

UNIVERSITÀ
DEGLI STUDI
DI PADOVA



COVID-19: La statistica racconta la pandemia

Laura Ventura

Dipartimento di Scienze Statistiche
Università degli Studi di Padova

ventura@stat.unipd.it

Martedì 15 dicembre 2020



L'ombra di Galileo



Qualche ben nota informazione iniziale

- L'**epidemia da COVID-19** (*CoronaVirus Disease 2019*) è dovuta a una malattia infettiva respiratoria causata dal virus SARS-CoV-2 della famiglia dei coronavirus.
- I primi casi di COVID-19 risalirebbero al mese di dicembre 2019 quando si sono registrati alcuni ricoveri per polmonite da causa sconosciuta tra la popolazione di Wuhan, nella provincia di Hubei in Cina.
- In Italia il virus arriva inizialmente da casi indiretti provenienti dalla Cina (turisti a Roma) per poi diffondersi con casi diretti in alcune località del Veneto e della Lombardia (20-25 febbraio). Ma casi con ceppo di Wuhan erano forse già presenti nel nord Italia a fine novembre 2019.
- **Al 14 dicembre (dopo 295 giorni di pandemia dal 24 febbraio) in Italia sono stati accertati 1.855.737 casi e 65.011 decessi attribuibili al COVID-19.**

Qualche ben nota informazione iniziale

- Una persona infetta può presentare sintomi dopo un periodo di incubazione che varia tra 2 e 14 giorni circa, durante i quali può comunque essere contagiosa.
- I sintomi più frequenti sono febbre, stanchezza e tosse secca. Ma quasi un infetto su due non presenta sintomi (ad esempio, il 43% dei positivi era asintomatico nello studio* di Vo', Padova).
- Alcuni aspetti peculiari di questa malattia sono l'elevata contagiosità e la trasmissione in assenza di sintomi e/o due giorni prima della manifestazione dei sintomi.
- **Per limitare la diffusione del virus devono essere prese precauzioni, come lavarsi frequentemente le mani, indossare la mascherina, rispettare il distanziamento ed evitare gli assembramenti.**

(*Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo', *Nature*)

I dati disponibili

**«Se torturi i numeri abbastanza a lungo,
confesseranno qualunque cosa.»**

Ronald Coase

I dati disponibili

- I dati sono forniti dal **Ministero della Salute**.
- Sono elaborati e gestiti dal **Dipartimento della Protezione Civile**.
- L'**Istituto Superiore di Sanità (ISS)** offre report e analisi sull'andamento dell'epidemia.
- I **dati aggregati** con dettaglio nazionale, regionale, provinciale, comunale (aree metropolitane e grandi città), sul conteggio dei casi di contagio, ospedalizzazioni, ricoveri in terapia intensiva, decessi, guarigioni, tamponi, ecc, sono disponibili nel sito <https://github.com/pcm-dpc/COVID-19>.
- Altri dati relativi alle caratteristiche demografiche (sesso, età) e cliniche (presenza di altre comorbidity) sono desumibili dai report pubblicati dall'ISS.
- Purtroppo, alla comunità scientifica non sono resi disponibili **dati disaggregati**.



I dati comunicati

(c'è bisogno di diffondere *data literacy*)

I dati comunicati

Nel bollettino dell'ISS del 18/08/2020 si legge “L'età mediana dei casi diagnosticati è di 30 anni”, ma sui giornali ...

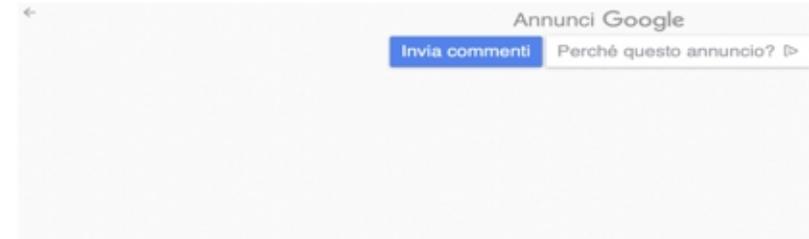
MENU | CERCA

la Repubblica

Coronavirus, età media dei nuovi casi intorno ai 30 anni



CORRIERE DELLA SERA / CRONACHE



IL BOLLETTINO DELL'ISTITUTO SUPERIORE DI SANITÀ
Covid, l'età più frequente dei contagi scesa a 30 anni. «Casi meno gravi e in maggioranza asintomatici». L'indice Rt è a 0,83

COVID-19

Iss: età dei contagi scesa a 30 anni. Importato il 28% dei nuovi casi

Il monitoraggio dell'Istituto Superiore di Sanità (Iss) e del ministero della Salute per la settimana dal 10 al 16 agosto: 1.077 focolai attivi



I dati comunicati

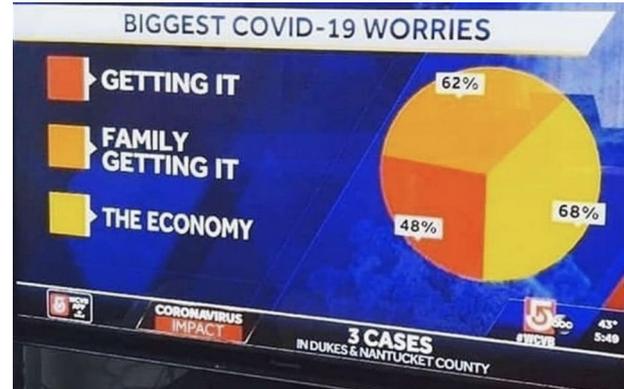
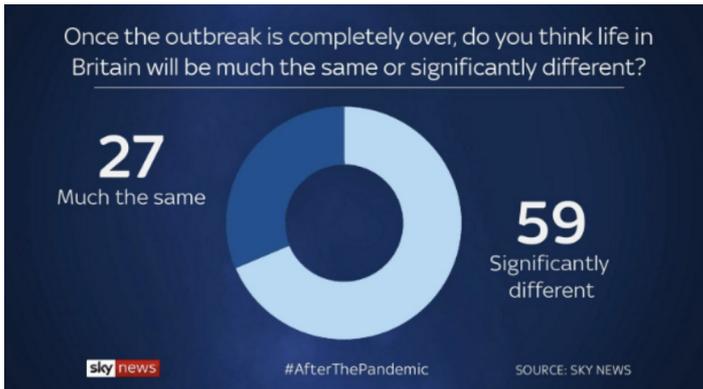
HOME CRONACA SPORT TERRITORIO RUBRICHE SALUTE IL CONSULENTE TEMPO LIBE

Se la curva si raffredda

Questo è il tempo non tanto di chiedersi con chi fare il cenone a Natale o se sarà possibile bridare a capodanno, ma piuttosto di resistere fino ad un abbassamento delle positività e a come, dopo, si potrà gestire una nuova fase evitando terze ondate.

Il termometro dell'epidemia - Hume

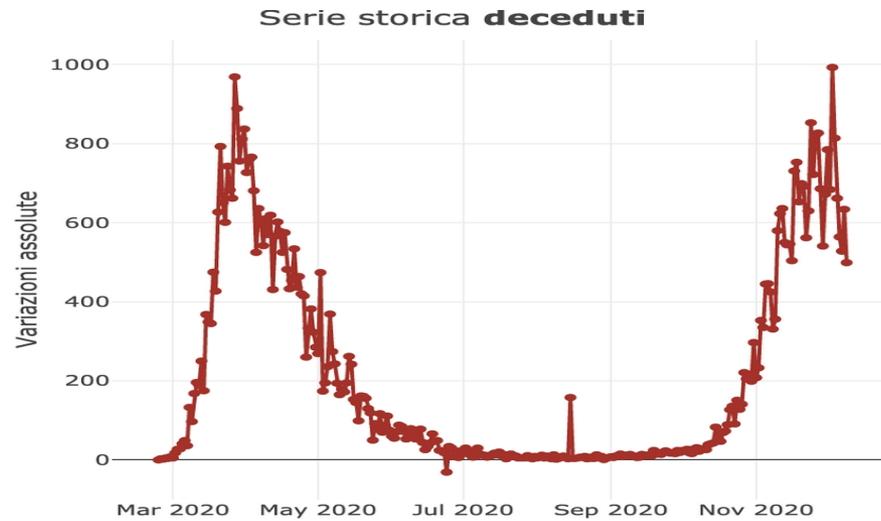
Anche oggi (ultimo dato disponibile, ore 18.00 del 9 luglio) la temperatura dell'epidemia è diminuita in misura modesta (-0.1 gradi), passando da 2.1 a 2.0 gradi pseudo-Kelvin.



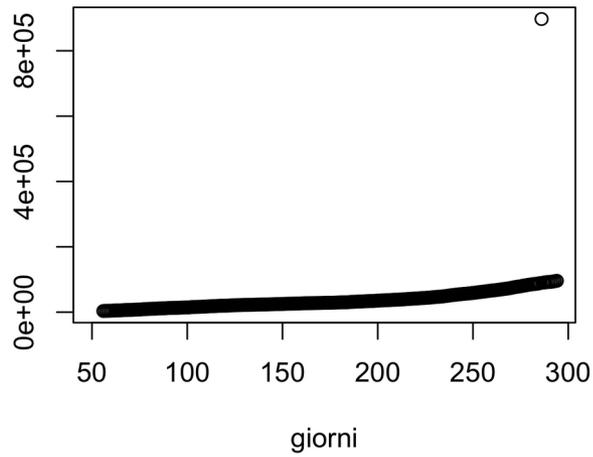
E nelle pubblicità:



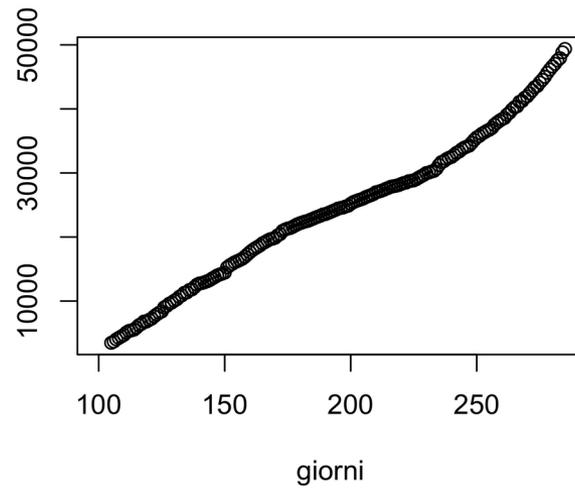
I dati



Casi testati Molise



Casi testati Molise



Dati di qualità e competenze

(firmate la petizione!)



https://www.change.org/o/societ%C3%A0_italiana_di_statistica



La Petizione



- **Lotta al COVID-19: sono necessari dati di alta qualità per le analisi e competenze adeguate per analizzarli.**
- **Perché dati accessibili?** Sulla base dei dati disponibili non è possibile svolgere alcune attività cruciali:
 - riprodurre le basi quantitative delle decisioni istituzionali;
 - valutare ex-post, in modo quantitativo e rigoroso, gli effetti delle decisioni;
 - comprendere aspetti ancora oscuri del fenomeno.
- Nel documento della SIS si chiede l'accesso a dati dettagliati.
- **Perché competenze adeguate?** Il processo scientifico richiede dei passaggi essenziali:
 - Definizione del problema;
 - Analisi dell'informazione;
 - Condivisione dell'informazione;
 - Diffusione dell'informazione.
- Le giuste competenze sono fondamentali per l'analisi di un fenomeno tanto complesso come la pandemia di COVID-19: è necessario saper incorporare nelle valutazioni l'enorme variabilità osservata e saper distinguere tra associazione e relazioni causali.

Dati di qualità e competenze

PAROLA AGLI STATISTICI

Affrontare il COVID-19 con dati e competenze

Una conversazione con
Francesca Bassi, Corrado Crocetta, Gianpiero Dalla Zuanna,
Paolo Girardi, Gianfranco Lovison, Antonietta Mira

Moderata da: Telmo Pievani

Martedì 22 dicembre 2020 alle 16:30
Live Youtube @scienzestatisticheunipd

800
ANNI



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

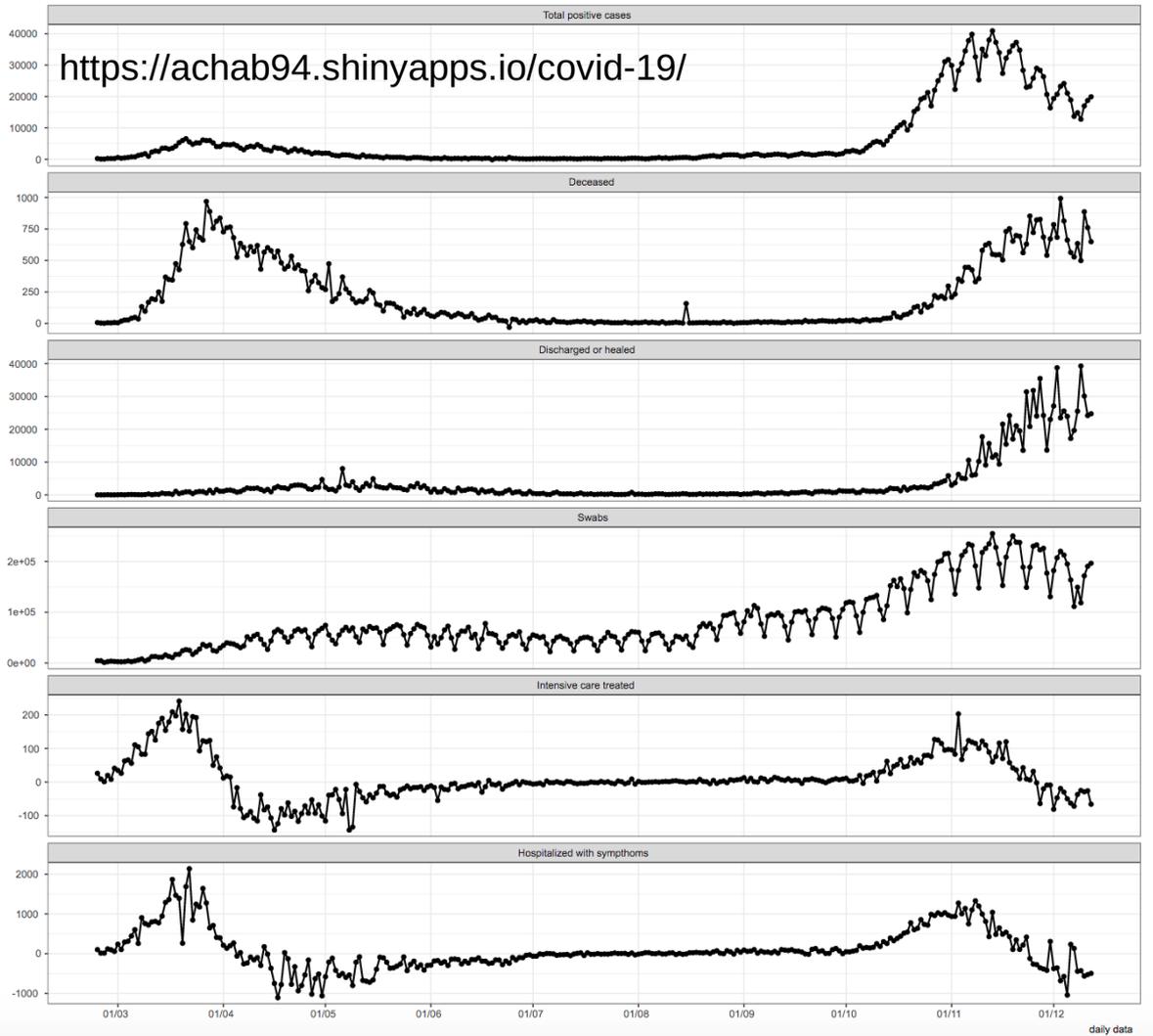


<https://www.facebook.com/events/234267261375682/>

@scienzestatisticheunipd

Allora con i dati disponibili...

Overall monitoring of COVID-19



"ricoverati_con_sintomi"
"terapia_intensiva"
"totale_ospedalizzati"
"isolamento_domiciliare"
"totale_positivi"

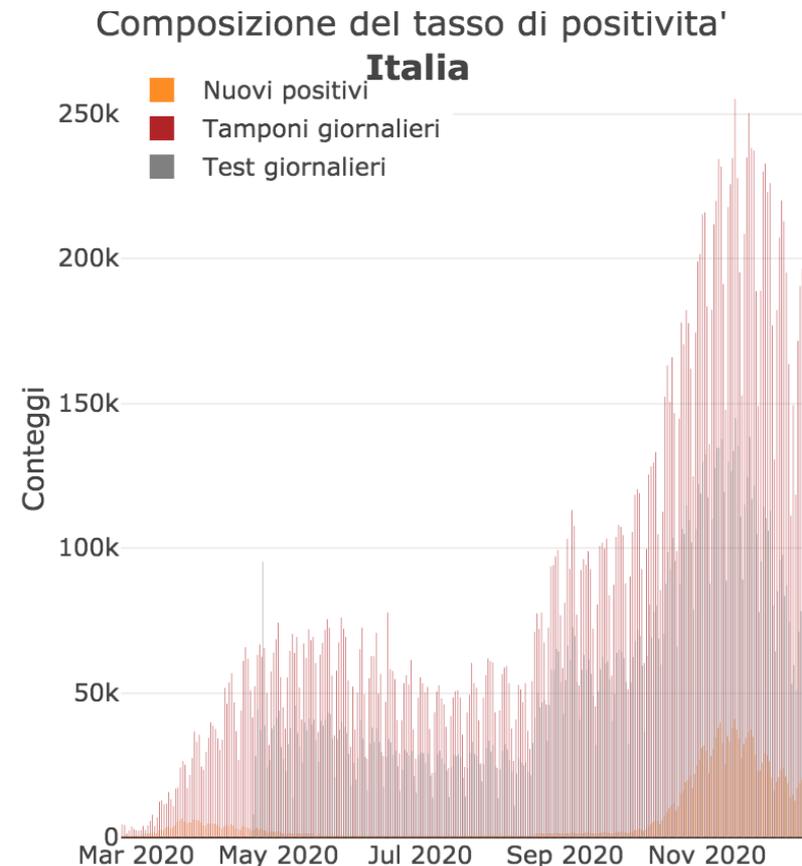
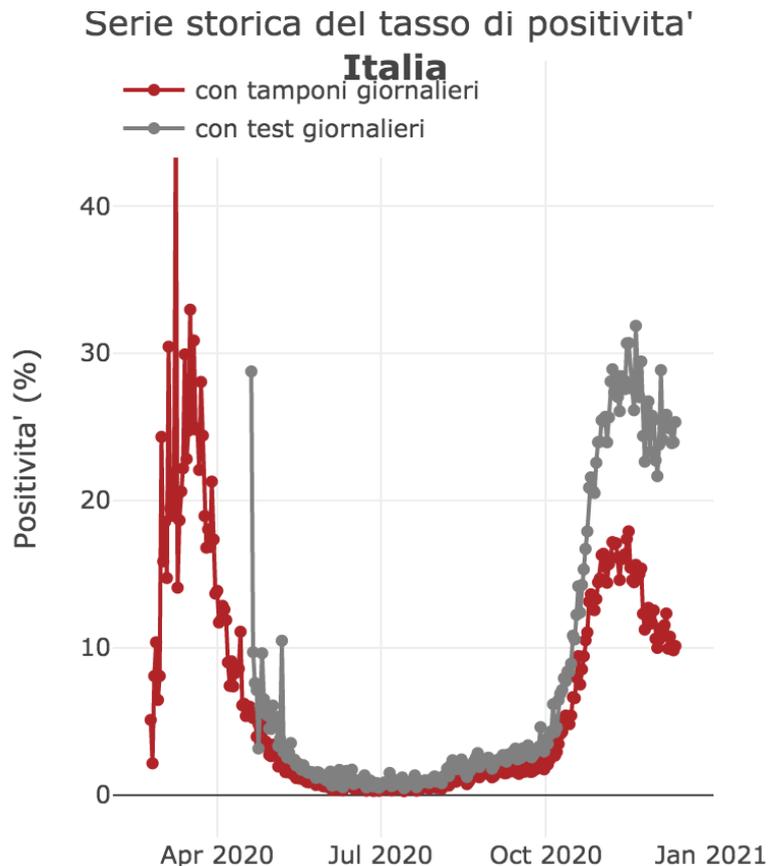
"dimessi_guariti"
"deceduti"

"totale_casi"
"tamponi"
"casi_testati"

... cosa ci può raccontare la statistica?

Diffidate di alcuni numeri assoluti: il tasso di positività

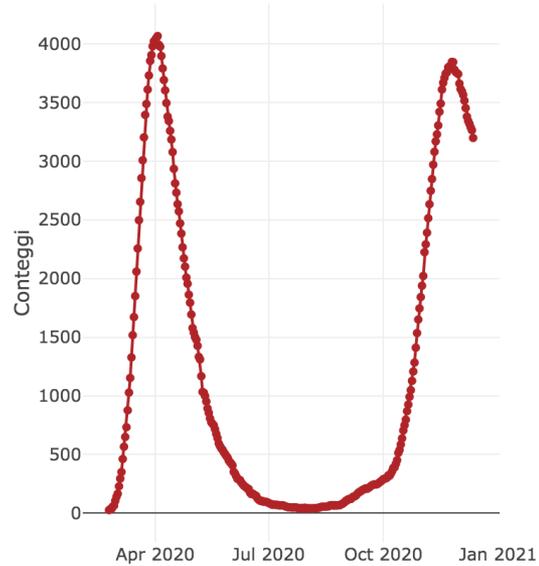
- Il numero di casi di contagio **dipende** dal numero di tamponi processati.
- Ma quali tamponi? Tamponi o casi testati?



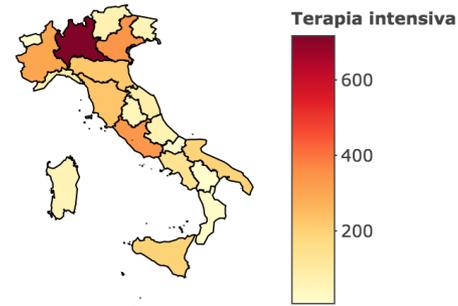
Fonte: <https://statgroup19.shinyapps.io/Covid19App/>

Presenze in terapia intensiva

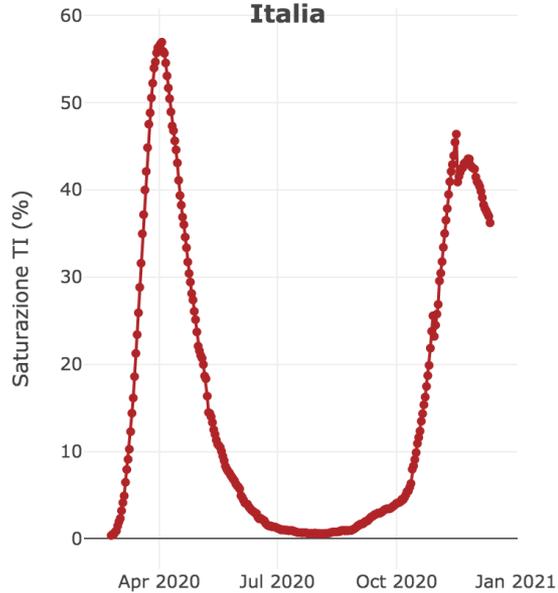
Serie storica **terapia intensiva**



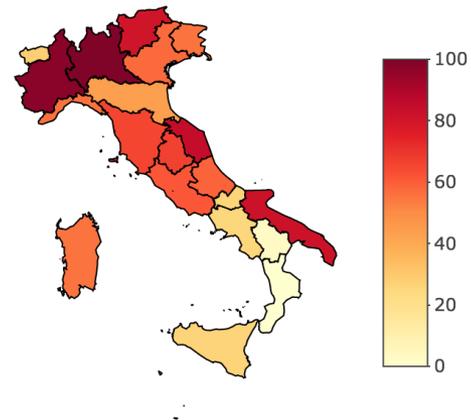
2020-12-12



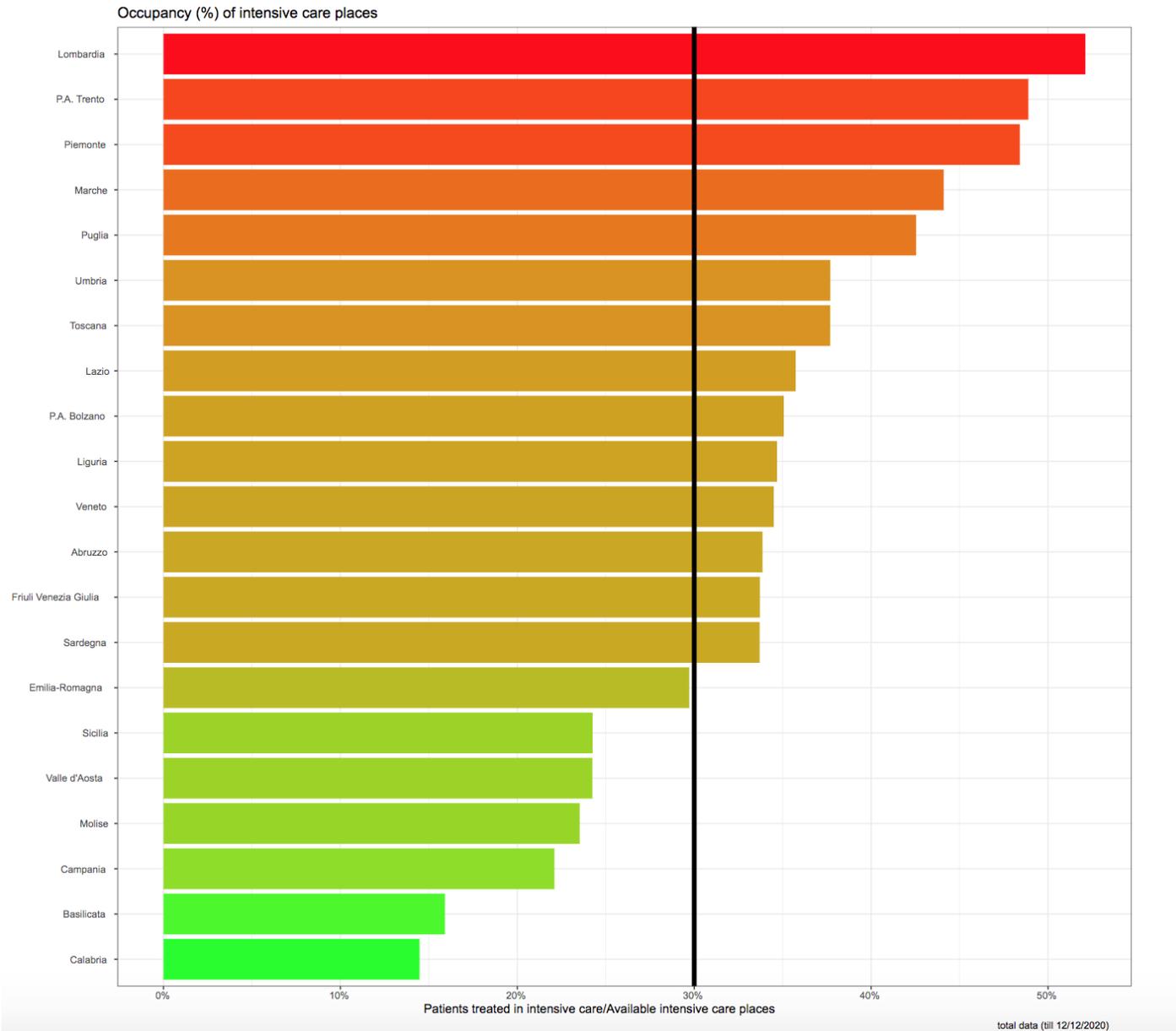
Serie storica del tasso di saturazione ti
Italia



Saturazione terapie intensive
2020-12-12



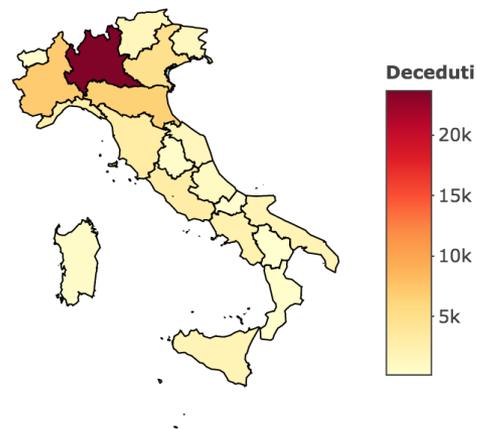
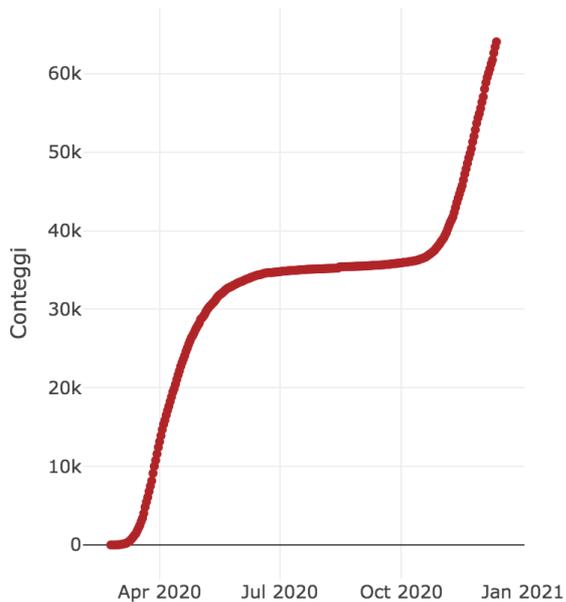
Presenze in terapia intensiva



Deceduti

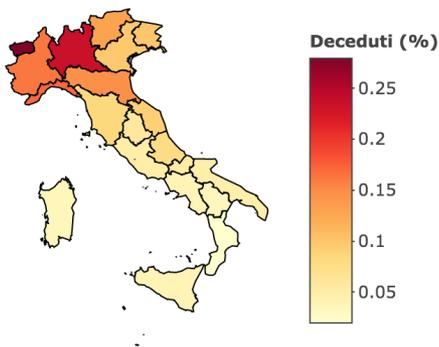
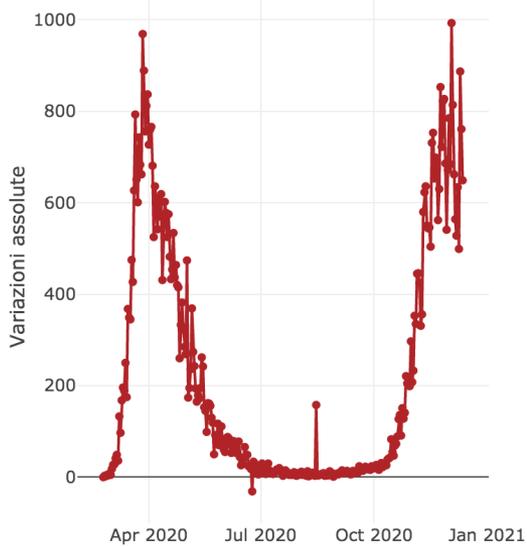
Serie storica **deceduti**

2020-12-12

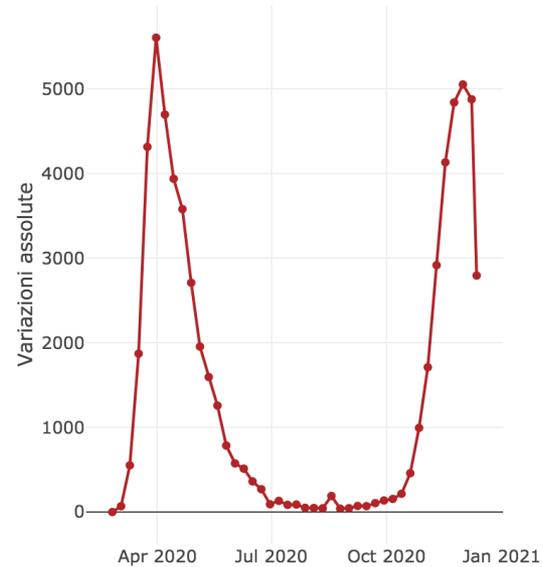


Serie storica **deceduti**

2020-12-12

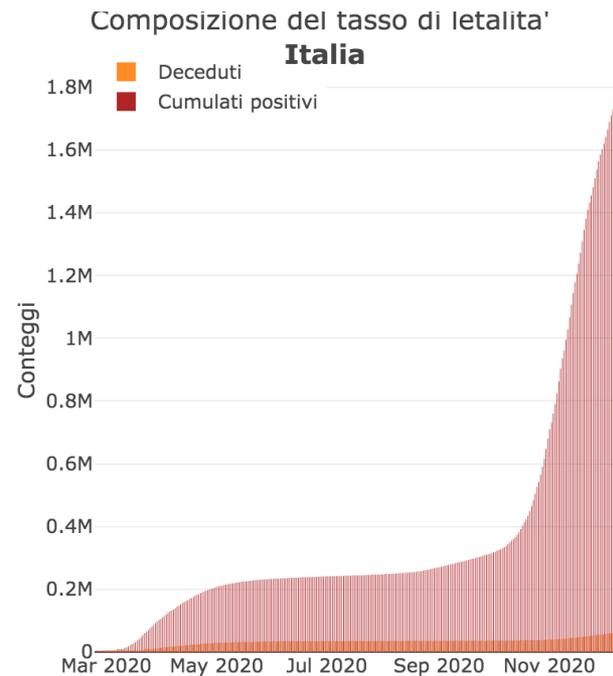
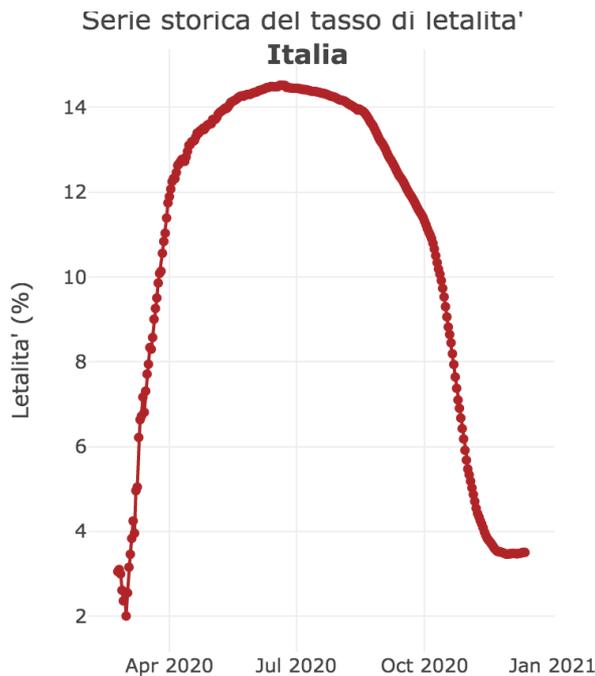


Serie storica **deceduti**



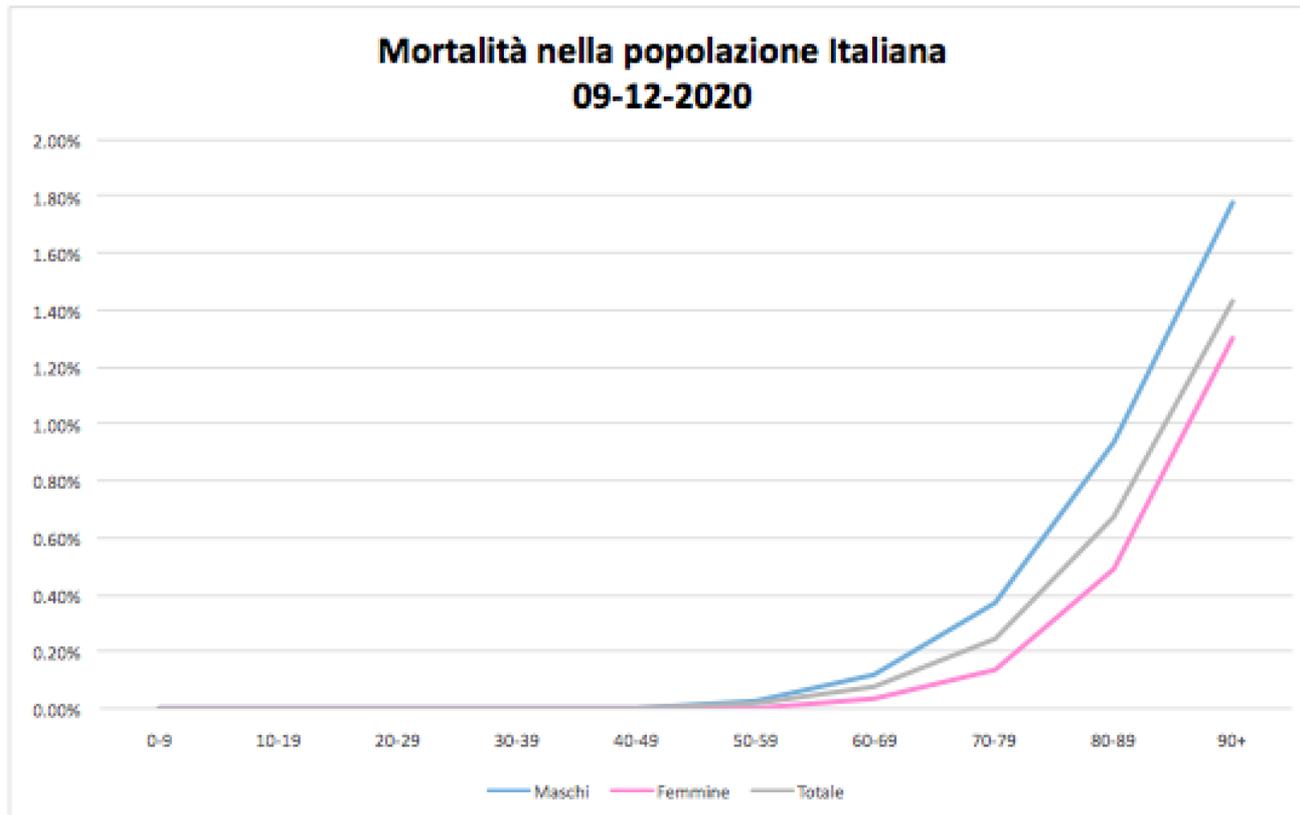
Letalità

- Il **tasso di letalità** è il rapporto tra il numero delle persone decedute a causa della malattia e il totale dei malati.
- Questo dato può oscillare molto a seconda del modo in cui si decide di rilevare quante persone sono malate: ad esempio, a causa delle scelte fatte dall'Italia, il nostro tasso di letalità risulta maggiore di quello di altri paesi che hanno più contagi.



Mortalità

- Il tasso di mortalità calcola quante persone sono morte sul totale delle persone esposte, cioè sugli abitanti del nostro paese.



La mortalità:

- cresce con l'età;
- è più alta per i maschi in tutte le classi d'età

Alcune considerazioni

- Premessa: i dati dipendono anche dalle situazioni (condizioni) in cui avviene la rilevazione e dai criteri adottati.
- Migliore capacità diagnostica: sono sottoposti a tampone tutti i casi sospetti e non soltanto i sintomatici.
- Cure più efficaci e maggiore diffusione delle norme anti-contagio.
- Maggiore mobilità (lavoro e scuola).
- Maggiore protezione degli anziani e delle categorie a rischio, in particolare degli operatori sanitari.
- Minore letalità.
- Con dati disaggregati si potrebbero dire molte più cose.

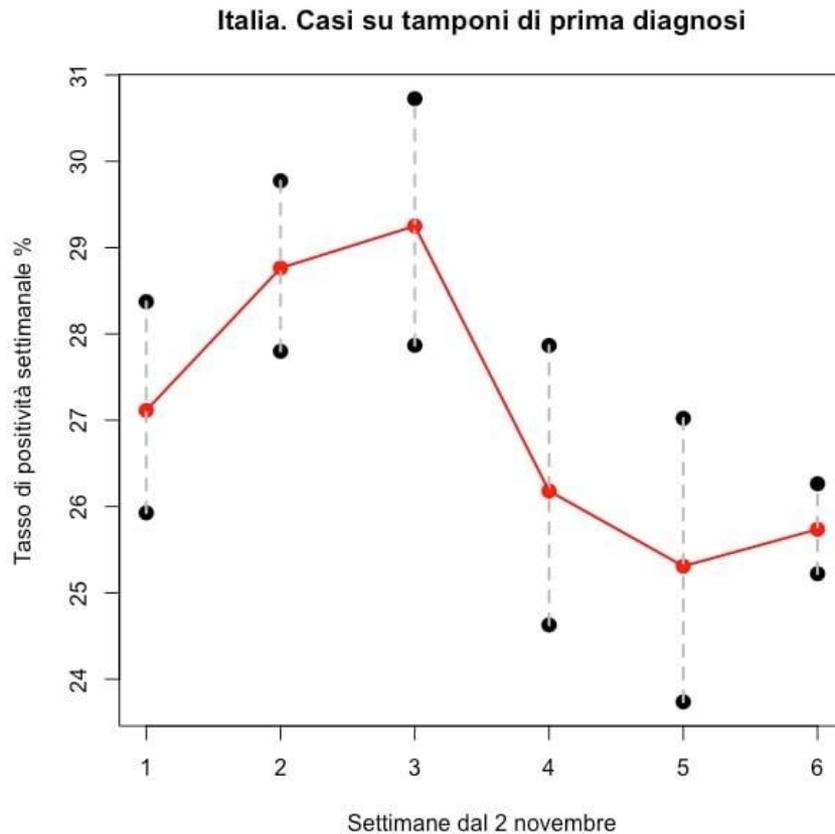
Incertezza

**«L'incertezza domina ovunque.
Tutta la nostra vita è immersa nell'incertezza;
nulla – all'infuori di ciò – si può affermare con certezza.»**

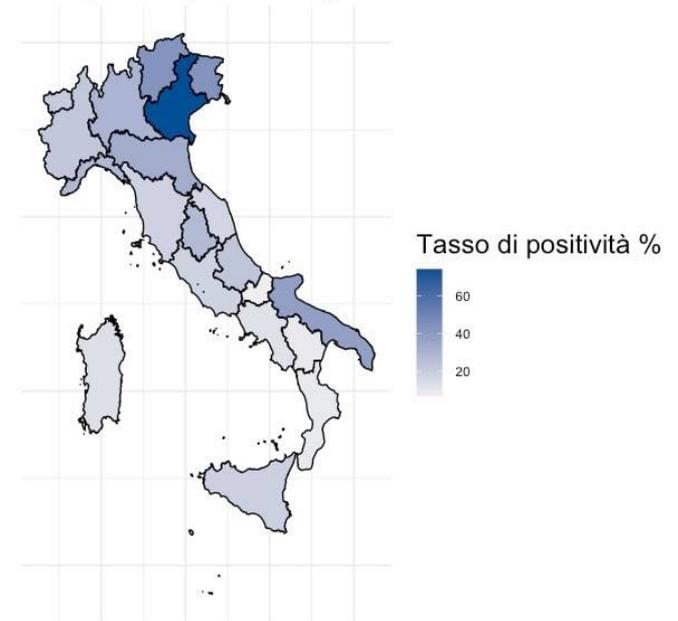
Bruno de Finetti

Monitoraggio del tasso di positività

Stime puntuali e stime intervallari

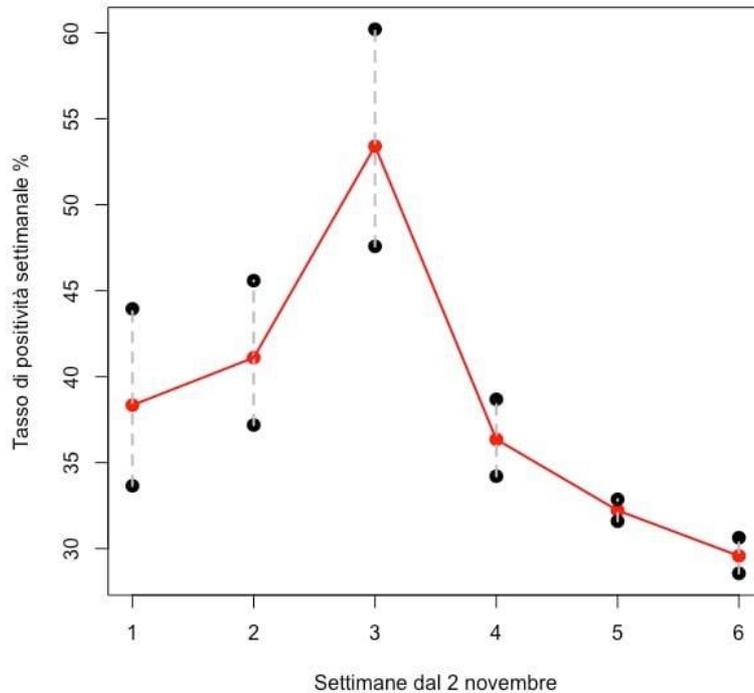


Tamponi di prima diagnosi 07.12 - 13.12

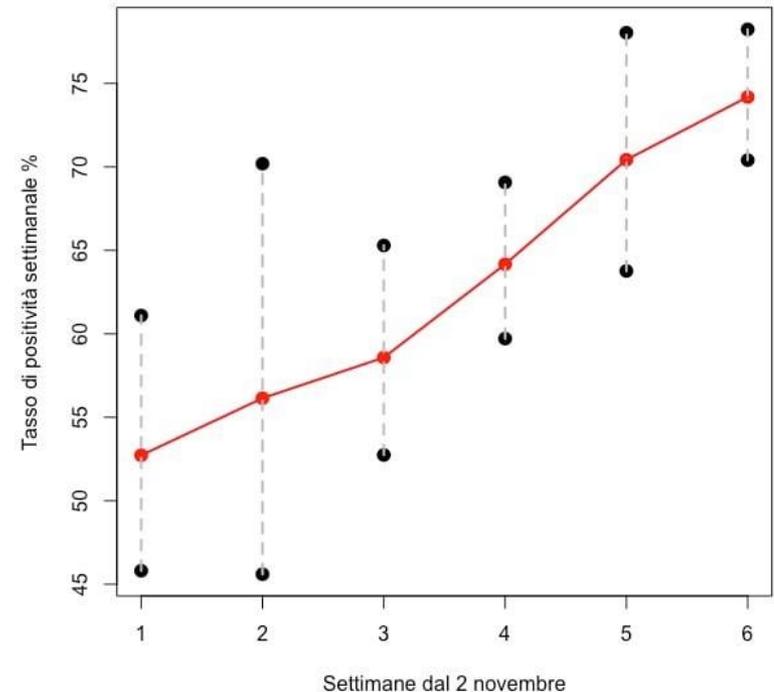


Monitoraggio del tasso di positività

Lombardia. Casi su tamponi di prima diagnosi

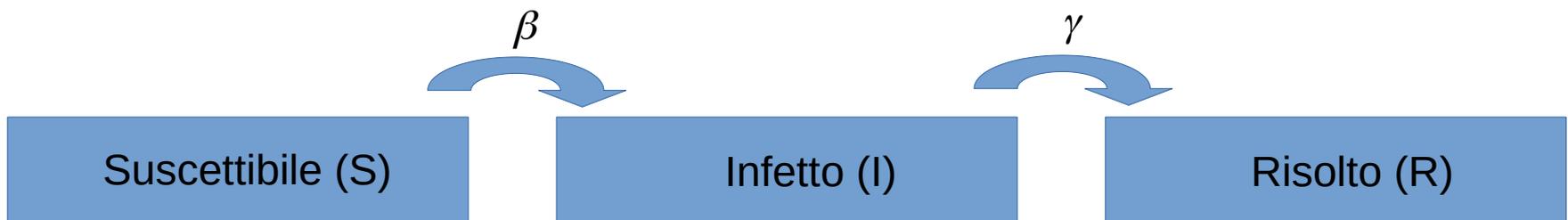


Veneto. Casi su tamponi di prima diagnosi



Monitoraggio dell'indice R_t

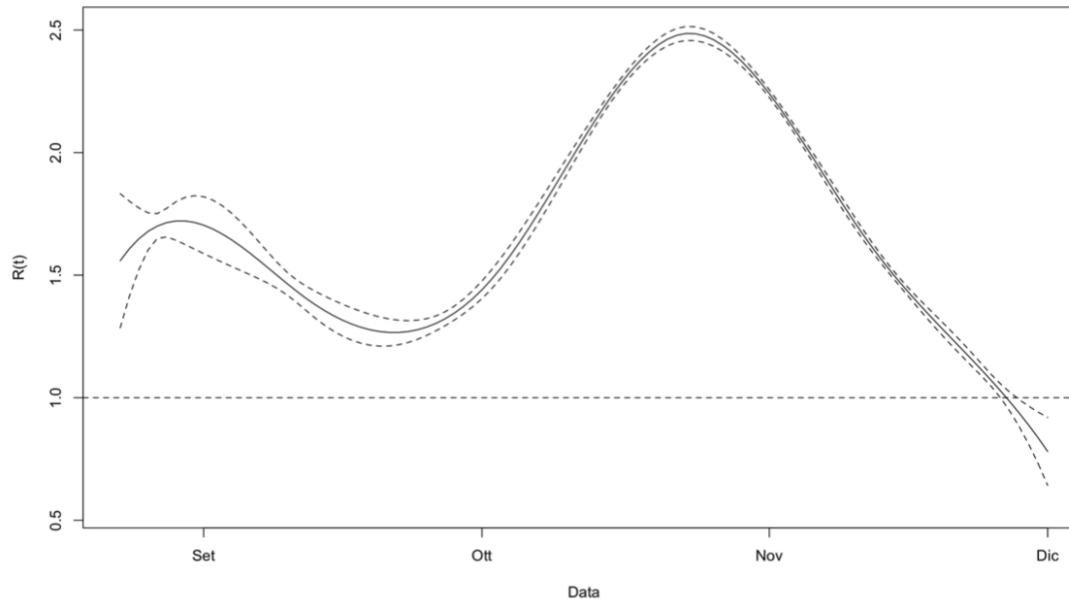
- L'indice R_t è il numero di riproduzione dell'infezione, che ci dice “quante persone in media possono essere contagiate da un singolo infetto e in un certo periodo di tempo”.
- Ad esempio, $R_t = 2$ indica che ogni infetto, in un determinato tempo, può contagiare due persone che a loro volta ne possono contagiare altre due a testa nel periodo successivo. Quando $R_t < 1$ l'epidemia è sotto controllo.
- Modello SIR:



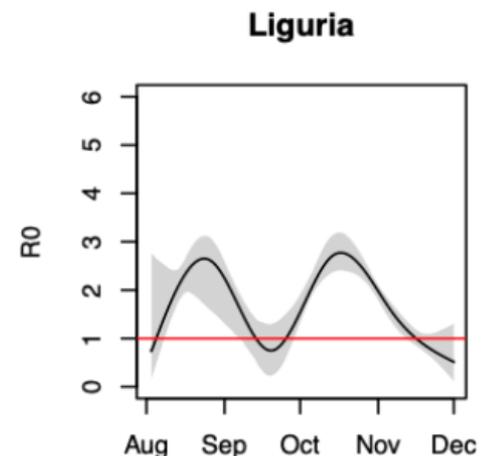
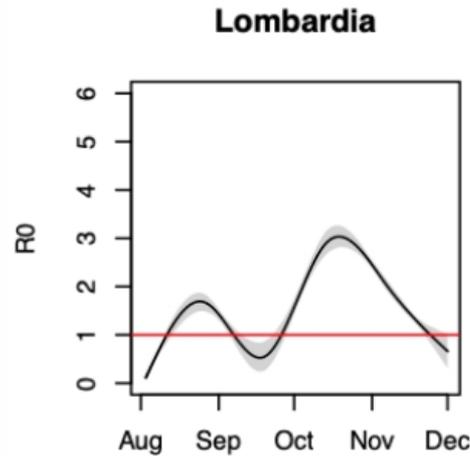
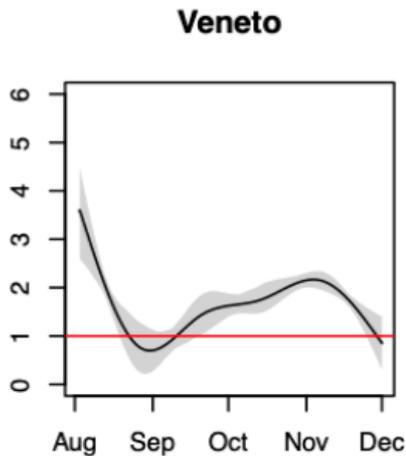
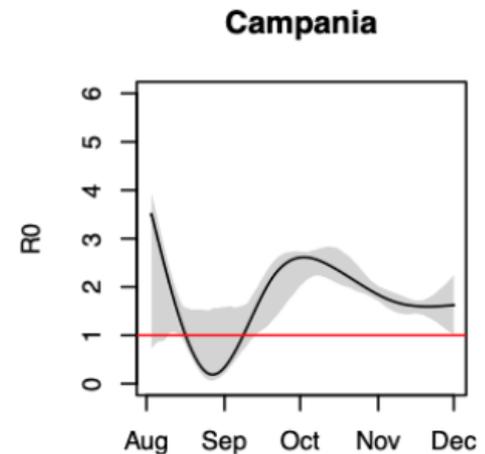
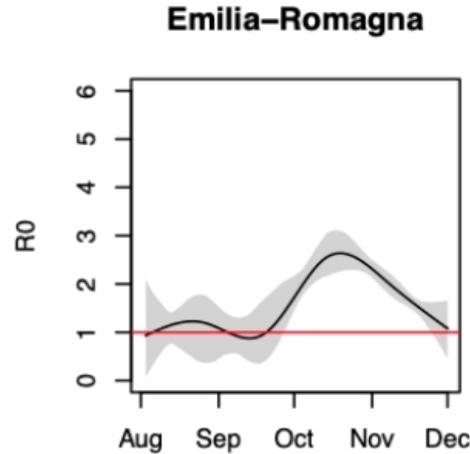
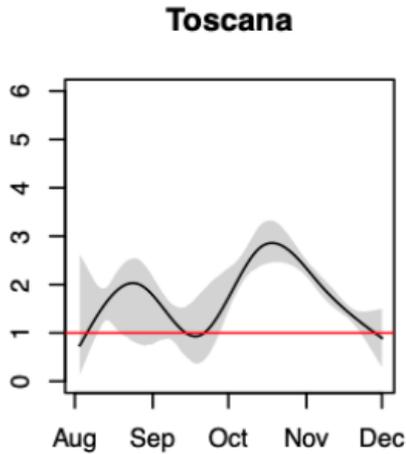
- γ è il tasso di guarigione e β è l'indice di trasmissione.
- Per il modello SIR $R_t = \beta/\gamma$.

Monitoraggio dell'indice Rt

- Il tasso di guarigione γ può essere fissato da precedenti pubblicazioni in $1/18$ (Wang *et al.*, 2020), la malattia dura in media 18 giorni.
- Quindi l'indice Rt varia nel tempo... si deve usare un approccio dinamico, dove l'indice Rt possa assumere diversi valori nel tempo. Pertanto nel modello SIR si deve stimare l'indice di trasmissione $\beta(t)$.



Monitoraggio dell'indice Rt



dai dati ai modelli statistici

«Tutti i modelli sono sbagliati, ma alcuni sono utili.»

George E.P. Box

dai dati ai modelli statistici

Cumulative confirmed COVID-19 deaths

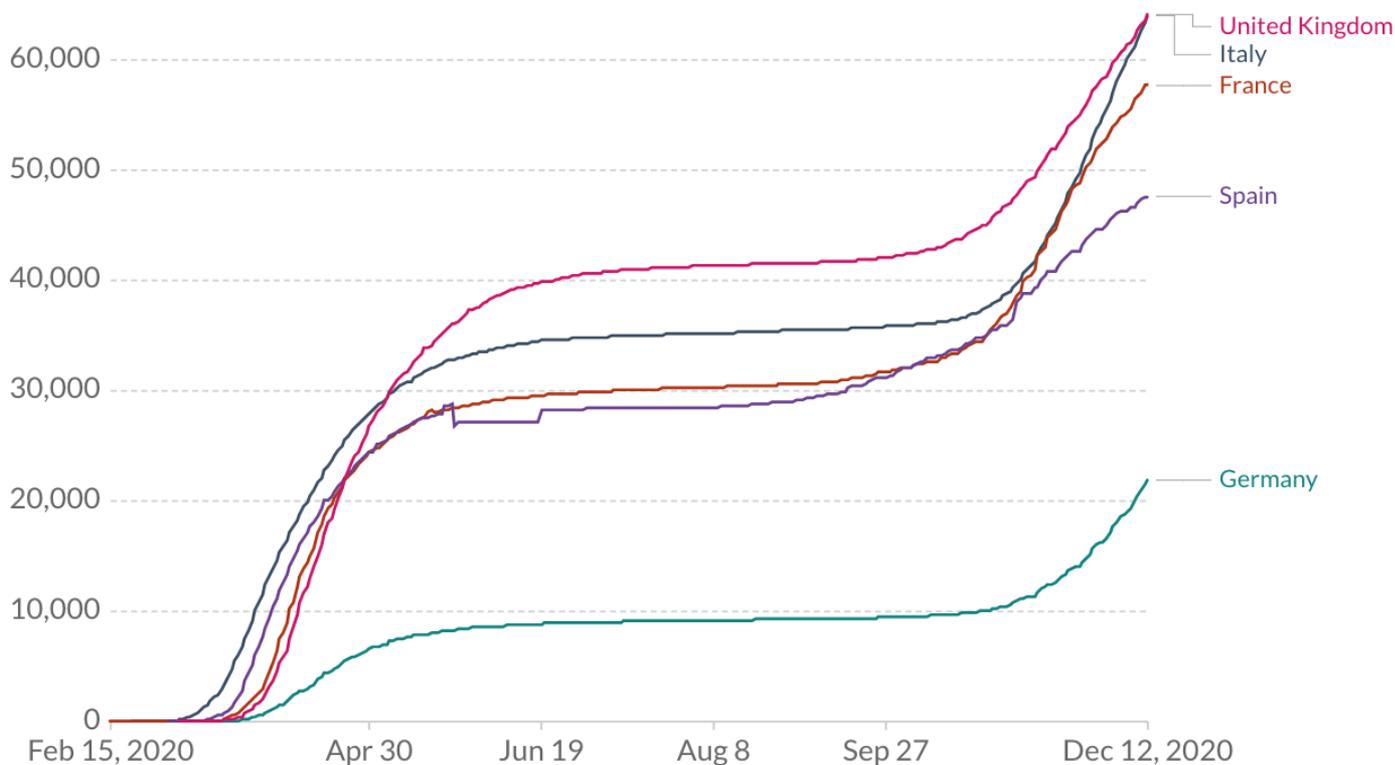
Limited testing and challenges in the attribution of the cause of death means that the number of confirmed deaths may not be an accurate count of the true number of deaths from COVID-19.

Our World
in Data

LINEAR

LOG

+ Add country



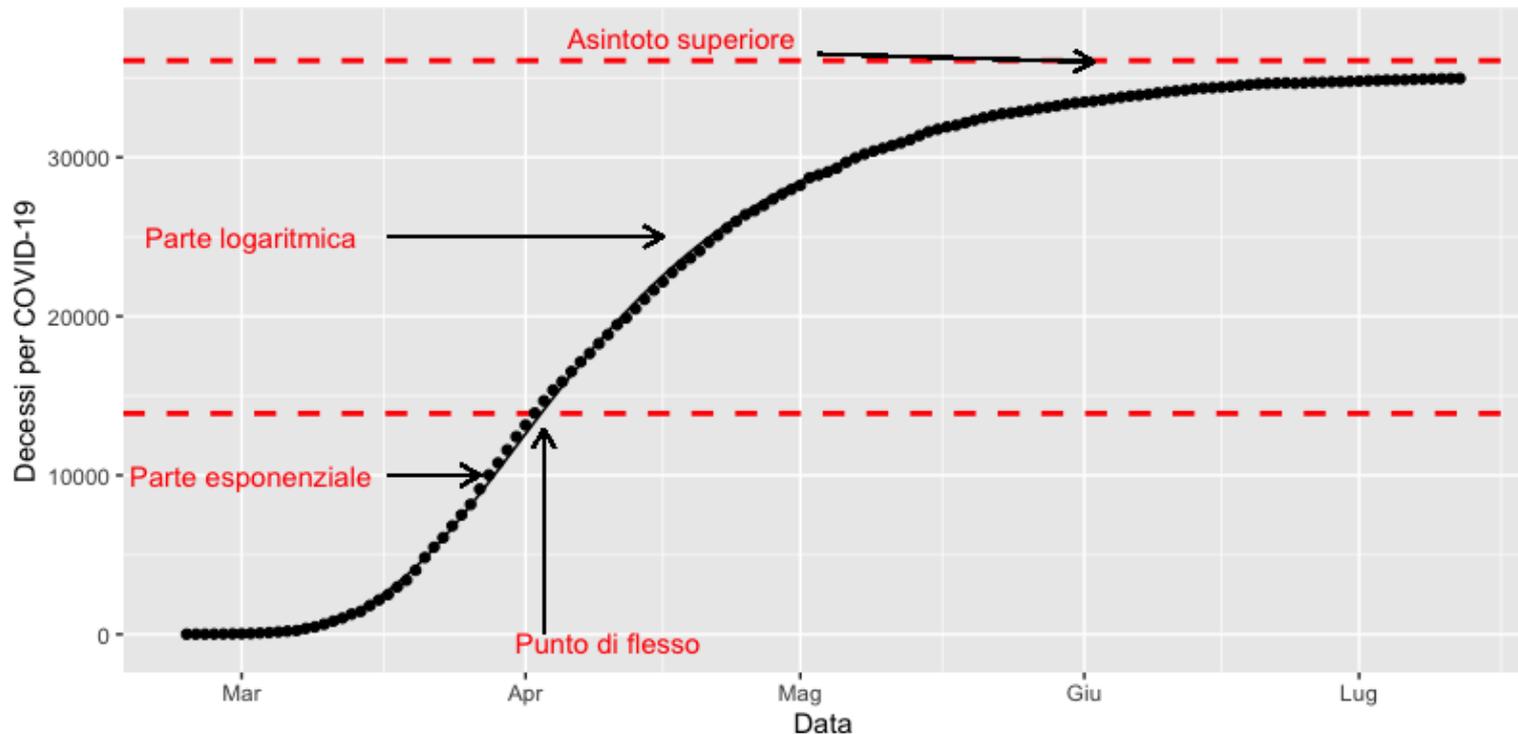
Source: Johns Hopkins University CSSE COVID-19 Data - Last updated 13 December, 06:06 (London time)

CC BY

▶ Feb 15, 2020 Dec 12, 2020

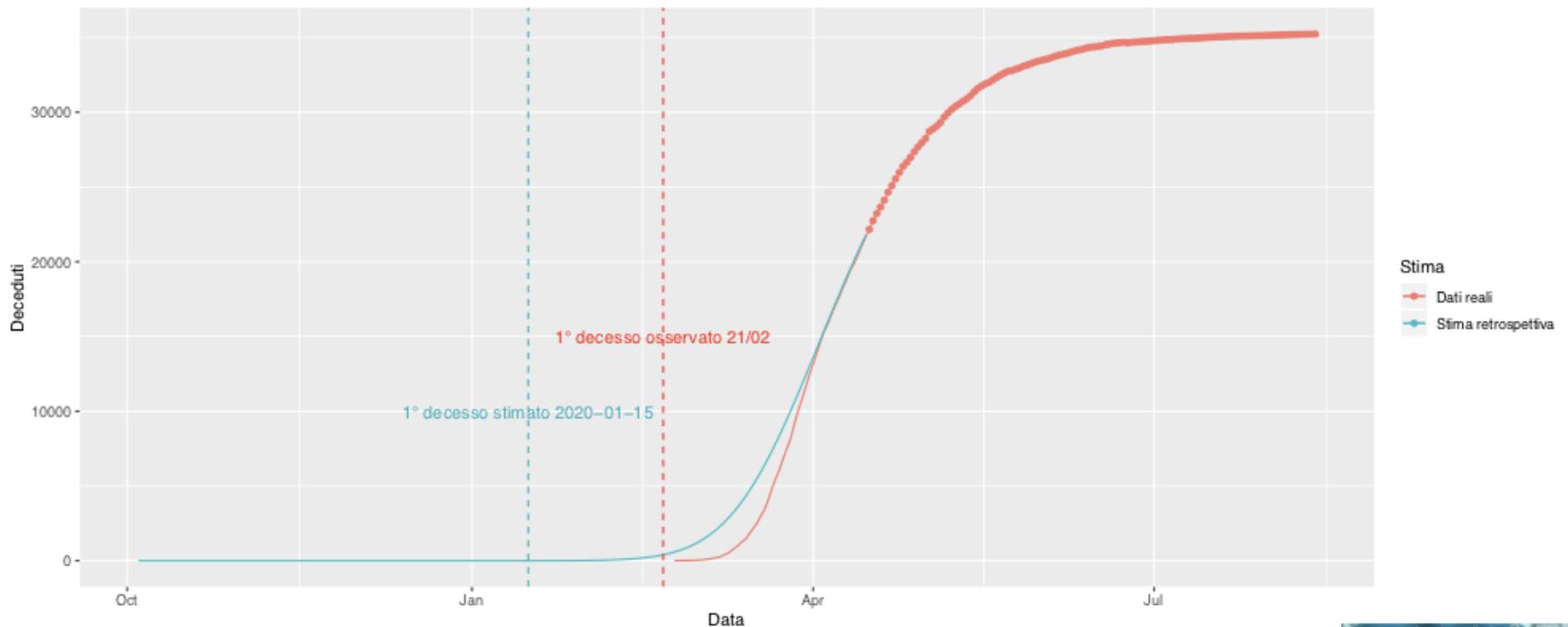
Modello non-lineare per i dati cumulati

Nel modello per il numero cumulato dei decessi (Y) giornalieri si considera una funzione $f(X;c,d,e,b)$ non lineare rispetto alla variabile $X =$ tempo (in giorni): la funzione è usualmente caratterizzata da un asintoto orizzontale inferiore c , da un asintoto orizzontale superiore d e da un punto di flesso e . Infine, b rappresenta la pendenza della curva.



uno sguardo al passato – ritorno al futuro

- ▶ Utilizzando i modelli non-lineari proviamo a rispondere a una domanda che ha interessato la comunità scientifica: a quando risale il primo decesso COVID-19 in Italia?
- ▶ Abbiamo considerato solamente i dati dei decessi dal 15 Aprile e stimato a ritroso la curva con un modello non-lineare.



- ▶ Si stima il primo decesso già a metà gennaio 2020.
- ▶ Questo risultato è compatibile con altri studi in corso.



una nuova ondata

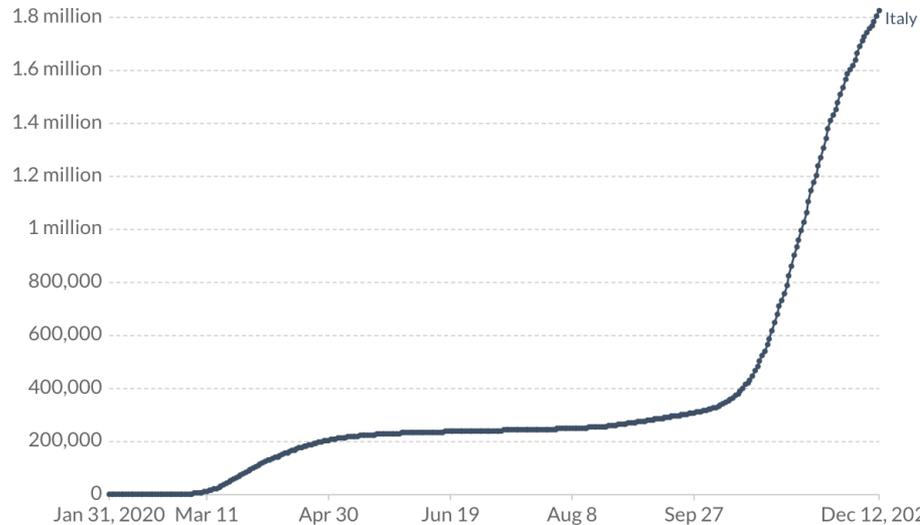
<<È molto difficile fare previsioni, specialmente sul futuro.>>
(Vecchio proverbio danese, spesso attribuito a Niels Bohr)

Cumulative confirmed COVID-19 cases

The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.

Our World
in Data

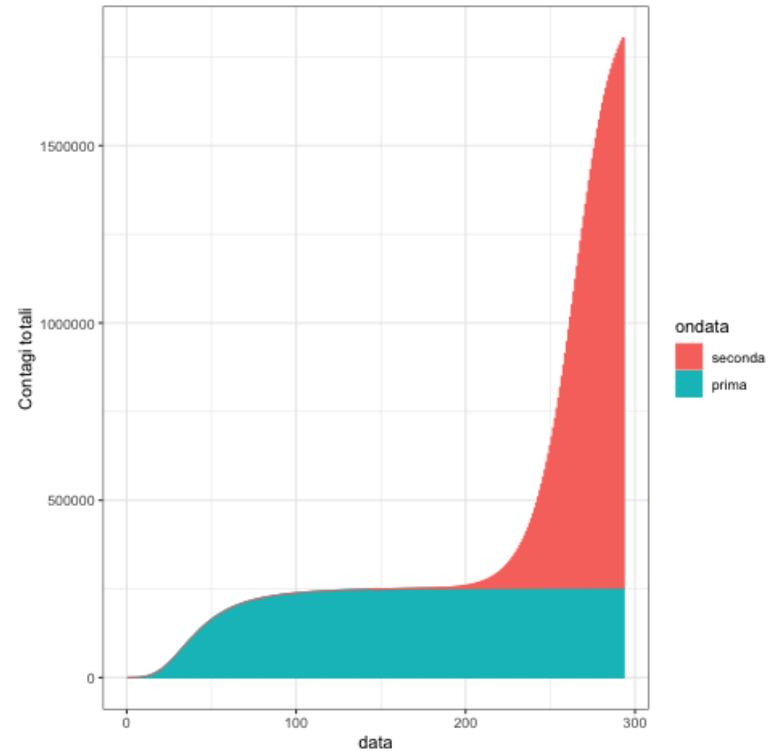
LINEAR LOG + Add country



Source: Johns Hopkins University CSSE COVID-19 Data - Last updated 13 December, 06:06 (London time)

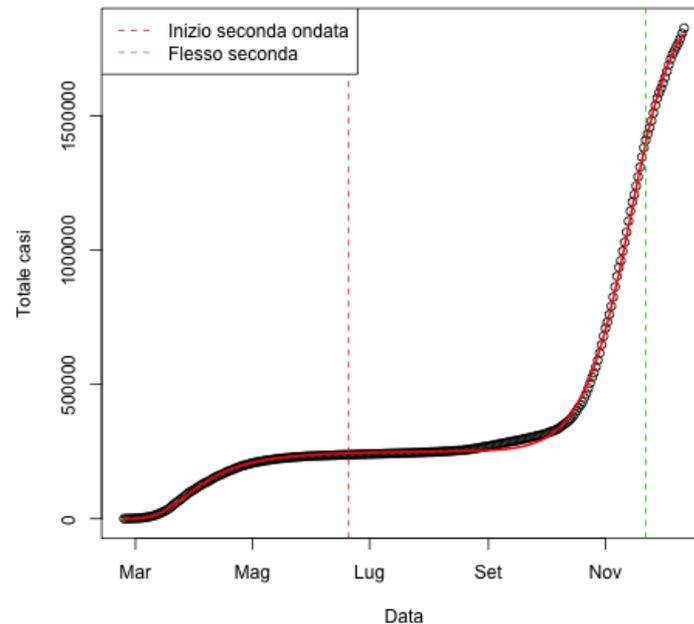
CC BY

▶ Jan 31, 2020 ○ Dec 12, 2020



Una nuova ondata – Mistura di modelli non-lineari

- ▶ Siamo nella seconda ondata di contagi.
- ▶ Utilizzo di una “fusione” di due modelli non-lineari (detta in gergo “mistura”) per trarre informazioni dai dati osservati.



Il modello fornisce alcune indicazioni sulla seconda ondata:

- ▶ È iniziata a fine giugno
- ▶ I casi sono più diluiti nel tempo
- ▶ Abbiamo da poco concluso la fase esponenziale (circa 22.11)
- ▶ Consistente numero di casi totali



un esempio di cosa si può fare con dati clinici

«Da un piccolo campione, possiamo giudicare l'intero pezzo.»

Miguel de Cervantes

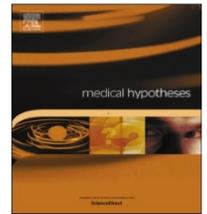
Cosa si può fare con dati clinici?



Contents lists available at [ScienceDirect](#)

Medical Hypotheses

journal homepage: www.elsevier.com/locate/mehy



Letter to Editors

Smoking and COVID-19, the paradox to discover: An Italian retrospective, observational study in hospitalized and non-hospitalized patients

ARTICLE INFO

Keywords

Smoking
SARS-CoV2
COVID-19 disease

Carlo Lombardi^{a*}, Elena Roca^b, Laura Ventura^c, Marcello Cottini^d
^a COVID-19 Unit & Departmental Unit of Pneumology & Allergology,
Istituto Ospedaliero Fondazione Poliambulanza, Brescia, Italy
^b COVID-19 Unit & Neurosurgery Departmental Unit, Istituto Ospedaliero
Fondazione Poliambulanza, Brescia, Italy
^c Department of Statistical Sciences, University of Padova, Padova, Italy
^d Allergy and Pneumology Outpatient Clinic, Bergamo, Italy

STUDIO RETROSPETTIVO SUL FUMO E COVID-19

	DECEDUTO	GUARITO	STATO DI SALUTE (variabile risposta)
FUMO	13	26	39
NON FUMO	50	96	146
	63	122	185

FATTORE
DI
RISCHIO

Proporzione di DECEDUTI tra chi fuma: $13/39 = 0.33$

Proporzione di GUARITI tra chi fuma: $26/39 = 0.67$

STUDIO RETROSPETTIVO SUL FUMO E COVID-19

	DECEDUTO	GUARITO	
FUMO	a = 13	b = 26	a+b = 39
NON FUMO	c = 50	d = 96	c+d = 146
	63	122	185

$$\left\{ \begin{array}{l} a/(a+b) = 13/39 = 0.33 \\ b/(a+b) = 26/39 = 0.67 \end{array} \right.$$

$$\text{Rapporto tra due proporzioni (odds)} = \frac{a/(a+b)}{b/(a+b)} = \frac{a}{b} = \frac{13}{26} = 0.5$$

Odds (quota) = rappresenta il rapporto fra la frequenza con la quale un evento si verifica (o si è verificato) e la frequenza con la quale l'evento non si verifica (o non si è verificato).

In una "scommessa" è il rapporto tra la probabilità p di successo e la probabilità $(1-p)$ di insuccesso.

STUDIO RETROSPETTIVO SUL FUMO E COVID-19

	DECEDUTO	GUARITO	
FUMO	a = 13	b = 26	a+b = 39
NON FUMO	c = 50	d = 96	c+d = 146
	63	122	185

$$\left\{ \begin{array}{l} c/(c+d) = 50/146 = 0.34 \\ d/(c+d) = 96/146 = 0.66 \end{array} \right.$$

Rapporto tra due proporzioni (*odds*) = $\frac{c/(c+d)}{d/(c+d)} = \frac{c}{d} = \frac{50}{96} = 0.52$

STUDIO RETROSPETTIVO SUL FUMO E COVID-19

	DECEDUTO	GUARITO	
FUMO	a = 13	b = 26	39
NON FUMO	c = 50	d = 96	146
	63	122	185

Rapporto tra i due *odds*: $OR = \frac{a/b}{c/d} = \frac{ad}{bc} = \frac{13 \cdot 96}{26 \cdot 50} = 0.96$
(detto anche rapporto incrociato)

La *quota* di morte tra i fumatori è 0.96 volte quello dei non fumatori

$OR = 1$ Non c'è associazione tra *Malattia* ed *Esposizione*

$OR < 1$ Associazione negativa: il fattore può proteggere dalla malattia.
La proporzione di *Malati* tra gli *Esposti* è "modesta" rispetto alla proporzione tra i *Non Esposti*

$OR > 1$ Associazione positiva: il fattore può causare la malattia.
La proporzione di *Malati* tra gli *Esposti* è "elevata" rispetto alla proporzione tra i *Non Esposti*

STUDIO RETROSPETTIVO SUL FUMO E COVID-19

Table 1

See text.

Patients	Variable	Deaths	OR	95% CI OR	P-value
Hospitalized	Smoking	No 0.280	1.01	(0.45;2.18)	0.99
		Yes 0.283			
Not-hospitalized	Smoking	No 0.310	0.74	(0.26;1.89)	0.66
		Yes 0.250			

per concludere

Insomma, i metodi statistici:

- Permettono di descrivere l'evoluzione nel tempo della pandemia.
- Permettono di monitorare l'andamento del contagio attraverso la stima di indici (stime puntuali e stime intervallari) e di modelli opportuni.
- Permettono di prevedere, almeno nel breve periodo, l'evoluzione del contagio.
- Permettono di simulare, attraverso tecniche Monte Carlo, scenari di contagio.
- Ahimè al momento, sulla base dei dati disponibili, non si è in grado di valutare le decisioni istituzionali e ex-post, in modo quantitativo e rigoroso, gli effetti delle decisioni.

Alcuni siti e riferimenti utili

- Siti:

DSS: <https://www.stat.unipd.it/come-ci-si-orienta-tra-i-dati>

Robbayes-C19: <https://www.facebook.com/Robbayes-C19-111971610458527>

StatGroup-19: <https://www.facebook.com/statgroup19>

Piattaforma: <https://achab94.shinyapps.io/covid-19/>

CovStat_it: <https://covstat.it/>

SIS-COVID 19: <http://www.sis-statistica.it/ita/20/index.php?p=9996>

- Alcune letture spettinate:

- Massarenti, Mira (2020), La pandemia dei dati. Ecco il vaccino. Mondadori.

- Ventura, Racugno (2017), Biostatistica. Casi di studio in R. Egea.

- “Robust inference for nonlinear regression models from the Tsallis score: application to Covid-19 contagion in Italy”, P. Girardi, E. Ruli, L. Ventura, L. Greco et al., *Stat*, 9:e309.

- “Quanto è diminuita l'aspettativa di vita delle persone dopo le morti da Covid-19?”, S. Mazzuco, Info Data Sole24ore, 7/11/2020.

- “The evolution of the endemic stage of the Covid-19 outbreak in Italy during Summer 2020”, P. Girardi, L. Greco, L. Ventura et al., *Significance*, 05/11/2020

- “Observed and estimated prevalence of Covid-19 in Italy: How to estimate the total cases from medical swabs data”, F. Bassi et al., *Science of the Total Environment*, 29/09/2020

Ad esempio, la suddivisione delle regioni in fasce

- Gli indicatori riguardano tre dimensioni fondamentali della pandemia da COVID-19:
 1. il processo di rilevazione dei dati attinenti al contagio (% di schede in cui è indicata la data di inizio contagio).
 2. l'entità del contagio vero e proprio e l'impatto sulle strutture sanitarie;
 3. le risorse del sistema sanitario impegnate nella rilevazione dei dati.
- 21 indicatori sono troppi: ne bastano 3 per spiegare l'impatto del virus?
- Basterebbero Rt, posti letto disponibili in ospedale e in terapia intensiva per dare l'idea della evoluzione dell'epidemia nel territorio di interesse e a comunicare ai tecnici e agli scienziati, e così pure alla popolazione, l'incidenza dell'epidemia al fine di potersi assumere le rispettive responsabilità.

(Luigi Fabbris, 4/12/2020, Il Sussidiario)

Conclusioni



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



- ▶ Avere comportamenti prudenti e rispettosi.
- ▶ Aumentare il numero di tamponi e test rapidi.
- ▶ Recupero di Immuni?

- ▶ Usare la mascherina o no? ... un ragionamento matematico e Bayes?

Sia p la probabilità di contagio a meno di 1 metro per almeno 15 minuti:

- con mascherina 3,1 %
- senza mascherina 17,4 %*

Sia q la probabilità di incontrare una persona infetta (asintomatica).
Tale probabilità dipende dal momento, diciamo: tra 1 % e 5 %.

Per calcolare la probabilità di infettarsi in n incontri a rischio, si usa:

$$PI = 1 - (1 - pq)^n$$

Quando c'è 1 % di infetti in media bastano

- 400 incontri senza mascherina
- 2300 incontri con mascherina

per avere $PI \geq 50$ %.

* The Lancet: June 01, 2020 DOI: [https://doi.org/10.1016/S0140-6736\(20\)31142-9](https://doi.org/10.1016/S0140-6736(20)31142-9)