

# Previsioni morti Covid per il Veneto

*Gabriele Carrara, matricola 814720; Davide Porcellini, matricola 816586; Simone Tufano, matricola 816984*

*06 maggio 2020*

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(lubridate)
```

```
data <- read.csv('comune_giorno.csv')
```

Si opta per la stima dei morti per la Regione Veneto

```
data_veneto <- data[which(data$NOME_REGIONE == 'Veneto'),]
```

Si verifica che siano presenti tutti i comuni del Veneto

```
length(unique(data_veneto$NOME_COMUNE))
```

```
## [1] 563
```

Il numero coincide con i comuni ufficiali. Si procede eliminando le colonne non necessarie per l'analisi.

```
data_veneto <- data_veneto[,c(3:9,22:27)]
```

Si trasforma la data in un formato più corretto e si elimina la variabile GE.

```
data_veneto %>%
  mutate(DATA = as.Date(paste0("0", GE, "2020"),
    format = "%m%d%Y")) -> data_veneto
```

```
data_veneto <- data_veneto[, -7]
```

si separano le variabili SESSO e ANNO e si rende il dataset verticale.

```
data_veneto <- data_veneto %>%
  gather(key = "SESSO_ANNO", value = "DECESSI", TOTALE_15:TOTALE_20)

data_veneto %>%
  separate(SESSO_ANNO,
    c("SESSO", "ANNO"), "_") -> data_veneto
```

```
summary(data_veneto)
```

```
##          NOME_REGIONE    NOME_PROVINCIA    NOME_COMUNE    COD_PROVCOM
## Veneto      :435948    Treviso:82482    Venezia: 7758    Min.   :23001
## Abruzzo      :      0    Padova :79008    Verona : 7500    1st Qu.:24061
## Basilicata    :      0    Vicenza:78360    Padova : 7104    Median :26050
## Calabria     :      0    Verona :74142    Vicenza: 5754    Mean   :25835
## Campania     :      0    Venezia:63480    Treviso: 4962    3rd Qu.:28001
## Emilia-Romagna:      0    Rovigo :30660    Rovigo : 3984    Max.   :29052
## (Other)      :      0    (Other):27816    (Other):398886
##          DATA_INIZIO_DIFF    CL_ETA    DATA
## 1 aprile      :112146    Min.   : 0.0    Min.   :2020-01-01
## 16 aprile     : 34764    1st Qu.:15.0    1st Qu.:2020-01-28
## 8 aprile      : 40026    Median :17.0    Median :2020-02-25
## Dati 2020 n.d.:249012    Mean   :16.6    Mean   :2020-02-26
##                                     3rd Qu.:19.0    3rd Qu.:2020-03-26
##                                     Max.   :21.0    Max.   :2020-04-30
##
##          SESSO          ANNO          DECESSI
## Length:435948    Length:435948    Min.   : 0
## Class :character    Class :character    1st Qu.: 0
## Mode  :character    Mode  :character    Median : 0
##                                     Mean   :1074
##                                     3rd Qu.: 1
##                                     Max.   :9999
##
```

Si trasforma la variabile DATA e si unisce con la variabile ANNO.

```
data_dmy <- paste0((
  paste(
    day(data_veneto$DATA),
    ... = lubridate::month(data_veneto$DATA),
    '20', sep = '/'),
  data_veneto$ANNO)
```

Si trasforma la variabile precedente in un formato data corretto.

```
data_veneto$DATA <- lubridate::dmy(data_dmy)
```

```
summary(data_veneto)
```

```
##          NOME_REGIONE    NOME_PROVINCIA    NOME_COMUNE    COD_PROVCOM
## Veneto      :435948    Treviso:82482    Venezia: 7758    Min.   :23001
## Abruzzo      :      0    Padova :79008    Verona : 7500    1st Qu.:24061
## Basilicata    :      0    Vicenza:78360    Padova : 7104    Median :26050
## Calabria     :      0    Verona :74142    Vicenza: 5754    Mean   :25835
## Campania     :      0    Venezia:63480    Treviso: 4962    3rd Qu.:28001
## Emilia-Romagna:      0    Rovigo :30660    Rovigo : 3984    Max.   :29052
## (Other)      :      0    (Other):27816    (Other):398886
##          DATA_INIZIO_DIFF    CL_ETA    DATA
## 1 aprile      :112146    Min.   : 0.0    Min.   :2015-01-01
## 16 aprile     : 34764    1st Qu.:15.0    1st Qu.:2016-02-25
## 8 aprile      : 40026    Median :17.0    Median :2017-08-31
```

```
## Dati 2020 n.d.:249012    Mean    :16.6    Mean    :2017-08-27
##                          3rd Qu.:19.0    3rd Qu.:2019-02-25
##                          Max.     :21.0    Max.     :2020-04-30
##                          NA's     :876
##      SESSO              ANNO              DECESSI
## Length:435948          Length:435948    Min.     : 0
## Class :character       Class :character  1st Qu.: 0
## Mode  :character       Mode  :character  Median : 0
##                               Mean      :1074
##                               3rd Qu.: 1
##                               Max.      :9999
##
```

Si eliminano le variabili inutili per le analisi.

```
data_veneto <- data_veneto[,-c(1,9)]
```

Si costruisce il nuovo dataframe con i dati aggregati per data e comune per eliminare la componente “classi età”.

```
data_veneto_gb <- data_veneto %>%
  group_by(DATA, NOME_COMUNE) %>%
  summarize(DECESSI = sum(DECESSI))
```

Dataset orizzontale per sistemare le date mancanti.

```
data_veneto_spread <- data_veneto_gb %>%
  spread(key = "DATA", value = "DECESSI")
```

Vengono sostituiti i valori mancanti con 0.

```
data_veneto_spread[is.na(data_veneto_spread)] <- 0
```

Una volta inseriti gli 0 al posto dei valori mancanti si torna al dataset tidy

```
data_veneto_tidy <- data_veneto_spread %>%
  gather(key = "DATA", value = "DECESSI", `2015-01-01`:`2020-04-30`)
data_veneto_tidy <- data_veneto_tidy[-2]
```

si estraggono i comuni

```
comuni <- data_veneto_tidy %>%
  .$NOME_COMUNE %>%
  unique()
```

Si crea la finestra previsionale (primi 3 mesi 2020) e la lista che conterrà i risultati

```
giorni_previsione <- 91
finestra_prev <- as.Date("2019-12-31") %m+% days(1:giorni_previsione)
all_pred <- list()
```

Tramite un modello auto arima si stimano i morti per comune

```

for (i in 1:length(comuni)) {

  current_comune <- comuni[i]

  pred_death <- data_veneto_tidy %>%
    filter(NOME_COMUNE == current_comune &
           DATA<= as.Date("2019-12-31")) %>%
    arrange(DATA) %>%
    .$DECESSI %>%
    ts(frequency = 365) %>%
    forecast::auto.arima() %>%
    forecast::forecast(h = giorni_previsione)

  out_pred <- tibble(NOME_COMUNE = rep(current_comune, giorni_previsione),
                    DATA = finestra_prev,
                    PREVISIONE = pred_death$mean)

  all_pred[[i]] <- out_pred
}

```

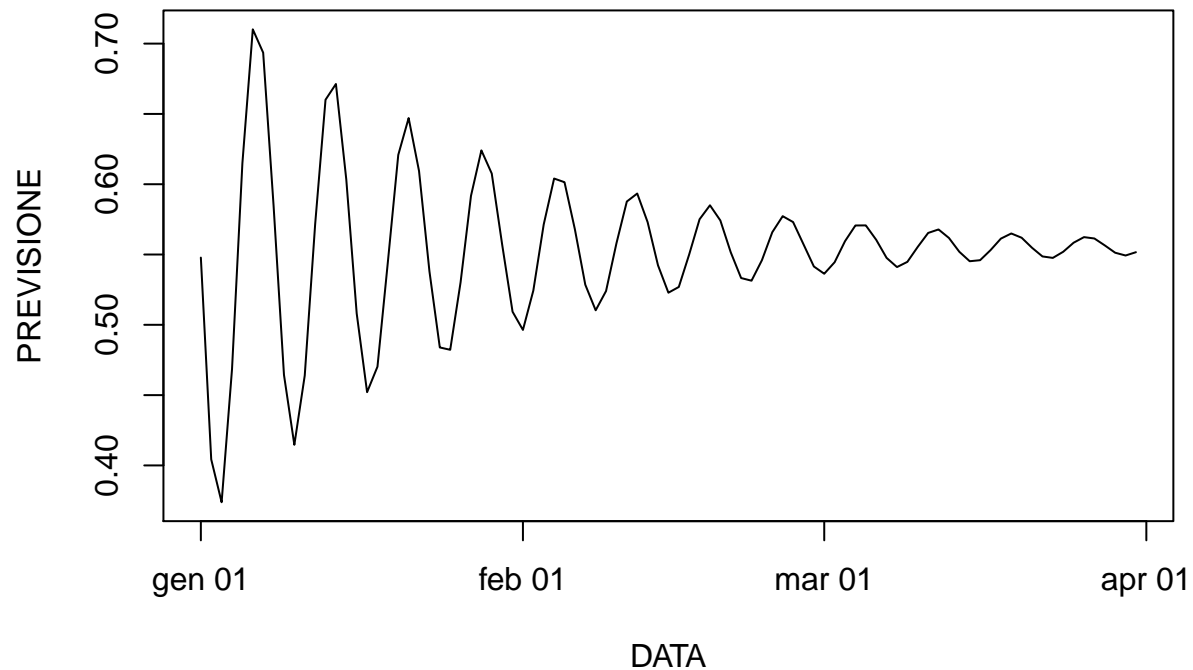
Si uniscono tutte le previsioni

```
previsioni_veneto<- bind_rows(all_pred)
```

Esempio di una previsione

```
plot(all_pred[[1]][2:3], type= 'l', main = "Abano Terme")
```

## Abano Terme



Si nota che il valore puntuale della stima si avvicina nel tempo alla media, tuttavia mantiene una quota di varianza lungo tutta la finestra previsionale. Si opta per non arrotondare la cifra stimata per evitare di introdurre un Bias sistematico.

Si aggiunge la colonna relativa ai decessi effettivi per il 2020

```
confronto<- merge(previsioni_veneto,
                  filter(data_veneto_tidy,
                        DATA >= "2019-12-31" &
                        DATA <= "2020-03-31"),
                  by= c("NOME_COMUNE", "DATA"))
head(confronto)
```

##	NOME_COMUNE	DATA	PREVISIONE	DECESSI
## 1	Abano Terme	2020-01-01	0.5478239	0
## 2	Abano Terme	2020-01-02	0.4042795	0
## 3	Abano Terme	2020-01-03	0.3738158	1
## 4	Abano Terme	2020-01-04	0.4685408	1
## 5	Abano Terme	2020-01-05	0.6149640	0
## 6	Abano Terme	2020-01-06	0.7101463	1

Si unisce la provincia al dataset. Questo al fine di applicare ai comuni in cui non sono presenti i dati un tasso d'incremento provinciale

```
comune_provincia<-unique(data_veneto[c("NOME_COMUNE", "NOME_PROVINCIA")])
confronto<- merge(confronto, comune_provincia,
                  by= "NOME_COMUNE")
head(confronto)
```

```
##   NOME_COMUNE      DATA PREVISIONE DECESSI NOME_PROVINCIA
## 1 Abano Terme 2020-01-01  0.5478239      0      Padova
## 2 Abano Terme 2020-01-02  0.4042795      0      Padova
## 3 Abano Terme 2020-01-03  0.3738158      1      Padova
## 4 Abano Terme 2020-01-04  0.4685408      1      Padova
## 5 Abano Terme 2020-01-05  0.6149640      0      Padova
## 6 Abano Terme 2020-01-06  0.7101463      1      Padova
```

```
comune_presenza_dati<- unique(data_veneto[c("NOME_COMUNE", "DATA_INIZIO_DIFF")])
confronto<- merge(confronto, comune_presenza_dati,
                  by= "NOME_COMUNE")
```

Si crea il dataset per stimare il tasso d'incremento dei decessi reali rispetto agli stimati dove si hanno dati certi per provincia

```
tasso<- confronto%>%
  filter(DECESSI<5000)%>%
  filter(DATA_INIZIO_DIFF!= "Dati 2020 n.d.")%>%
  group_by(NOME_PROVINCIA)%>%
  summarise(SOMMA_DECESSI=sum(DECESSI), SOMMA_PREVISTI=sum(PREVISIONE))
head(tasso)
```

```
## # A tibble: 6 x 3
##   NOME_PROVINCIA SOMMA_DECESSI SOMMA_PREVISTI
##   <fct>          <int>          <dbl>
## 1 Belluno        196          167.
## 2 Padova         710          550.
## 3 Rovigo         355          328.
## 4 Treviso       1139          989.
## 5 Venezia        950          824.
## 6 Verona       1728         1453.
```

Si calcola il tasso in questione

```
tasso$tasso_incremento<- tasso$SOMMA_DECESSI/tasso$SOMMA_PREVISTI
head(tasso)
```

```
## # A tibble: 6 x 4
##   NOME_PROVINCIA SOMMA_DECESSI SOMMA_PREVISTI tasso_incremento
##   <fct>          <int>          <dbl>          <dbl>
## 1 Belluno        196          167.          1.17
## 2 Padova         710          550.          1.29
## 3 Rovigo         355          328.          1.08
## 4 Treviso       1139          989.          1.15
## 5 Venezia        950          824.          1.15
## 6 Verona       1728         1453.          1.19
```

Si nota che il numero di decessi è maggiore in ogni provincia ma il valore del tasso non è molto elevato

```
confronto<- merge(confronto, tasso,  
                  by= "NOME_PROVINCIA")
```

Dove non erano presenti i dati si stimano i decessi moltiplicando le previsioni del paese per il tasso provinciale

```
da_prevedere<- confronto %>%  
  filter(DATA_INIZIO_DIFF== "Dati 2020 n.d.")  
da_prevedere$DECESSI<- da_prevedere$PREVISIONE* da_prevedere$tasso_incremento
```

Si uniscono i risultati

```
altri<- confronto %>%  
  filter(DATA_INIZIO_DIFF!= "Dati 2020 n.d.")  
data_finale<-bind_rows(altri, da_prevedere)
```

```
colSums(data_finale[4:5])
```

```
## PREVISIONE    DECESSI  
##    13218.42    15698.63
```

Il numero di morti da noi stimato per il Veneto è 2480, mentre il numero ufficiale è 1565 al 06/05.