

Dai dati al monitoraggio: cos'è l'indice di trasmissione del contagio R_t e come viene calcolato

Maria Teresa Giraudo
Università degli Studi di Torino
Dipartimento di Matematica

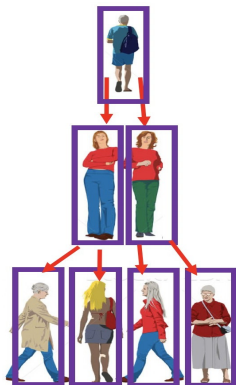
L'analisi dei dati al tempo del Coronavirus
8 marzo 2021

SOMMARIO

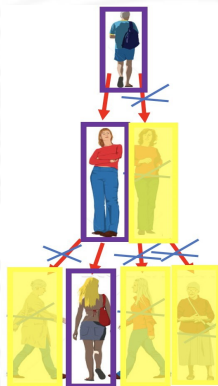
- **Definizioni**
- **A proposito di tempi**
- **Stima dell'indice di replicazione**
- **Cosa presenta l'ISS**
- **Altri possibili metodi di stima**

Definizioni

- **Numero di riproduzione di base R_0** : *numero atteso di casi secondari derivanti da un caso primario tipico in una popolazione completamente suscettibile*
- Quando l'infezione si diffonde nella popolazione è invece opportuno considerare il **numero di riproduzione effettivo R** (o meglio **R_t**):
numero medio attuale di casi secondari del contagio prodotti da un singolo infettore durante il suo periodo infettivo
- Andamento di **R_t** nel tempo: monitoraggio della progressione del contagio e feedback sulle misure di contenimento
 $R_t < 1$, meglio **$R_t \ll 1$** , per epidemia sotto controllo



*L'indice R_0 misura l'intensità di replicazione (la ragione) **in assenza** di misure di contenimento*



*L'indice R_t misura l'intensità di replicazione al tempo t **in presenza** di misure di contenimento*

Il numero di riproduzione R_t è definito nel contesto dei modelli del tipo "time-to-infection".

Metodi di stima basati su dati di sorveglianza:

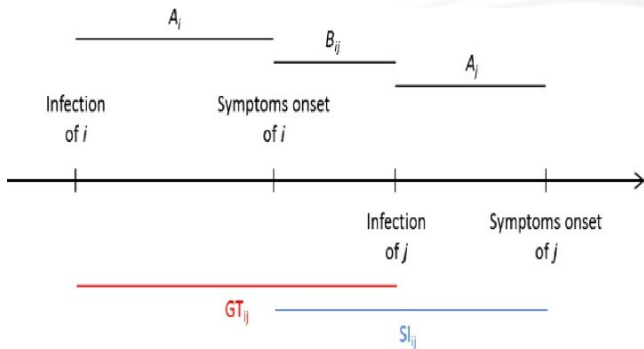
- approssimazione con crescita esponenziale dei casi nel tempo
- fitting di modelli meccanicistici a dati di incidenza (difficili da generalizzare)
- metodo basati su ricostruzione probabilistica degli alberi di trasmissione dell'epidemia
- **metodi basati su serie temporali di incidenza**
 - facilmente implementabili
 - tengono conto dell'incertezza nella distribuzione degli intervalli temporali tra casi primari e secondari

A proposito di tempi...

Intervallo di generazione τ (o **intervallo seriale** o **tempo di generazione**): tempo tra sintomi in un caso primario e sintomi in un caso secondario corrispondente

- **Tempo di generazione:** intervallo tra l'infezione di un caso primario e quella dei casi generati
- **Intervallo seriale:** intervallo tra i sintomi
I dati di sorveglianza riportano il secondo, quando presente, e non il primo
- **Periodo di incubazione:** intervallo tra infezione e sintomi; in esso l'individuo non è sintomatico

La distribuzione degli intervalli seriali viene considerata come proxy della distribuzione del profilo di infettività



- Tempo di generazione

$$GT_{ij} = A_i + B_{ij}$$

- Intervallo seriale

$$SI_{ij} = A_j + B_{ij}$$

Metodo di stima di R_t utilizzato dall'ISS

- **Metodo statistico** consolidato per valutare la trasmissibilità del virus [1]
- Implementato dalla **Fondazione Bruno Kessler** (FBK, [2]) e impiegato da gennaio 2020 per stimare la trasmissibilità in Cina e, da febbraio 2020, in Italia [3]
- Si basa su una semplice statistica ottenuta dai dati delle nuove infezioni e **non richiede ipotesi sui meccanismi di trasmissione del virus**
- Si utilizzano le curve epidemiche dei **casi sintomatici organizzati per data inizio sintomi**. La definizione di caso sintomatico nel sistema di sorveglianza integrate include tutti i casi a cui sia associata una data inizio sintomi

[1] A Cori, NM Ferguson, C Fraser, S Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol, 178(9):1505-1512, 2013

[2] <https://www.fbk.eu/en/>

[3] <https://www.epicentro.iss.it/coronavirus/open-data/rt.pdf>

Relativamente al calcolo dell' R_t , il flusso del dato su cui viene basato, e cioè quello relativo ai pazienti sintomatici, è il seguente:

Ogni volta che viene rilevato un caso clinico, viene compilato il relativo campo "stato clinico" nel quale viene indicato il grado di severità dei sintomi, da paucisintomatico a severo e, quando possibile, anche la data della loro insorgenza.

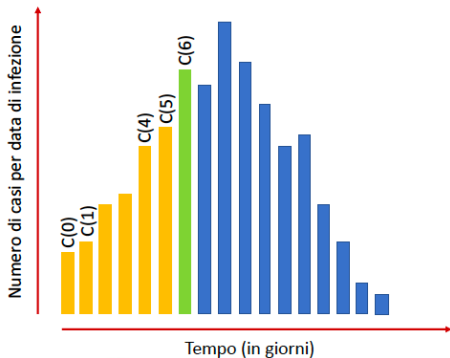
Presenza di sintomi e data di insorgenza sono le due variabili che definiscono i soggetti da considerare per il calcolo di R_t . Può capitare, anche se è raro, che alla data di inizio sintomi non sia associato uno stato clinico e in questo caso i pazienti vengono inizialmente considerati sintomatici perché si riconosce il dato fornito dalla regione come indicativo della presenza di sintomi in assenza di altre informazioni.

Nel caso in cui invece il campo "stato clinico" non venga mai compilato e così nei successivi aggiornamenti del database fino a quando sia documentata la guarigione o il decesso, il caso si considera asintomatico nonostante la presenza di una data di inizio sintomi. Non è infatti plausibile che ci sia una data di inizio sintomi di una persona senza alcun sintomo documentato fino alla guarigione o alla morte.

Infine vengono anche considerati asintomatici i rari casi in cui vi sia una data inizio sintomi ma il caso sia dichiarato nello stato clinico "asintomatico" in quanto i due dati sono in chiara contraddizione.

Dal Comunicato Stampa N° 6/2021 ISS

Esemplificazione del metodo per il calcolo di R_t (o $R(t)$)



Quanto vale per esempio $R(t)$ il giorno 6, ovvero $R(6)$?

Potenziali infettori: $C(0), C(1), \dots, C(5)$

- $C(s)$ no di casi per data di infezione a partire da un giorno fissato fino a $t-1$
- $\varphi(s)$ *profilo di infettività* s giorni dopo l'infezione

$$C(6) \approx R(6)[\varphi(6)C(0) + \varphi(5)C(1) + \varphi(4)C(2) + \varphi(3)C(3) + \varphi(2)C(4) + \varphi(1)C(5)]$$

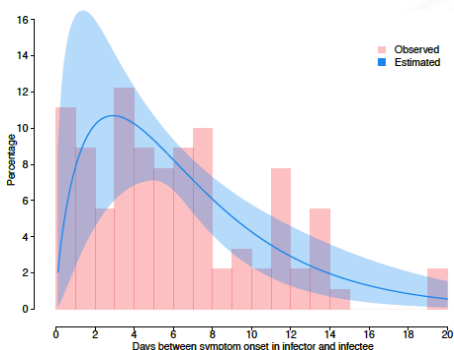
[4] https://www.epicentro.iss.it/coronavirus/open-data/rt_diapo.pdf

Con questo schema si può stimare la **distribuzione della trasmissibilità giornaliera $R(t)$** con un algoritmo Markov Chain Monte Carlo (MCMC) che utilizza la seguente distribuzione:

$$P(C(t) - I(t), R(t) \mid \sum_s \varphi(s)C(t - s))$$

- $P(k; \lambda)$ è la **densità di una distribuzione di Poisson** valutata in k di parametro (media) λ
- $C(t)$ è il numero di casi sintomatici con data di inizio sintomi al giorno t , con $t=0, \dots, T$
- $I(t)$ è il numero di **casi sintomatici importati** (da un'altra Regione o dall'estero) aventi data inizio sintomi nel giorno t , $t=0, \dots, T$
- $\varphi(t)$ è la distribuzione del tempo di generazione, approssimato con l'intervallo seriale

Distribuzione degli intervalli seriali



Distribuzione degli intervalli seriali utilizzata [5]

Dall'analisi di **dati di cluster in Lombardia all'inizio 2020**:

fit con **distribuzione Gamma** per i serial interval con parametri shape 1.87 (sd 0.26), scale 0.28 (sd 0.04)

valore medio 6.6 (giorni) e deviazione standard 4.88 (giorni)

[5] D Cereda, M Tirani, F Rovida, V Demicheli, M Ajelli, P Poletti, F Trentini, G Guzzetta, V Marziano, A Barone, M Magoni, S Deandrea, G Diurno, M Lombardo, M Faccini, A Pan, R Bruno, E Pariani, G Grasselli, A Piatti, M Gramegna, F Baldanti, A Melegaro, S Merler. The early phase of the COVID-19 outbreak in Lombardy, Italy. ArXiv:2003.09320v1, 2020

Stima della trasmissibilità puntuale

La trasmissibilità giornaliera $R(t)$ può essere influenzata da eventi particolari che la rendano artificialmente alta o bassa, quindi si considera il **valore medio dell'ultima settimana** (7 giorni, circa un tempo di generazione):

$$\underline{R_{sett}(t)} = \frac{1}{7} \sum_{s=0}^6 R(t-s)$$

Si stima anche per $R_{sett}(t)$ la distribuzione a posteriori con la relativa **media**.

E' detta stima puntuale ed è il valore utilizzato dalla Cabina di Regia per il monitoraggio settimanale:

$$\bar{R}_{sett}(t) = \frac{1}{N} \sum_{i=1}^N R_{sett, i}(t)$$

Regione.PA	Incidenza per 100.000 ab 7gg	Nuovi casi segnalati nella settimana	Trend settimanale COVID-19		Stima di Rt-puntuale (calcolato al 17/02/2021)
			Casi (Fonte ISS)	Focolai	
Abruzzo	245,92	3182	↑	↑	0.96 (CI: 0.91-1)
Basilicata	143,51	794	↑	↑	1.16 (CI: 0.95-1.37)
Calabria	62,56	1185	↑	↑	0.81 (CI: 0.72-0.92)
Campania	228,29	13040	↑	↑	0.96 (CI: 0.88-1.04)
Emilia-Romagna	342,08	15271	↑	↑	1.13 (CI: 1.11-1.16)
FVG	177,25	2138	↑	↑	0.92 (CI: 0.87-0.98)
Lazio	151,80	8737	↑	↑	0.98 (CI: 0.95-1.02)
Liguria	149,26	2276	↑	↑	0.96 (CI: 0.91-1.02)
Lombardia	254,44	25514	↑	↑	1.13 (CI: 1.11-1.15)†
Marche	265,16	4011	↑	↑	1.08 (CI: 0.98-1.19)

Regione.PA	Incidenza per 100.000 ab 7gg	Nuovi casi segnalati nella settimana	Trend settimanale COVID-19		Stima di Rt-puntuale (calcolato al 17/02/2021)
			Casi (Fonte ISS)	Focolai	
Molise	193,00	580	↓	↑	1.66 (CI: 1.27-2.07)
Piemonte	191,06	8237	↑	↑	1.15 (CI: 1.12-1.19)
PA Bolzano/Bozen	376,99	2008	↓	↓	0.75 (CI: 0.71-0.79)
PA Trento	385,02	2100	↑	↑	1.1 (CI: 1.03-1.16)
Puglia	169,73	6710	↑	↑	0.93 (CI: 0.9-0.96)
Sardegna	27,36	441	↓	↓	0.67 (CI: 0.6-0.75)
Sicilia	68,14	3322	↑	↓	0.79 (CI: 0.75-0.84)
Toscana	204,33	7545	↑	↑	1.18 (CI: 1.14-1.22)
Umbria	192,26	1673	↓	↓	0.79 (CI: 0.76-0.84)
V.d'Aosta/V.d'Aoste	60,78	76	↑	↑	1.21 (CI: 0.9-1.57)
Veneto	151,30	7382	↑	↑	1.08 (CI: 1.04-1.12)

Dal Report settimanale di monitoraggio dell'ISS del 5 marzo [6]

[6] https://www.iss.it/documents/20126/0/Monitoraggio+Fase+2_+report_nazionale_42.pdf/6724e46a-7eb5-31d8-68fc-cc5e062151a1?t=1614965943898

Stima della trasmissibilità media

E' utile anche calcolare la **variabilità dei valori medi** $\bar{R}_{sett}(t)$ per capire ad esempio se le stime

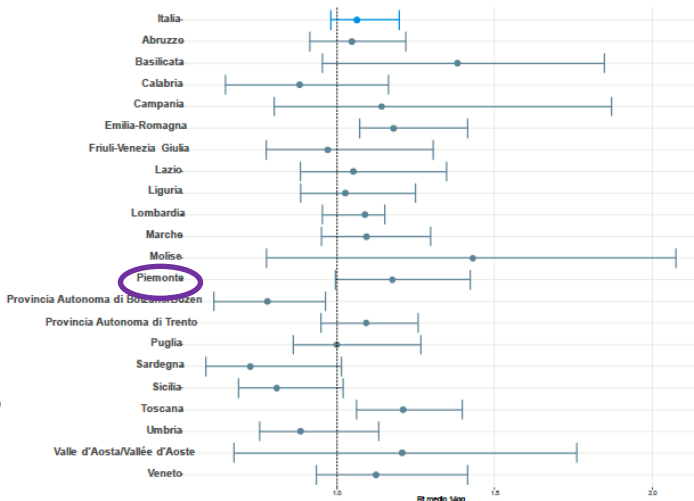
- sono **stabili nel tempo**
- sono caratterizzate da qualche **trend di crescita o decrescita**
- **fluttuano in un intervallo molto ampio** (ad esempio perchè il numero di casi $C(t)$ è molto basso)

Si considera allora il valore

$$\bar{M}(t) = \frac{1}{14} \sum_{s=0}^{13} \bar{R}_{sett}(t-s)$$

I bollettini ISS riportano oltre a $\bar{M}(t)$ il minimo e il massimo sui 14 giorni del 95% percentile delle distribuzioni di $R_{sett}(t)$.

La **Figura 9** riporta la stima del numero di riproduzione netto R_t medio nazionale a 14 giorni nel tempo in relazione alla curva epidemica.



Dal Bollettino
dell'ISS del 3
marzo

[7] https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_3-marzo-2021.pdf

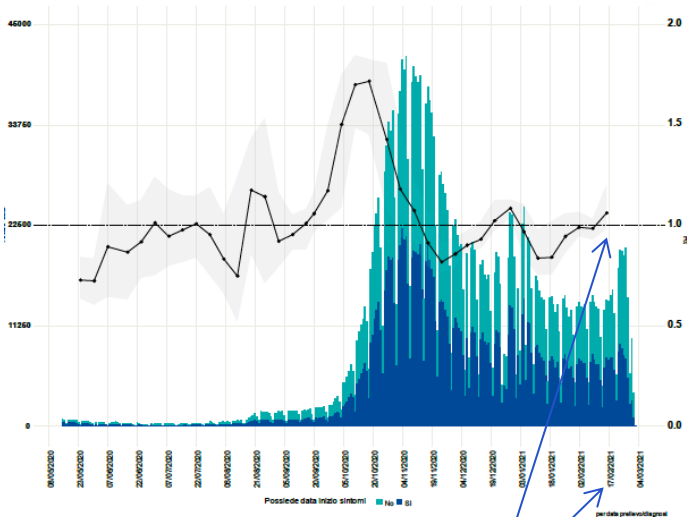


FIGURA 9 - STIME DELL' $R_{TMEIO14GG}$ E NUMERO DI CASI PER DATA PRELIEVO/ DIAGNOSI DISTINTI PER PRESENZA O ASSENZA DELLA DATA DI INIZIO SINTOMI.

Il punto (valore di $\bar{M}(t)$) è collocato nel giorno centrale dell'intervallo di 14 giorni

Algoritmo e software per il calcolo dell'indice di trasmissibilità

- FBK utilizza come codice di base un algoritmo scritto in **codice C** e disponibile all'indirizzo <https://github.com/majelli/Rt> distribuito pubblicamente
- L'8 giugno 2020 è stato distribuito a tutte le Regioni uno script in **codice R** che permette il calcolo di $R(t)$ a partire dall'algoritmo **EpiEstim** nel pacchetto omonimo <https://cran.r-project.org/web/packages/EpiEstim/index.html> disponibile pubblicamente

Dal sito dell'ISS:

<https://www.epicentro.iss.it/coronavirus/sars-cov-2-dashboard>



L'epidemiologia per la sanità pubblica
Istituto Superiore di Sanità

Indice A-Z dei contenuti

Cerca..



Coronavirus

Home | EpiCentro



Dati della Sorveglianza integrata COVID-19 in Italia

[English \(Inglese\)](#)

Per approfondire i dati contenuti nella dashboard e nei bollettini settimanali della sorveglianza scarica:

- il file xlsx "[COVID-19 ISS open data](#)" (*aggiornamento giornaliero*)
- la nota metodologica "[Stime della trasmissibilità di SARS-CoV-2 in Italia](#)" (pdf 559 kb)
- presentazione esplicativa del [metodo per l'estima degli Rt](#) (pdf 735 kb), a cura della Fondazione Bruno Kessler, Trento
- [dati nazionali e script per il calcolo dell'Rt nazionale](#) (zip 6 kb) (*aggiornamento settimanale*)

Le informazioni ed i dati relativi all'epidemia da COVID-19 vengono condivisi dall'Istituto Superiore di Sanità e periodicamente aggiornati e verificati, al fine di mantenere la loro correttezza ed attualità. L'Istituto Superiore di Sanità è, in ogni caso, sollevato da ogni responsabilità per eventuali errori, imprecisioni o inesattezze nel contenuto di tali informazioni e dati provenienti da enti terzi. In nessun caso l'Istituto Superiore di Sanità sarà ritenuto responsabile per qualsiasi danno, diretto o indiretto, derivante dall'uso improprio delle informazioni e dei dati condivisi. L'Istituto Superiore di Sanità è, inoltre, svincolato da ogni responsabilità riguardo all'interpretazione delle informazioni e dei dati da parte di terzi, così come anche, da qualsivoglia utilizzo improprio e/o illecito degli stessi.

[Dati ultimi 30 giorni](#)

[Dati cumulativi](#)



Altri metodi disponibili per la stima della trasmissibilità: 1.

$R^*(t)$

Strumento sperimentale aggiornato settimanalmente per interpretare i dati Covid-19, con particolare riguardo alla stima dell'impatto dell'epidemia da Sars-CoV-2 in termini di proiezione a breve termine del numero di infetti attivi nelle Regioni e nelle singole Province.

Si fonda sul calcolo di un indice riproduttivo semplificato denominato **$R^*(t)$, non confrontabile con l' $R(t)$ pubblicato dall'ISS**, poichè viene calcolato con differenti algoritmi matematici e su una diversa base dati, costituita da quanto pubblicato giornalmente dalla Protezione Civile.

Il lavoro è frutto della collaborazione tra Agenas e l'Osservatorio dei dati epidemiologici del Dipartimento di Fisica - Università di Trento.

<https://www.agenas.gov.it/covid19/web/index.php?r=site%2Fprevisionecontagi&q=ITA>

Altri metodi disponibili per la stima della trasmissibilità: 2.

RD_t

L' RD_t non prende in considerazione il momento del contagio bensì la data della diagnosi di positività e per questo motivo viene chiamato **indice di replicazione diagnostica** in quanto descrive l'intensità dello sviluppo della frequenza di casi riconosciuti come positivi nella popolazione sia a livello nazionale che regionale.

Si calcola come rapporto tra una media mobile centrata non pesata di n elementi al tempo t e una media mobile centrata non pesata di n elementi al tempo $t-s$, con s solitamente pari a 4, 5, 6, o 7 giorni. **Aggiornato giornalmente**, consente di valutare l'andamento dell'epidemia in tempo pressochè reale.

Il lavoro (nella piattaforma **MADE**) è frutto della collaborazione tra l'Associazione Italiana di Epidemiologia, la rivista Epidemiologia e Prevenzione e Scienza in Rete.

<https://www.ep.epiprev.it/apps/made.php>

GRAZIE DELL'ATTENZIONE E... AD MELIORA!

RD(t) Indice di Replicazione Diagnostica a lag 7 giorni (medie mobili a 7 giorni)

