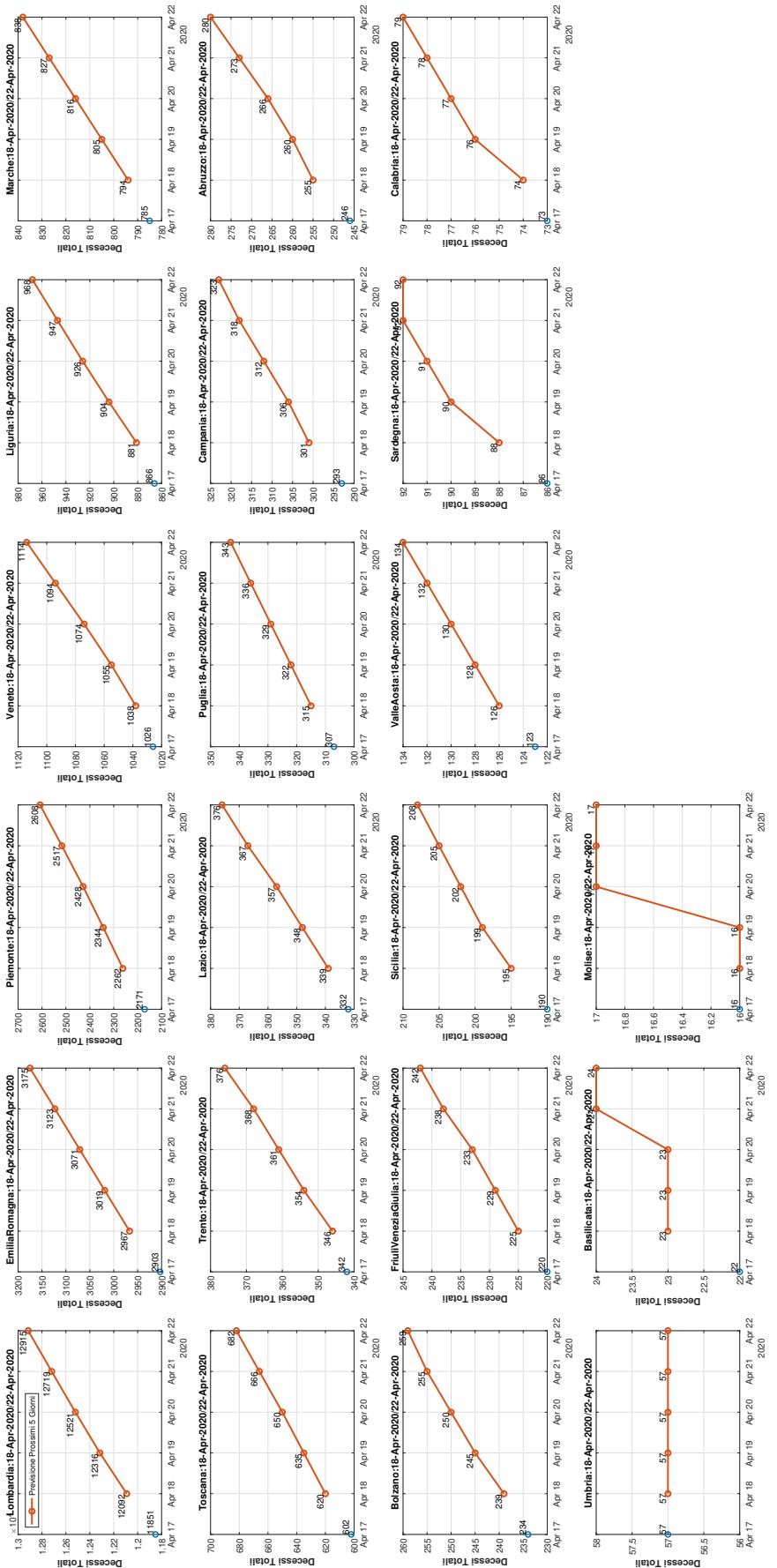


Appendice

Evoluzione Tasso di Letalità Stimato - Regioni



Previsione Decessi Totali - Prossimi 5 Giorni - Regioni



Previsione Decessi Giornalieri - Prossimi 5 Giorni - Regioni

