

Stima_Morti_Emiliana

Davoli_Asperti

6/5/2020

Si considerano 3 dataset differenti:

- Dataset sulla mortalità in tutti i comuni italiani.
- Dataset ISTAT, con solo alcuni tra i comuni più colpiti dal virus.
- Dataset contenete le popolazioni totali di tutti i comuni italiani.

```
tot_comune<-read.csv('C:/Users/davol/Documents/Magistrale/DataScience_LAB/assignment/prova.csv', header=TRUE)
covid<-read.csv('C:/Users/davol/Documents/Magistrale/DataScience_LAB/assignment/dati-giornalieri-comuni.csv', header=TRUE)
pop_totale<- read.csv('C:/Users/davol/Documents/Magistrale/DataScience_LAB/assignment/Popolazione_comuni.csv', header=TRUE)
```

Si considera come regione d'interesse l'Emilia-Romagna. Si estrapolano dai primi 2 dataset, le informazioni relative a questa regione (usando la variabile NOME_REGIONE) e si trattano i dati relativi ad alcune variabili.

Nel dataset 1 sono presenti per ogni giorno e ogni comune il numero totale dei morti fino ad aprile, negli anni dal 2015-2019. Si effettuano dei raggruppamenti per ottenere il numero dei morti totale per ogni anno per ogni comune, suddivisa nei vari mesi.

Anche nel dataset 2 si effettua un raggruppamento per ottenere il numero totale di morti nei vari mesi, nei vari comuni. Non si tiene conto della suddivisione tra uomini e donne. I dati forniti dall'ISTAT offrono informazioni riguardanti 178 comuni dell'Emilia-Romagna su un totale di 328.

L'analisi è effettuata per i mesi da gennaio a marzo.

```
tot_emilia<- filter(tot_comune, NOME_REGIONE == 'Emilia-Romagna')
tot_emilia$MESE_DECESSO<-as.numeric(tot_emilia$MESE_DECESSO)
tot_emilia<- tot_emilia %>%
  group_by(NOME_COMUNE, MESE_DECESSO, add=FALSE)%>%
  summarise(somma_2015=sum(DECESSI_2015), somma_2016=sum(DECESSI_2016),somma_2017=sum(DECESSI_2017),somma_2018=sum(DECESSI_2018),somma_2019=sum(DECESSI_2019))

decessi_comuni<- filter(tot_emilia, MESE_DECESSO<=3)

covid_20<- filter(covid, NOME_REGIONE == 'Emilia-Romagna')
covid_20<- filter(covid_20, GE<=399)
covid_20 <- covid_20 %>%
  mutate(DATA = as.Date(paste0("0", GE, "2020"), format = "%m%d%Y"))

covid_20<- covid_20 %>%
  gather(key = "SESSO_ANNO", value = "DECESSI", MASCHI_15:TOTALE_20)
```

```
covid_20<-filter(covid_20, SESSO_ANNO=="TOTALE_20")

covid_20<- filter(covid_20, DECESSI<9999)
covid_20$MESE_DECESSO<-month(covid_20$DATA)

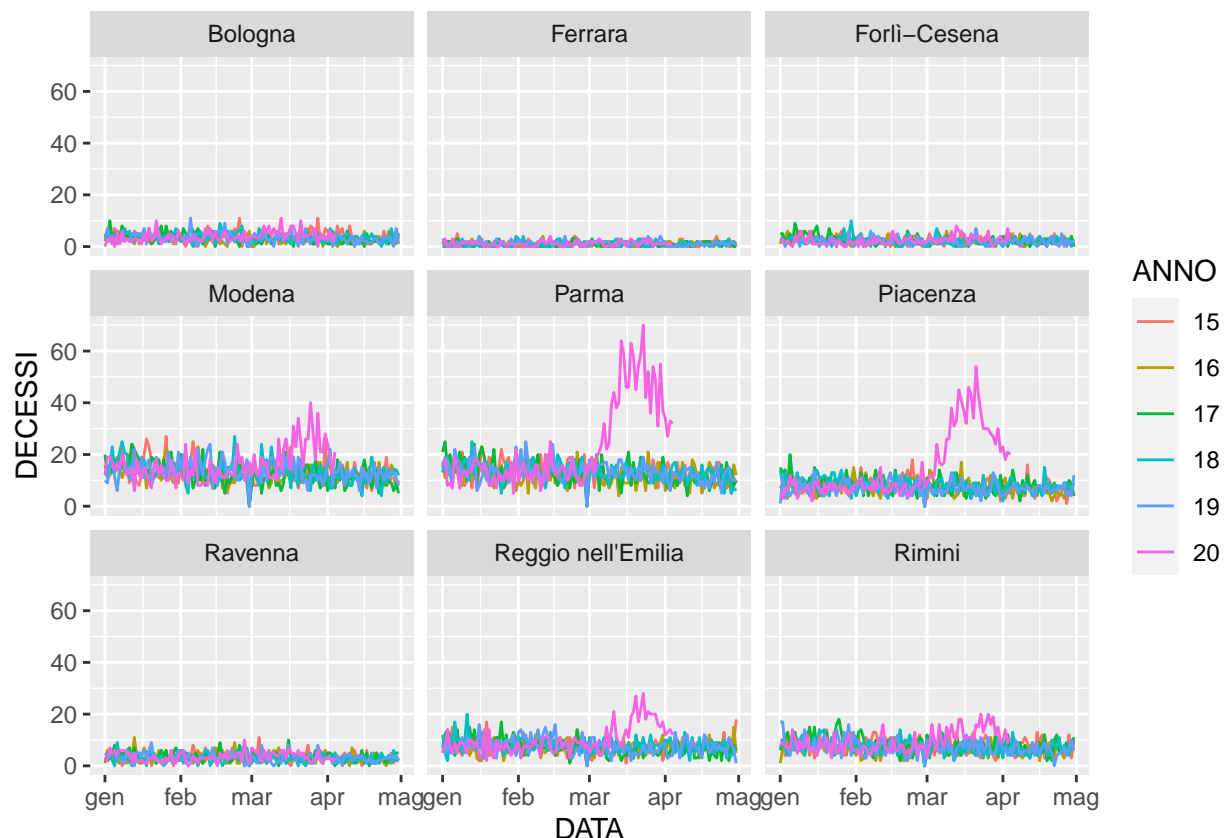
covid_20<- covid_20 %>%
  group_by(NOME_COMUNE, MESE_DECESSO ) %>%
  summarise(DECESSI_2020 = sum(DECESSI))
```

Per avere una stima dei morti degli anni fino al 2019 si calcola la media nei vari mesi per i vari comuni.

```
decessi_comuni$morti_medi_2015_2019<- rowMeans(decessi_comuni[,3:7], na.rm=FALSE)
```

Si crea un Dataset unico, usando come chiave di merge il nome del comune e il mese del decesso. In questo modo è già possibile vedere il cambiamento dagli anni passati al 2020. Si nota che nel mese di gennaio il cambiamento è spesso nel senso opposto di quello atteso. Ci si aspettava infatti un aumento dei morti fin da gennaio. Evidentemente la Regione non era ancora stata colpita dall'epidemia. Si mostrano i valori per le province dell'Emilia-Romagna, in cui si considerano sia la città provincia che i comuni che fanno parte della provincia. Si effettua anche un controllo sulle singole città provincia, per vedere se quello che appare nei plot è giustificato dai valori alti nella città o in tutta la provincia.

```
wdt %>% ggplot(aes(x = DATA, y = DECESSI, color = ANNO)) +
  geom_line() +
  facet_wrap(~NOME_PROVINCIA)
```



```
covid_bologna<-filter(covid_tot, NOME_COMUNE=='Bologna')
covid_bologna<-select(covid_bologna, NOME_COMUNE, MESE_DECESSO, morti_medi_2015_2019, DECESSI_2020)
head(covid_bologna)
```

```
##  NOME_COMUNE MESE_DECESSO morti_medi_2015_2019 DECESSI_2020
## 1 Bologna 1 501.8 400
## 2 Bologna 2 409.8 416
## 3 Bologna 3 419.6 504
```

```
covid_parma<-filter(covid_tot, NOME_COMUNE=='Parma')
covid_parma<-select(covid_parma, NOME_COMUNE, MESE_DECESSO, morti_medi_2015_2019, DECESSI_2020)
head(covid_parma)
```

```
##  NOME_COMUNE MESE_DECESSO morti_medi_2015_2019 DECESSI_2020
## 1 Parma 1 200.4 188
## 2 Parma 2 178.0 169
## 3 Parma 3 175.0 507
```

```
covid_modena<-filter(covid_tot, NOME_COMUNE=='Modena')
covid_modena<-select(covid_modena, NOME_COMUNE, MESE_DECESSO, morti_medi_2015_2019, DECESSI_2020)
head(covid_modena)
```

```
##  NOME_COMUNE MESE_DECESSO morti_medi_2015_2019 DECESSI_2020
## 1 Modena 1 216.6 181
## 2 Modena 2 177.8 186
## 3 Modena 3 184.2 265
```

```
covid_rimini<-filter(covid_tot, NOME_COMUNE=='Rimini')
covid_rimini<-select(covid_rimini, NOME_COMUNE, MESE_DECESSO, morti_medi_2015_2019, DECESSI_2020)
head(covid_rimini)
```

```
##  NOME_COMUNE MESE_DECESSO morti_medi_2015_2019 DECESSI_2020
## 1 Rimini 1 159.6 155
## 2 Rimini 2 122.4 128
## 3 Rimini 3 133.4 219
```

```
covid_piacerenza<-filter(covid_tot, NOME_COMUNE=='Piacenza')
covid_piacerenza<-select(covid_piacerenza, NOME_COMUNE, MESE_DECESSO, morti_medi_2015_2019, DECESSI_2020)
head(covid_piacerenza)
```

```
##  NOME_COMUNE MESE_DECESSO morti_medi_2015_2019 DECESSI_2020
## 1 Piacenza 1 121.4 102
## 2 Piacenza 2 113.2 122
## 3 Piacenza 3 115.6 458
```

Nelle varie province si individuano due comportamenti differenti. Confrontando il plot con i valori della singola città, appare che se nella provincia di Bologna appare un andamento costante, non vale per la singola città, che nel mese di marzo ha avuto un incremento di morti di circa 25%. Al contrario per Parma i valori alti della provincia vanno insieme ai valori elevati della città.

Si unisce anche l'informazione relativa alla popolazione nei vari comuni per poter analizzare l'effetto del virus in percentuale. La popolazione Emiliana è pari a 4459477 abitanti.

```
## Warning: si è prodotto un NA per coercizione
```

```
#merge
covid_tot<- merge(covid_tot, pop_totale, by='NOME_COMUNE')

emilia<-(sum(covid_tot$Value, na.rm = TRUE))/3
emilia
```

```
## [1] 4459477
```

```
#popolazione con morti
emilia_covid<- merge(covid_tot,covid_20, by=c('NOME_COMUNE','MESE_DECESSO' ))
pop_istat<-sum(emilia_covid$Value)/3
pop_istat
```

```
## [1] 2718973
```

```
diff<-emilia-pop_istat
diff
```

```
## [1] 1740504
```

Le informazioni fornite dall'ISTAT riguardano solamente 2718973 persone. Ne restano 1740504 per le quali è necessario stimare il valore dei morti nel 2020.

Si decide di suddividere i comuni in base al numero di abitanti, ritenendo che in comuni di diverse grandezze il virus si diffonda in maniera differente. Si utilizzano come fasce (0:10000, 10001:50000, 50000:infinito)

Per ognuna di queste suddivisioni si calcola la percentuale dei decessi nel 2020 nei comuni conosciuti, per poi applicarla come stima per i comuni di cui non si dispone dell'informazione. In tale stima si suppone che da mese a mese la popolazione rimanga costante (numero morti = numero nati).

```
covid_tot$Type<-ifelse(covid_tot$Value<=10000, "1",
  ifelse(covid_tot$Value<=50000, "2",
    ifelse(covid_tot$Value>50000, "3", NA )))
emilia_covid$Type<-ifelse(emilia_covid$Value<=10000, "1",
  ifelse(emilia_covid$Value<=50000, "2",
    ifelse(emilia_covid$Value>50000, "3", NA )))
table(covid_tot$Type)
```

```
##
##   1   2   3
## 684 261  39
```

```
paesi<-filter(covid_tot, Type==1)
citta<-filter(covid_tot, Type==2)
metropoli<-filter(covid_tot, Type==3)
paesi_c<-filter(emilia_covid, Type==1)
citta_c<-filter(emilia_covid, Type==2)
metropoli_c<-filter(emilia_covid, Type==3)
```

```
#calcolo della percentuale dei decessi nelle varie categorie, nei 3 mesi supponendo che la popolazione
```

```
p_paesi<-sum(paesi$DECESSI_2020 , na.rm=TRUE)/(sum(paesi_c$Value)/3)*100
p_citta<-sum(citta$DECESSI_2020 , na.rm=TRUE)/(sum(citta_c$Value)/3)*100
p_metropoli<-sum(metropoli$DECESSI_2020 , na.rm=TRUE)/(sum(metropoli_c$Value)/3)*100

p_paesi
```

```
## [1] 0.4319454
```

```
p_citta
```

```
## [1] 0.3786004
```

```
p_metropoli
```

```
## [1] 0.3780121
```

Infine si stimano i morti effettivi causati dal coronavirus in tutti i comuni dell'Emilia-Romagna. Si ottiene la stima, nei comuni per i quali manca l'informazione, come differenza tra i morti medi degli 5 anni precedenti e i morti stimati causati dal virus ottenuti come prodotto tra la percentuale ottenuta prima e la popolazione del paese.

Si effettua tale stima applicando in ogni suddivisione (paese, città, metropoli) la percentuale relativa trovata prima.

Per ottenere un range di variazione si considera non solo la stima puntuale ma anche un arrotondamento per eccesso, considerando che un valore medio non ha molto senso quando si parla di mortalità.

```
paesi$covid_morti<-ifelse(is.na(paesi$DECESSI_2020)==TRUE, paesi$Value*p_paesi/100, ifelse(paesi$DECESSI_2020,
#stima arrotondata per eccesso
paesi$morti_int<-as.numeric(paesi$covid_morti)
paesi$morti_int<-ceiling(paesi$covid_morti)

#stima puntuale
paesi$covid_morti<-as.numeric(paesi$covid_morti)

stima_covid_int<-sum(paesi$morti_int)/3
stima_covid<-sum(paesi$covid_morti)/3
media_noncovid<-sum(paesi$morti_medi_2015_2019)/3
df_paesi_int=stima_covid_int-media_noncovid
df_paesi=stima_covid-media_noncovid
df_paesi
```

```
## [1] 1523.65
```

```
df_paesi_int
```

```
## [1] 1578
```

```

citta$covid_morti<-ifelse(is.na(citta$DECESSI_2020)==TRUE, citta$Value*p_citta/100, ifelse(citta$DECESSI_2020==TRUE,
#stima arrotondata
citta$morti_int<-as.numeric(citta$covid_morti)
citta$morti_int<-ceiling(citta$covid_morti)

#stima non arrotondata
citta$covid_morti<-as.numeric(citta$covid_morti)

stima_covid_int<-sum(citta$morti_int)/3
stima_covid<-sum(citta$covid_morti)/3
media_noncovid<-sum(citta$morti_medi_2015_2019)/3
df_citta_int=stima_covid_int-media_noncovid
df_citta=stima_covid-media_noncovid
df_citta

```

```
## [1] 2076.339
```

```
df_citta_int
```

```
## [1] 2096.8
```

```

metropoli$covid_morti<-ifelse(is.na(metropoli$DECESSI_2020)==TRUE, metropoli$Value*p_metropoli/100, ifelse(metropoli$DECESSI_2020==TRUE,
#stima arrotondata
metropoli$morti_int<-as.numeric(metropoli$covid_morti)
metropoli$morti_int<-ceiling(metropoli$covid_morti)

#stima non arrotondata
metropoli$covid_morti<-as.numeric(metropoli$covid_morti)

stima_covid_int<-sum(metropoli$morti_int)/3
stima_covid<-sum(metropoli$covid_morti)/3
media_noncovid<-sum(metropoli$morti_medi_2015_2019)/3

df_metropoli_int=stima_covid_int-media_noncovid
df_metropoli=stima_covid-media_noncovid
df_metropoli

```

```
## [1] 2035.87
```

```
df_metropoli_int
```

```
## [1] 2038.4
```

```

morti_tot_covid=df_metropoli+df_citta+df_paesi
morti_tot_covid_int=df_metropoli_int+df_citta_int+df_paesi_int
morti_tot_covid

```

```
## [1] 5635.86
```

```
morti_tot_covid_int
```

```
## [1] 5713.2
```

```
sum(covid_tot$morti_medi_2015_2019)-sum(covid_tot$DECESSI_2020, na.rm = TRUE)
```

```
## [1] 3536.4
```

Si ottengono valori tra 1524 e 1578 per i possibili morti dovuti al covid nei paesi, tra 2076 e 2097 per le città e un range tra 2036 e 2038 per le metropoli.

In totale i decessi dichiarati nel documento ISTAT relativi all'Emilia-Romagna sono 3536 nei mesi da gennaio a marzo. La stima da noi calcolata è pari ad un numero di morti compresi tra 5636 e 5713.