



I FONDAMENTALI

*Le parole dell'epidemia spiegate dai docenti
di Scienze Statistiche di Padova*

COVID-19
dai dati ai modelli



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



DIPARTIMENTO
DI SCIENZE
STATISTICHE

Prof.ssa
LAURA VENTURA

docente di
STATISTICA MEDICA
e di
MODELLI STATISTICI



I FONDAMENTALI

*Le parole dell'epidemia spiegate dai docenti
di Scienze Statistiche di Padova*

COVID-19
dai dati ai modelli



UNIVERSITÀ
DELLI STUDI
DI PADOVA



I dati della protezione civile

Dati comunicati quotidianamente dal Dipartimento della Protezione Civile e disponibili nel sito <https://github.com/pcm-dpc/COVID-19>.

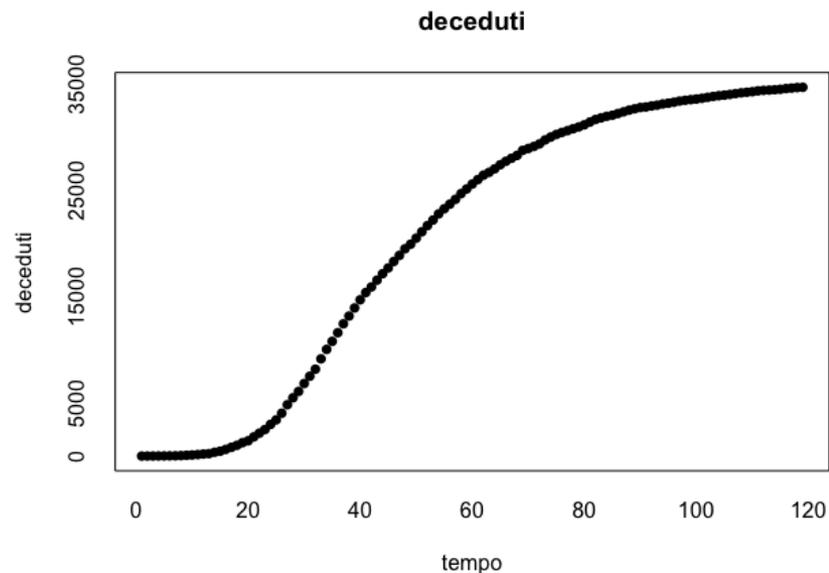
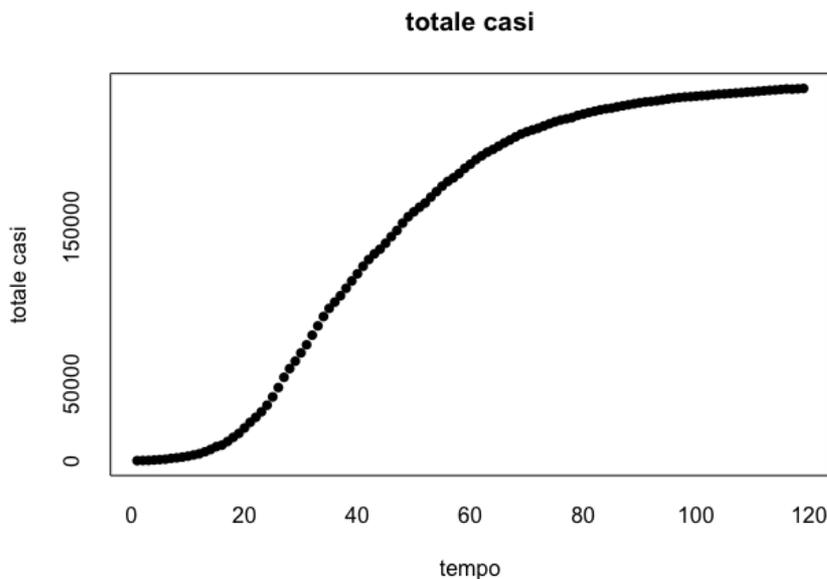


24 febbraio 2020 – giorno 1

24 giugno 2020 – giorno 121

Cosa si osserva dai grafici dei dati cumulati?

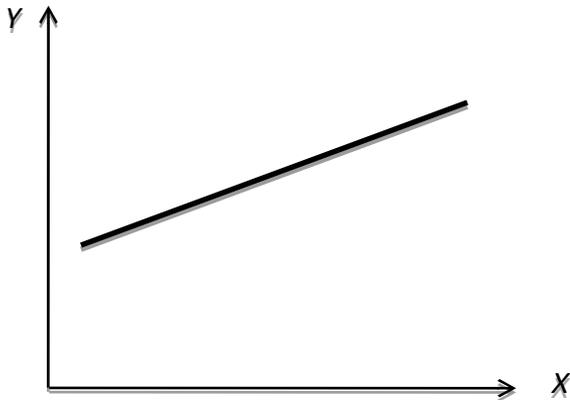
(somma dei dati giorno per giorno)



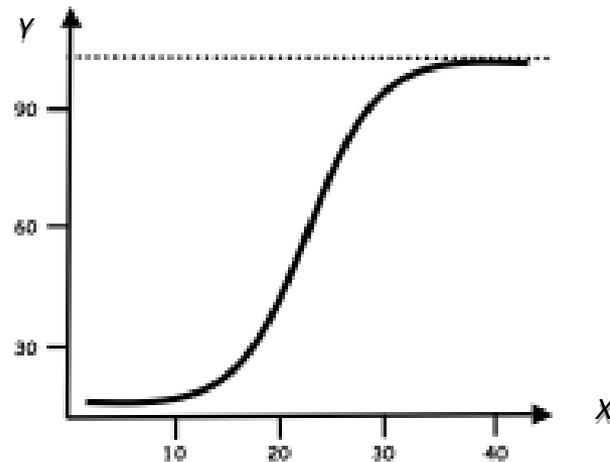
- I dati all'inizio crescono lentamente, poi si registra un'accelerazione e la velocità con la quale i conteggi crescono aumenta; intorno al 40/45-esimo giorno si nota una flessione dei casi e dei deceduti e la velocità dei conteggi diminuisce sempre più.
- Se i conteggi non aumentano si è raggiunto l'estremo superiore dei casi/deceduti totali e la diffusione epidemica si può ritenere esaurita.
- L'estremo superiore dei casi totali rappresenta l'estensione massima del contagio, cioè il numero massimo di individui risultati positivi al contagio (inclusi guariti e deceduti).
- Così come l'estremo superiore dei deceduti totali rappresenta il numero complessivo di morti dovute al COVID-19.

Dal modello lineare al modello non-lineare

- Nella regressione lineare i punti del diagramma di dispersione tendono a disporsi secondo una linea retta: $f(X;a,b) = a + b X$.
- Nel modello non-lineare la relazione tra X e Y è modellabile tramite una funzione crescente (o decrescente), caratterizzata da un andamento che all'inizio è "quasi" esponenziale, per poi appiattirsi tendendo a una retta orizzontale.



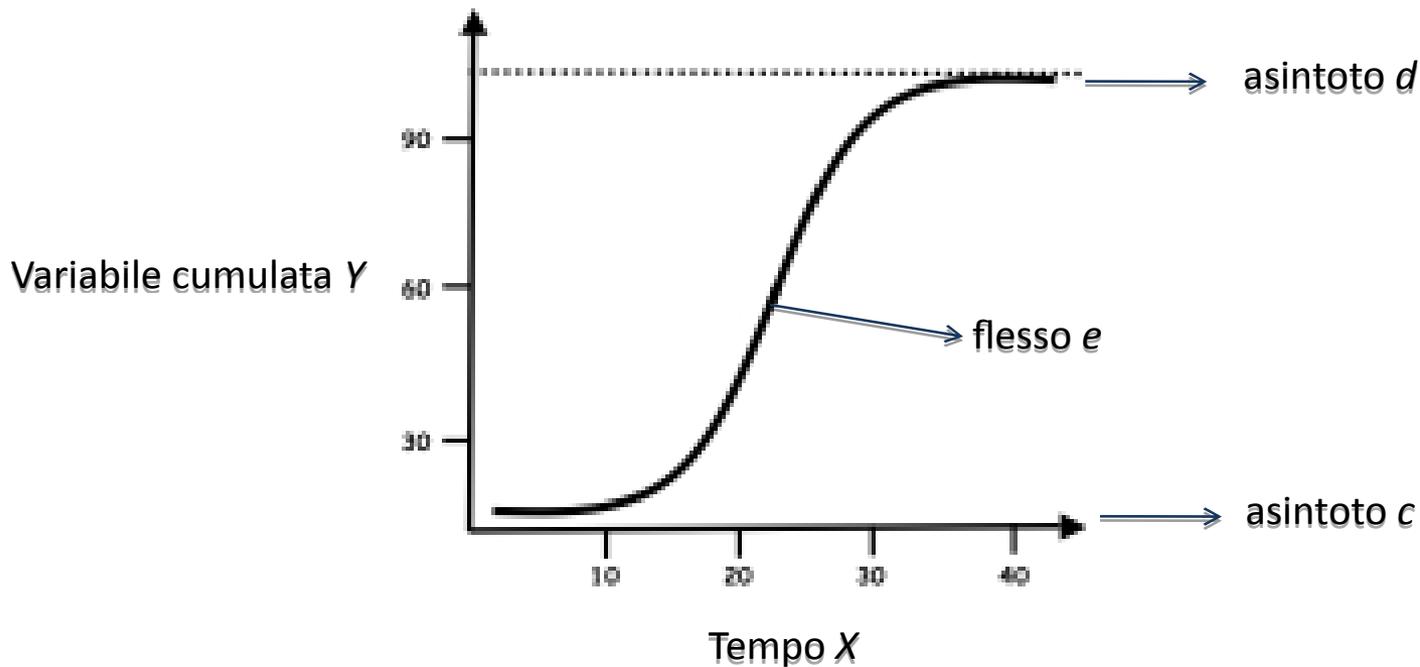
modello lineare



modello non-lineare

Modello non-lineare per i dati cumulati

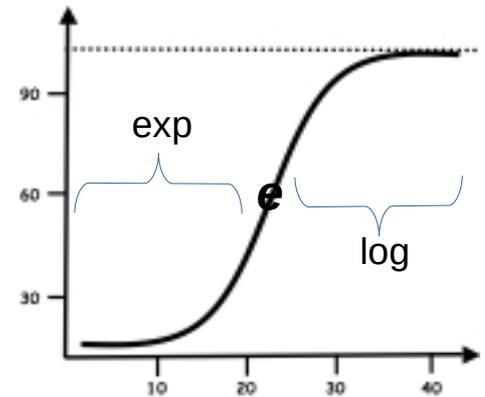
- Nel modello per il numero cumulato dei casi/deceduti giornalieri si considera una funzione $f(X;c,d,e,b)$ non lineare rispetto alla variabile $X =$ tempo (giorni): la funzione è usualmente caratterizzata da un asintoto orizzontale inferiore c , da un asintoto orizzontale superiore d e da un punto di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



Modello log-logistico

- Un esempio particolarmente utile per descrivere l'andamento del processo epidemico è la funzione di crescita *log-logistica*:

$$f(x; b, c, d, e) = c + \frac{d - c}{1 + \exp(b(\log(x) - e))}$$



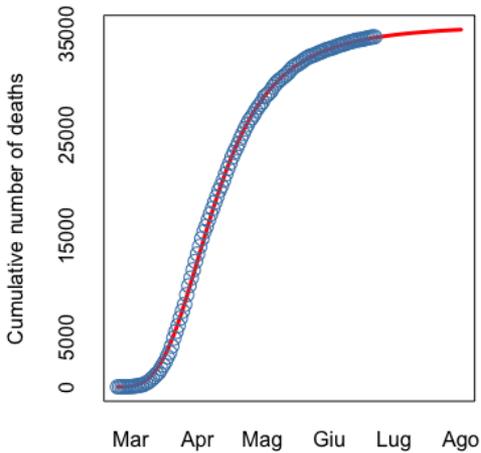
- È una curva a forma di S allungata, formata da due rami: il primo esponenziale (exp) e il secondo logaritmico (log) uniti in corrispondenza del punto e .
- La stima dei parametri della funzione $f(x; b, c, d, e)$ può avvenire con il metodo dei minimi quadrati (come per la retta di regressione), ossia minimizzando

$$distanza = \sum_{i=1}^n (y_i - f(x_i; b, c, d, e))^2$$

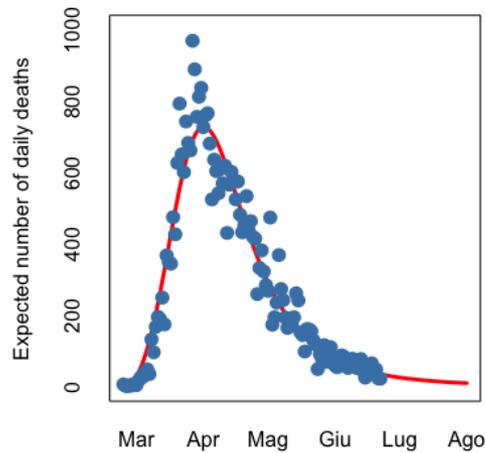
o con altre tecniche statistiche più raffinate.

Deceduti e totali positivi per COVID-19 in Italia al 24 giugno 2020

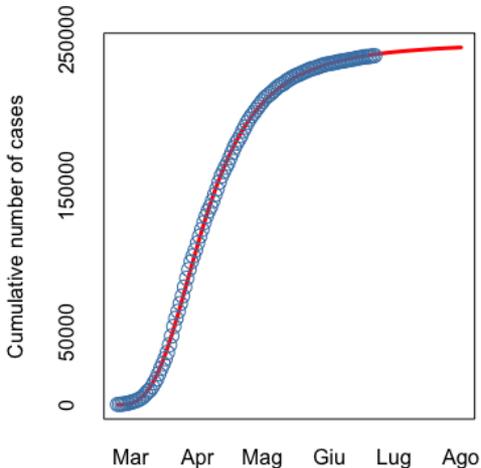
Decessi cumulati



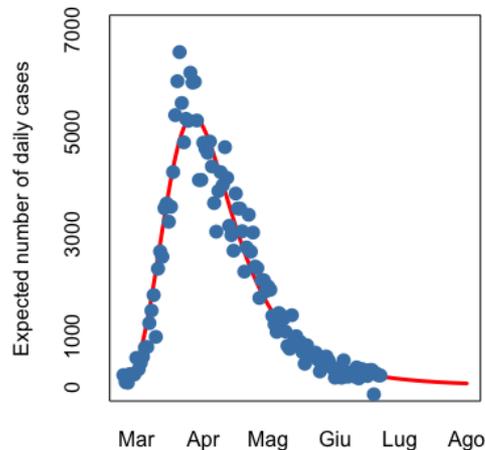
Decessi giornalieri



Casi cumulati



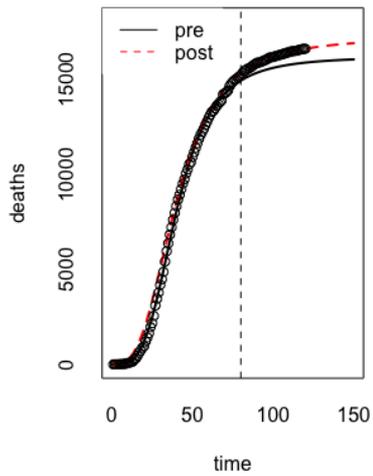
Casi giornalieri



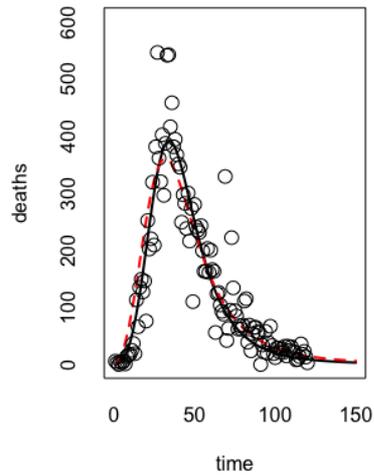
- I punti blu sono i dati osservati, le linee rosse sono i modelli non-lineari stimati.
- I grafici a sinistra si riferiscono ai deceduti/casi giornalieri cumulati.
- I grafici a destra mostrano i deceduti/casi giornalieri: si noti il "picco" (*moda* della distribuzione) e il conseguente calo. Dalla asimmetria della curva si evince che la decrescita è molto più lenta della crescita.

E dopo la riapertura del 3 giugno 2020?

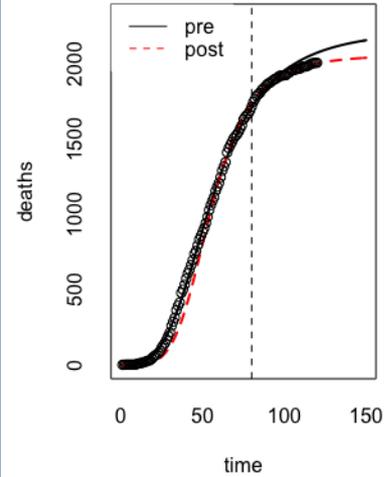
Decessi cumulati Lombardia



Decessi giornalieri Lombardia



Decessi cumulati Veneto



Decessi giornalieri Veneto

