

Appendice per Rubrica su “I dati sul Covid-19 in Italia a livello regionale”

di Mauro Maltagliati

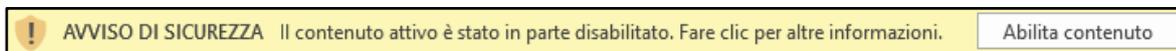
Appendice (aggiornata al 2 maggio 2020)

Preparazione e funzionamento del Data Base e di programmi di elaborazione flessibili

I lettori sono autorizzati ad usare il data base per la elaborazione dei grafici per le serie ivi riportate. Comunque, chiediamo a chi lo utilizza sia per costruire le serie che per le elaborazioni per produrre lavori (articoli, ecc.) di citare sia il data base che il Manuale.

Manuale per l’uso del file autoDB.xlsm

Il file “autoDB.xlsm” è un file di Excel che contiene delle macro (da cui l’ultima lettera dell’estensione del file) scritte in linguaggio Visual Basic, Applications Edition (VBA). Le macro in Excel sono disabilitate per impostazione predefinita, dato che sono suscettibili di essere utilizzate per la consegna di codice dannoso. Per poter usare tutte le macro del file si deve cliccare su “Abilita contenuto” alla sua apertura. In particolare, in questo file, le macro sono necessarie per avviare la macro di aggiornamento automatico del database, come descritto più avanti.



Per eventuali domande sull’utilizzo, ma anche per segnalare eventuali “bugs”, contattare mauro.maltagliati@unifi.it.

Se il file viene utilizzato per elaborazioni , si prega di citare la

All’apertura del file, come al solito in Excel, in basso si trovano le “linguette-scheda” che consentono di navigare tra i fogli elettronici.

In questo file ce ne sono 13, raggruppate per colore a seconda della tipologia.



Aggiornamento database (linguette azzurre)



Le Linguette azzurre riguardano il database vero e proprio, che viene aggiornato cliccando sul pulsante che si trova nel foglio “AGGIORNA”.

AGGIORNA DATABASE

Cliccando sul pulsante si aziona una macro che carica il file dalla pagina web:

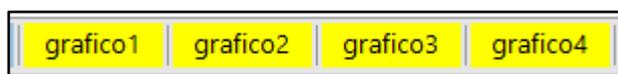
<https://github.com/pcm-dpc/COVID-19/blob/master/dati-regioni/dpc-covid19-ita-regioni.csv>.

Solitamente i nuovi dati sono presenti dalle ore 19 di ogni giorno. I dati vengono poi riordinati e immessi nel foglio "DBRegioni". Come si può notare, la tabella non inizia dalla riga 1 come ci si aspetterebbe. Per trovarla si deve scorrere il foglio fino alla riga 50-70 circa (a seconda dei computer sui quali abbiamo effettuato la prova). Questa "stranezza" è dovuta al fatto che l'operazione di riordinamento della tabella provocava errori nell'esecuzione della macro su alcuni computer su cui il file è stato caricato. In futuro contiamo di risolvere questa piccola imperfezione. Sui computer Mac l'aggiornamento non ha funzionato.

Comunque, una volta che sul proprio computer si individua la riga in cui la tabella dei dati regionali inizia, ad ogni aggiornamento la troveremo sempre nella stessa posizione.

Le tabelle DBItalia e DB_No_Lomb sono ricavate dalla DBRegioni e riportano i dati aggregati per l'Italia intera e per l'Italia con esclusione della Lombardia. Queste tabelle consentiranno di produrre grafici ed elaborazioni per questi due raggruppamenti territoriali (in aggiunta alla possibilità di avere i grafici per le singole regioni).

Grafici semplici (linguette gialle)



Su questi fogli non c'è molto da dire, sono grafici il cui contenuto informativo è facilmente comprensibile. L'unica opzione per l'utente è quella di scegliere la circoscrizione geografica tra le 23 opzionabili: le 21 regioni (le province del Trentino sono separate nel database originale), l'Italia e infine l'Italia con esclusione della Lombardia. La selezione avviene cliccando sulla voce desiderata nel menu a tendina.



In questi fogli l'unica ulteriore operazione che si può compiere è copiare i grafici per incollarli su un'altra applicazione. Se si intende trasferirli in Word, si consiglia (all'interno del documento di Word) di scegliere l'opzione "Incolla speciale- Immagine (Enhanced Metafile)". È un consiglio (valido in generale) che consente di trattare il grafico importato come una semplice immagine, pur perdendo il "collegamento" con Excel.

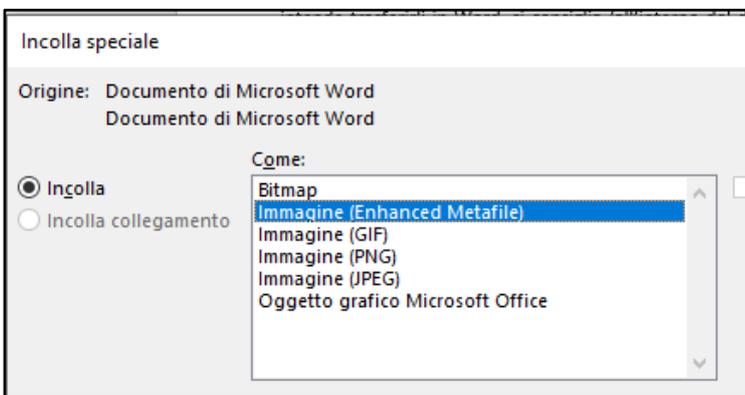
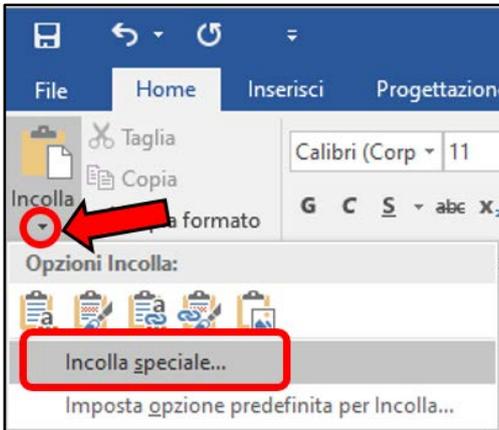


Grafico di confronto dei time plot per due diverse zone geografiche (linguetta arancione "DUE zone")



Su questo foglio è possibile mettere a confronto la stessa serie storica per due differenti aree geografiche. Le due aree sono selezionate attraverso i due menu gemelli ("SERIE1 (blu)" e "SERIE2 (arancione)") e il tipo di serie mediante i quattro menu sulla sinistra.

La combinazione "serie originaria-valori assoluti-scala naturale" dei primi tre menu rappresenta il dato originale caricato dal sito indicato in precedenza. Le altre combinazioni derivano da semplici elaborazioni. Anche per questo "foglio di lavoro" l'unica operazione a disposizione dell'utente (a parte la scelta della serie e delle zone geografiche) è poter copiare il grafico e incollarlo in un altro programma.

	SERIE1 (blu)	SERIE2 (arancione)
serie originaria serie differenziata	Abruzzo Basilicata P.A. Bolzano Calabria Campania Emilia-Romagna Friuli Venezia Giulia Lazio Liguria Lombardia Marche Molise Piemonte Puglia Sardegna Sicilia Toscana P.A. Trento Umbria Valle d'Aosta Veneto ITALIA ITA-Lomb	Abruzzo Basilicata P.A. Bolzano Calabria Campania Emilia-Romagna Friuli Venezia Giulia Lazio Liguria Lombardia Marche Molise Piemonte Puglia Sardegna Sicilia Toscana P.A. Trento Umbria Valle d'Aosta Veneto ITALIA ITA-Lomb
valori assoluti valori rel. (per milione di resid.)		
scala naturale scala logaritmica		
ricoverati_con_sintomi terapia_intensiva totale_ospedalizzati isolamento_domiciliare totale_positivi variazione_totale_positivi nuovi_positivi dimessi_guariti deceduti totale_casi tamponi		

Nella selezione delle serie si deve tener conto di alcune informazioni fondamentali:

le seguenti serie del database originale si riferiscono a valori cumulati:

- ricoverati_con_sintomi
- terapia_intensiva
- totale_ospedalizzati
- isolamento_domiciliare
- totale_positivi
- dimessi_guariti
- deceduti
- totale_casi
- tamponi

Solo su queste serie, pertanto, ha senso passare da “serie originaria” a “serie differenziata”.

serie originaria
serie differenziata

La serie originaria **variazione_totale_positivi** si riferisce, come dice il nome, alla differenza di **totale_positivi** con il giorno precedente. Ovvero: **variazione_totale_positivi** (tempo t)=**totale_positivi** (tempo t)-**totale_positivi** (tempo t-1).

La serie originaria **nuovi_positivi** si riferisce alla differenza di **totale_casi** con il giorno precedente. Ovvero: **nuovi_positivi** (tempo t)=**totale_casi** (tempo t)-**totale_casi** (tempo t-1).

Conseguentemente, la combinazione “serie differenziata”-“totale_positivi” coincide con la “serie originaria”-“variazione_totale_positivi”; analogamente, serie differenziata”-“totale_casi” coincide con la “serie originaria”-“nuovi_positivi”

Inoltre valgono le seguenti relazioni:

ricoverati_con_sintomi+ terapia_intensiva= totale_ospedalizzati;

totale_ospedalizzati+ isolamento_domiciliare= totale_positivi;

totale_positivi+ deceduti+ dimessi_guariti= totale_casi.

Elaborazioni a fini previsivi di una serie scelta (etichette verdi)

Seleziona serie

Foglio “Seleziona serie”

Analogamente al foglio di lavoro “DUE zone”, mediante il foglio “Seleziona serie” possiamo scegliere una (sola) serie da analizzare. Nei tre fogli successivi la serie selezionata viene elaborata in tre modi differenti.

Holt-Winters

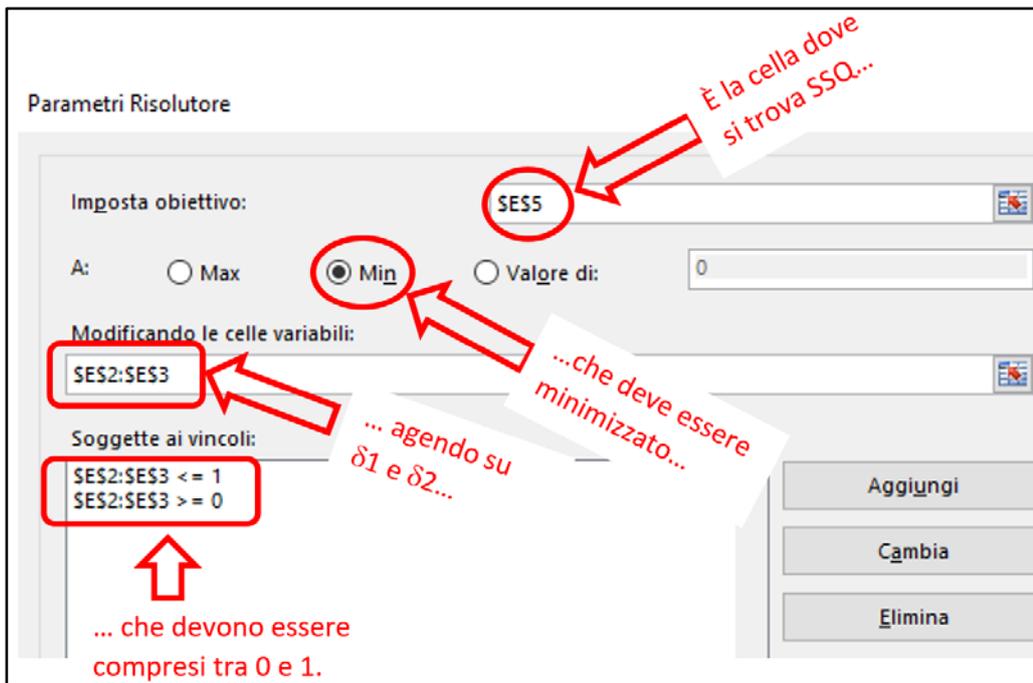
Foglio “Holt-Winters”

In questo foglio viene stimato un modello Holt-Winters di tipo semplificato: è presente il “livello” della serie e il suo “trend”, ma non la “stagionalità”. A questo proposito, non escludiamo di implementare il modello completo in un prossimo futuro, dato che, soprattutto ultimamente sembra presente un certo “effetto-domenica,” o più in generale “effetto-giorno festivo” di cui sono affetti i dati prodotti dall’ISS. Alternativamente potremmo considerare di poter analizzare una serie di medie mobili (a 3 o 5 termini ad esempio), per smorzare qualunque effetto dovuto alla variabilità della tempestività con cui i fenomeni vengono misurati dalle serie storiche ufficiali.

Nel box riquadrato in rosso viene ricordato il modello adottato. Riassumendo:

1. la previsione per il livello della serie al tempo $t+1$ deriva dalla somma di due sole componenti: livello m_t più trend T_t ;
2. ad ogni istante t
 - il livello m_t è la media ponderata (o combinazione lineare convessa) secondo δ_1 di due parti: m_{t-1} e T_{t-1} ;
 - il trend T_t è la media ponderata (o combinazione lineare convessa) secondo δ_2 di due parti: incremento del livello rispetto al tempo precedente ($m_t - m_{t-1}$) e trend precedente (T_{t-1}).

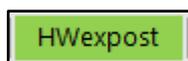
I parametri δ_1 e δ_2 sono individuati minimizzando la SSQ tra valori osservati e valori *fitted*. Per fare questo si può utilizzare lo strumento Risolutore di Excel, impostandolo secondo quanto è riassunto nella finestra qui sotto:



Lanciando il Risolutore si dovrebbero ottenere i parametri “ottimali”, che saranno in generale diversi per ogni serie analizzata.

Nel foglio compaiono le previsioni per i successivi k giorni, dove k è scelto nel menu a discesa. Il giorno 0 è l’ultimo giorno per il quale ci sono i dati nell’archivio. Pertanto per il giorno 0, oltre il valore “da modello”, si conosce anche il valore “osservato”.

Per la stima di δ_1 e δ_2 sono usati tutti i valori disponibili, dal 24 febbraio in poi. In futuro, probabilmente verrà aggiunta l’opzione di limitarsi a utilizzare gli ultimi n giorni, con n scelto dall’utente.



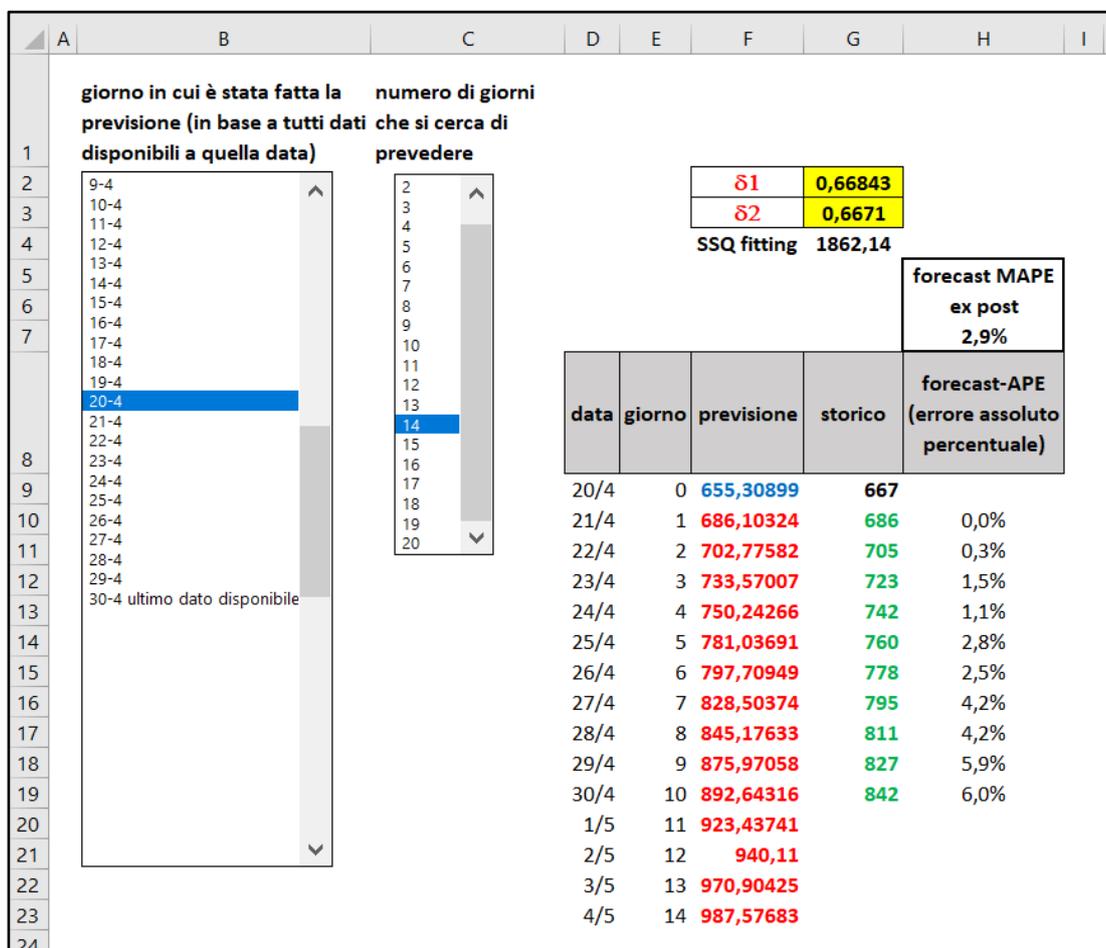
Foglio H-Wexpost

Anche in questo foglio la serie selezionata in “Seleziona serie” viene analizzata secondo lo stesso modello Holt -Winters di tipo semplificato visto in precedenza. La sola (importante) differenza è che in questo foglio ci si può posizionare fittiziamente in un giorno precedente all’ultimo per il quale sono disponibili i dati, e poi, coerentemente, stimare δ_1 e δ_2 con i dati che erano disponibili fino a quel giorno. Prolungando poi la previsione per k giorni, si può confrontare la previsione con i valori che si sono successivamente verificati.

Per chiarire meglio il funzionamento (che comunque è segue l’usuale logica del test di *goodness of forecast* condotto a posteriori) facciamo un esempio, corrispondente alla schermata riportata qui sotto (l’esempio si riferisce a “deceduti in Toscana (valori cumulati)”.

Supponiamo di avere a disposizione i dati fino al 30 aprile. Fissiamo la data “fittizia” dati al 20 aprile: in altri termini, “facciamo finta” di trovarci al 20 aprile, disponendo quindi dei soli dati disponibili fino a quel giorno (incluso). Lanciando il Risolutore (in questo caso, come si vede dalla figura, le soluzioni per la minimizzazione della cella SSQ che si trova in G4 sono contenute nelle celle G2 e G3), abbiamo trovato i parametri δ_1 e δ_2 ottimi pari, rispettivamente, a 0,66843 e 0,6671. Sulla base di questi parametri possiamo calcolare le stime per i 14 giorni successivi (dato che nel secondo menu a discesa abbiamo selezionato “14”); fino al 4 maggio dunque. Dato che in realtà disponiamo dei valori fino al 30 aprile, possiamo confrontare le previsioni coi

valori che abbiamo osservato dal 21 aprile al 30 aprile (in verde). Sulla base di questi 10 giorni che ci permettono un confronto previsione-storico, possiamo verificare il MAPE (Mean Absolute Percentage Error) di previsione. Ovviamente, le serie di valori “cumulati” (come nel caso dell’esempio qui riportato) avranno performance molto migliori rispetto alle serie di valori “giornalieri”. Analoghe considerazioni possono essere fatte per scala naturale-scala logaritmica. Alla luce di ciò dovremo calibrare le nostre frustrazioni ed entusiasmi nel trovare MAPE enormi o ridotti.



rette mobili

Rette mobili

Il foglio rette mobili produce un grafico in cui, ad ogni tempo, viene rappresentato un segmento ricavato da una retta dei minimi quadrati (variabile esplicativa è il tempo). Ciascuna retta è stimata sui dati della solita serie storica selezionata in “Seleziona serie”.

Per ogni tempo t :

- la retta è calcolata sui m valori della serie, di cui l’ultimo è quello del tempo t ; m è scelto mediante il primo menu, e va da 3 a 15;
- il segmento della retta dei m.q. inizia al tempo t ;

- il segmento della retta dei m.q. finisce al tempo al tempo $t+k$, dove k è selezionato dal secondo menu: ad esempio “segmento 8” significa che la lunghezza di ciascun segmento è 8, ovvero che va da t a $t+8$.

Appendice

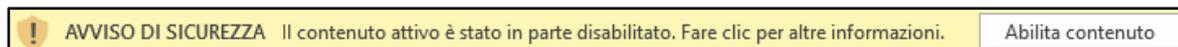
Preparazione e funzionamento del Data Base e di programmi di elaborazione flessibili

(di Mauro Maltagliati)

I lettori sono autorizzati ad usare il data base per la elaborazione dei grafici per le serie ivi riportate. Comunque, chiediamo a chi lo utilizza per le serie che per le elaborazioni per produrre lavori (articoli, ecc.) di citare sia il data base che il Manuale.

Manuale per l'uso del file autoDB.xlsm

Il file “autoDB.xlsm” è un file di Excel che contiene delle macro (da cui l'ultima lettera dell'estensione del file) scritte in linguaggio Visual Basic, Applications Edition (VBA). Le macro in Excel sono disabilitate per impostazione predefinita, dato che sono suscettibili di essere utilizzate per la consegna di codice dannoso. Per poter usare tutte le macro del file si deve cliccare su “Abilita contenuto” alla sua apertura. In particolare, in questo file, le macro sono necessarie per avviare la macro di aggiornamento automatico del database, come descritto più avanti.



Per eventuali domande sull'utilizzo, ma anche per segnalare eventuali “bugs”, contattare mauro.maltagliati@unifi.it.

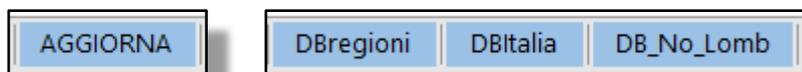
Se il file viene utilizzato per elaborazioni, si prega di citare la

All'apertura del file, come al solito in Excel, in basso si trovano le “linguette-scheda” che consentono di navigare tra i fogli elettronici.

In questo file ce ne sono 13, raggruppate per colore a seconda della tipologia.



Aggiornamento database (linguette azzurre)



Le Linguette azzurre riguardano il database vero e proprio, che viene aggiornato cliccando sul pulsante che si trova nel foglio "AGGIORNA".



Cliccando sul pulsante si aziona una macro che carica il file dalla pagina web:

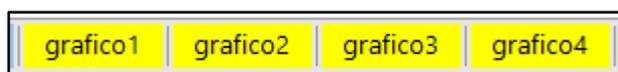
<https://github.com/pcm-dpc/COVID-19/blob/master/dati-regioni/dpc-covid19-ita-regioni.csv>.

Solitamente i nuovi dati sono presenti dalle ore 19 di ogni giorno. I dati vengono poi riordinati e immessi nel foglio "DBregioni". Come si può notare, la tabella non inizia dalla riga 1 come ci si aspetterebbe. Per trovarla si deve scorrere il foglio fino alla riga 50-70 circa (a seconda dei computer sui quali abbiamo effettuato la prova). Questa "stranezza" è dovuta al fatto che l'operazione di riordinamento della tabella provocava errori nell'esecuzione della macro su alcuni computer su cui il file è stato caricato. In futuro contiamo di risolvere questa piccola imperfezione. Sui computer Mac l'aggiornamento non ha funzionato.

Comunque, una volta che sul proprio computer si individua la riga in cui la tabella dei dati regionali inizia, ad ogni aggiornamento la troveremo sempre nella stessa posizione.

Le tabelle DBItalia e DB_No_Lomb sono ricavate dalla DBRegioni e riportano i dati aggregati per l'Italia intera e per l'Italia con esclusione della Lombardia. Queste tabelle consentiranno di produrre grafici ed elaborazioni per questi due raggruppamenti territoriali (in aggiunta alla possibilità di avere i grafici per le singole regioni).

Grafici semplici (linguette gialle)



Su questi fogli non c'è molto da dire, sono grafici il cui contenuto informativo è facilmente comprensibile. L'unica opzione per l'utente è quella di scegliere la circoscrizione geografica tra le 23 opzionabili: le 21 regioni (le province del Trentino sono separate nel database originale), l'Italia e infine l'Italia con esclusione della Lombardia. La selezione avviene cliccando sulla voce desiderata nel menu a tendina.



In questi fogli l'unica ulteriore operazione che si può compiere è copiare i grafici per incollarli su un'altra applicazione. Se si intende trasferirli in Word, si consiglia (all'interno del documento di Word) di scegliere l'opzione "Incolla speciale- Immagine (Enhanced Metafile)". È un consiglio (valido in generale) che consente di trattare il grafico importato come una semplice immagine, pur perdendo il "collegamento" con Excel.

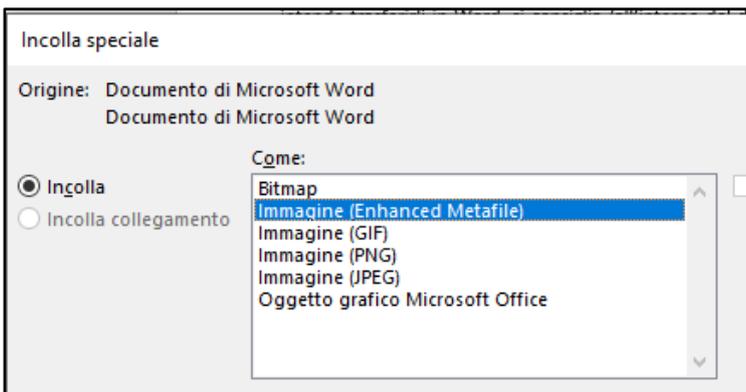
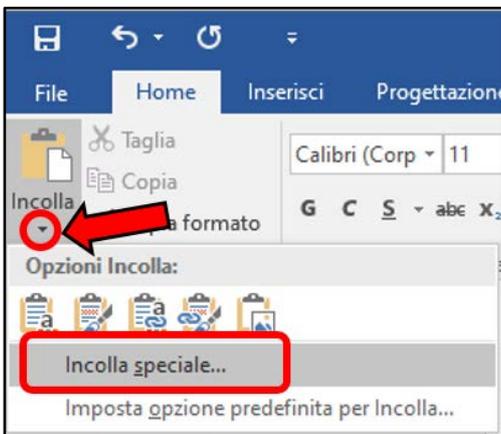


Grafico di confronto dei time plot per due diverse zone geografiche (linguetta arancione “DUE zone”)

DUE zone

Su questo foglio è possibile mettere a confronto la stessa serie storica per due differenti aree geografiche. Le due aree sono selezionate attraverso i due menu gemelli (“SERIE1 (blu)” e “SERIE2 (arancione)”) e il tipo di serie mediante i quattro menu sulla sinistra.

La combinazione “serie originaria-valori assoluti-scala naturale” dei primi tre menu rappresenta il dato originale caricato dal sito indicato in precedenza. Le altre combinazioni derivano da semplici elaborazioni. Anche per questo “foglio di lavoro” l’unica operazione a disposizione dell’utente (a parte la scelta della serie e delle zone geografiche) è poter copiare il grafico e incollarlo in un altro programma.

	SERIE1 (blu)	SERIE2 (arancione)
serie originaria serie differenziata	Abruzzo Basilicata P.A. Bolzano Calabria Campania Emilia-Romagna Friuli Venezia Giulia Lazio Liguria Lombardia Marche Molise Piemonte Puglia Sardegna Sicilia Toscana P.A. Trento Umbria Valle d'Aosta Veneto ITALIA ITA-Lomb	Abruzzo Basilicata P.A. Bolzano Calabria Campania Emilia-Romagna Friuli Venezia Giulia Lazio Liguria Lombardia Marche Molise Piemonte Puglia Sardegna Sicilia Toscana P.A. Trento Umbria Valle d'Aosta Veneto ITALIA ITA-Lomb
valori assoluti valori rel. (per milione di resid.)		
scala naturale scala logaritmica		
ricoverati_con_sintomi terapia_intensiva totale_ospedalizzati isolamento_domiciliare totale_positivi variazione_totale_positivi nuovi_positivi dimessi_guariti deceduti totale_casi tamponi		

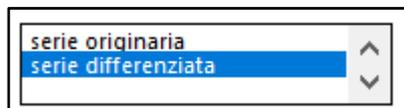
Nella selezione delle serie si deve tener conto di alcune informazioni fondamentali:

le seguenti serie del database originale si riferiscono a valori cumulati:

- [ricoverati_con_sintomi](#)
- [terapia_intensiva](#)
- [totale_ospedalizzati](#)
- [isolamento_domiciliare](#)
- [totale_positivi](#)

- dimessi_guariti
- deceduti
- totale_casi
- tamponi

Solo su queste serie, pertanto, ha senso passare da “serie originaria” a “serie differenziata”.



La serie originaria [variazione_totale_positivi](#) si riferisce, come dice il nome, alla differenza di [totale_positivi](#) con il giorno precedente. Ovvero: $\text{variazione_totale_positivi}(\text{tempo } t) = \text{totale_positivi}(\text{tempo } t) - \text{totale_positivi}(\text{tempo } t-1)$.

La serie originaria [nuovi_positivi](#) si riferisce alla differenza di [totale_casi](#) con il giorno precedente. Ovvero: $\text{nuovi_positivi}(\text{tempo } t) = \text{totale_casi}(\text{tempo } t) - \text{totale_casi}(\text{tempo } t-1)$.

Conseguentemente, la combinazione “serie differenziata”-“[totale_positivi](#)” coincide con la “serie originaria”-“[variazione_totale_positivi](#)”; analogamente, serie differenziata”-“[totale_casi](#)” coincide con la “serie originaria”-“[nuovi_positivi](#)”

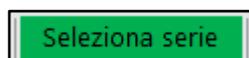
Inoltre valgono le seguenti relazioni:

$\text{ricoverati_con_sintomi} + \text{terapia_intensiva} = \text{totale_ospedalizzati}$;

$\text{totale_ospedalizzati} + \text{isolamento_domiciliare} = \text{totale_positivi}$;

$\text{totale_positivi} + \text{deceduti} + \text{dimessi_guariti} = \text{totale_casi}$.

Elaborazioni a fini previsivi di una serie scelta (etichette verdi)



Foglio “Seleziona serie”

Analogamente al foglio di lavoro “DUE zone”, mediante il foglio “Seleziona serie” possiamo scegliere una (sola) serie da analizzare. Nei tre fogli successivi la serie selezionata viene elaborata in tre modi differenti.



Foglio “Holt-Winters”

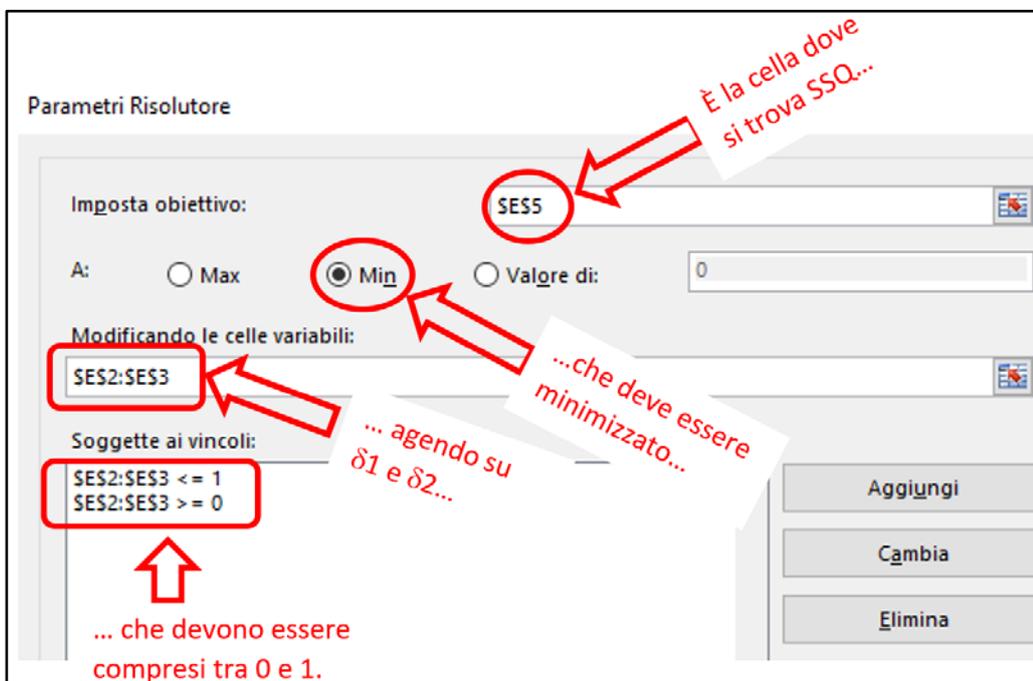
In questo foglio viene stimato un modello Holt-Winters di tipo semplificato: è presente il “livello” della serie e il suo “trend”, ma non la “stagionalità”. A questo proposito, non escludiamo di implementare il modello completo in un prossimo futuro, dato che, soprattutto ultimamente sembra presente un certo “effetto-domenica,” o più in generale “effetto-giorno festivo” di cui sono affetti i dati prodotti dall’ISS.

Alternativamente potremmo considerare di poter analizzare una serie di medie mobili (a 3 o 5 termini ad esempio), per smorzare qualunque effetto dovuto alla variabilità della tempestività con cui i fenomeni vengono misurati dalle serie storiche ufficiali.

Nel box riquadrato in rosso viene ricordato il modello adottato. Riassumendo:

3. la previsione per il livello della serie al tempo $t+1$ deriva dalla somma di due sole componenti: livello m_t più trend T_t ;
4. ad ogni istante t
 - il livello m_t è la media ponderata (o combinazione lineare convessa) secondo δ_1 di due parti: m_{t-1} e T_{t-1} ;
 - il trend T_t è la media ponderata (o combinazione lineare convessa) secondo δ_2 di due parti: incremento del livello rispetto al tempo precedente ($m_t - m_{t-1}$) e trend precedente (T_{t-1}).

I parametri δ_1 e δ_2 sono individuati minimizzando la SSQ tra valori osservati e valori *fitted*. Per fare questo si può utilizzare lo strumento Risolutore di Excel, impostandolo secondo quanto è riassunto nella finestra qui sotto:



Lanciando il Risolutore si dovrebbero ottenere i parametri “ottimali”, che saranno in generale diversi per ogni serie analizzata.

Nel foglio compaiono le previsioni per i successivi k giorni, dove k è scelto nel menu a discesa. Il giorno 0 è l’ultimo giorno per il quale ci sono i dati nell’archivio. Pertanto per il giorno 0, oltre il valore “da modello”, si conosce anche il valore “osservato”.

Per la stima di δ_1 e δ_2 sono usati tutti i valori disponibili, dal 24 febbraio in poi. In futuro, probabilmente verrà aggiunta l’opzione di limitarsi a utilizzare gli ultimi n giorni, con n scelto dall’utente.

Foglio H-Wexpost

Anche in questo foglio la serie selezionata in “Seleziona serie” viene analizzata secondo lo stesso modello Holt -Winters di tipo semplificato visto in precedenza. La sola (importante) differenza è che in questo foglio ci si può posizionare fittiziamente in un giorno precedente all’ultimo per il quale sono disponibili i dati, e poi, coerentemente, stimare δ_1 e δ_2 con i dati che erano disponibili fino a quel giorno. Prolungando poi la previsione per k giorni, si può confrontare la previsione con i valori che si sono successivamente verificati.

Per chiarire meglio il funzionamento (che comunque è segue l’usuale logica del test di *goodness of forecast* condotto a posteriori) facciamo un esempio, corrispondente alla schermata riportata qui sotto (l’esempio si riferisce a “deceduti in Toscana (valori cumulati)”).

Supponiamo di avere a disposizione i dati fino al 30 aprile. Fissiamo la data “fittizia” dati al 20 aprile: in altri termini, “facciamo finta” di trovarci al 20 aprile, disponendo quindi dei soli dati disponibili fino a quel giorno (incluso). Lanciando il Risolutore (in questo caso, come si vede dalla figura, le soluzioni per la minimizzazione della cella SSQ che si trova in G4 sono contenute nelle celle G2 e G3), abbiamo trovato i parametri δ_1 e δ_2 ottimi pari, rispettivamente, a 0,66843 e 0,6671. Sulla base di questi parametri possiamo calcolare le stime per i 14 giorni successivi (dato che nel secondo menu a discesa abbiamo selezionato “14”); fino al 4 maggio dunque. Dato che in realtà disponiamo dei valori fino al 30 aprile, possiamo confrontare le previsioni coi valori che abbiamo osservato dal 21 aprile al 30 aprile (in verde). Sulla base di questi 10 giorni che ci permettono un confronto previsione-storico, possiamo verificare il MAPE (Mean Absolute Percentage Error) di previsione. Ovviamente, le serie di valori “cumulati” (come nel caso dell’esempio qui riportato) avranno performance molto migliori rispetto alle serie di valori “giornalieri”. Analoghe considerazioni possono essere fatte per scala naturale-scala logaritmica. Alla luce di ciò dovremo calibrare le nostre frustrazioni ed entusiasmi nel trovare MAPE enormi o ridotti.

	A	B	C	D	E	F	G	H	I
1		giorno in cui è stata fatta la previsione (in base a tutti dati disponibili a quella data)	numero di giorni che si cerca di prevedere						
2		9-4	2			81	0,66843		
3		10-4	3			82	0,6671		
4		11-4	4						
5		12-4	5						
6		13-4	6						
7		14-4	7						
8		15-4	8						
9		16-4	9						
10		17-4	10						
11		18-4	11						
12		19-4	12						
13		20-4	13						
14		21-4	14						
15		22-4	15						
16		23-4	16						
17		24-4	17						
18		25-4	18						
19		26-4	19						
20		27-4	20						
21		28-4							
22		29-4							
23		30-4 ultimo dato disponibile							
24									

		SSQ fitting		1862,14	
				forecast MAPE	
				ex post	
				2,9%	

data	giorno	previsione	storico	forecast-APE (errore assoluto percentuale)
20/4	0	655,30899	667	
21/4	1	686,10324	686	0,0%
22/4	2	702,77582	705	0,3%
23/4	3	733,57007	723	1,5%
24/4	4	750,24266	742	1,1%
25/4	5	781,03691	760	2,8%
26/4	6	797,70949	778	2,5%
27/4	7	828,50374	795	4,2%
28/4	8	845,17633	811	4,2%
29/4	9	875,97058	827	5,9%
30/4	10	892,64316	842	6,0%
1/5	11	923,43741		
2/5	12	940,11		
3/5	13	970,90425		
4/5	14	987,57683		

rette mobili

Rette mobili

Il foglio rette mobili produce un grafico in cui, ad ogni tempo, viene rappresentato un segmento ricavato da una retta dei minimi quadrati (variabile esplicativa è il tempo). Ciascuna retta è stimata sui dati della solita serie storica selezionata in "Seleziona serie".

Per ogni tempo t :

- la retta è calcolata sui m valori della serie, di cui l'ultimo è quello del tempo t ; m è scelto mediante il primo menu, e va da 3 a 15;
- il segmento della retta dei m.q. inizia al tempo t ;
- il segmento della retta dei m.q. finisce al tempo al tempo $t+k$, dove k è selezionato dal secondo menu: ad esempio "segmento 8" significa che la lunghezza di ciascun segmento è 8, ovvero che va da t a $t+8$.