

# Homework di Analisi dei Dati e Statistica

De Massari Thomas

Numero di matricola: 226091

*#Carico il data-frame wvs e verifico se l'operazione è avvenuta correttamente  
#Se tutto avviene correttamente, nel dataset di R deve essere presente solo  
l'oggetto wvs*

```
rm(list=ls())  
load('C:\\Users\\Thomas De Massari\\Documents\\Homework  
ADES\\WorldValueSurvey.rdata'); ls()
```

```
## [1] "wvs"
```

*#Visualizzo la sintesi delle variabili e del dataset a video*  
names(wvs); str(wvs)

```
## [1] "v22"      "v46"      "v47"      "v68"      "v91"      "v92"      "v93"      "v94"  
## [9] "v114"     "v116"     "v117"     "v118"     "v119"     "v120"     "v121"     "v122"  
## [17] "v123"     "v162"     "v163"     "v178"     "v192"     "v198"     "v199"     "v200"  
## [25] "v201"     "v202"     "v203"     "v204"     "v205"     "v206"     "v207"     "v208"  
## [33] "v221"     "v235"     "v236"     "v237"     "v252"     "v255"     "etacat"   "area"
```

```
## 'data.frame':    1012 obs. of  40 variables:
```

```
## $ v22      : num  7 9 7 7 7 6 9 5 4 7 ...  
## $ v46      : num  10 3 7 8 7 5 1 5 5 7 ...  
## $ v47      : num  3 1 7 5 6 6 1 5 5 7 ...  
## $ v68      : num  5 8 7 8 10 5 6 3 7 7 ...  
## $ v91      : num  6 8 3 7 4 5 8 5 5 5 ...  
## $ v92      : num  5 9 4 7 8 NA 10 5 6 5 ...  
## $ v93      : num  8 8 7 2 9 8 5 10 8 5 ...  
## $ v94      : num  10 10 8 2 10 6 3 8 8 5 ...  
## $ v114     : num  NA 5 9 3 NA 4 10 2 NA 5 ...  
## $ v116     : num  1 10 8 3 7 8 10 7 7 4 ...  
## $ v117     : num  3 1 1 5 7 NA 1 5 5 4 ...  
## $ v118     : num  1 6 8 5 6 NA 10 2 5 7 ...  
## $ v119     : num  1 1 2 4 5 5 1 2 2 7 ...  
## $ v120     : num  10 3 4 7 4 5 1 6 7 5 ...  
## $ v121     : num  10 5 7 2 9 9 10 5 9 6 ...  
## $ v122     : num  10 7 6 7 7 6 3 3 9 4 ...  
## $ v123     : num  6 10 3 6 6 4 8 7 5 7 ...  
## $ v162     : num  8 10 9 10 9 10 10 3 8 7 ...  
## $ v163     : num  2 8 5 2 9 3 4 3 7 5 ...  
## $ v178     : num  1 7 6 3 9 9 2 5 7 8 ...  
## $ v192     : num  10 8 8 8 10 9 10 7 9 7 ...  
## $ v198     : num  1 1 1 2 2 1 1 3 1 3 ...  
## $ v199     : num  1 1 4 3 1 1 1 1 2 3 ...  
## $ v200     : num  1 1 6 1 2 1 1 1 2 3 ...  
## $ v201     : num  1 1 2 1 1 1 1 1 1 3 ...
```

```
## $ v202 : num 1 1 1 5 1 1 1 1 1 3 ...
## $ v203 : num 1 2 1 3 1 1 1 3 1 4 ...
## $ v204 : num 1 2 1 6 1 3 1 5 5 4 ...
## $ v205 : num 2 5 3 5 1 1 1 5 7 4 ...
## $ v206 : num 1 1 1 7 1 3 1 8 5 3 ...
## $ v207 : num 1 1 1 1 1 1 1 NA 5 3 ...
## $ v208 : num 1 1 1 1 1 1 1 1 1 1 ...
## $ v221 : num 6 8 6 9 5 NA 9 5 8 7 ...
## $ v235 : chr "female" "male" "female" "male" ...
## $ v236 : num 1969 1959 1964 1980 1937 ...
## $ v237 : num 36 46 41 25 68 44 53 60 49 64 ...
## $ v252 : num 4 4 2 2 4 4 3 3 4 3 ...
## $ v255 : num 4 4 4 4 4 4 4 4 4 4 ...
## $ etacat: Ord.factor w/ 7 levels "[18,19]"<"(19,29]"<...: 3 4 4 2 6 4 5 6 4 6
...
## $ area : chr "C" "C" "C" "C" ...
```

```
library(rmf) #carico La Libreria rmf
```

#### *#ESERCIZIO 1*

*#L'analisi dovrà essere condotta sulla variabile v119, che contiene le risposte, su una scala di Likert, alla domanda "la competizione è positiva?". Alla risposta "si" è attribuito il valore 1, mentre alla risposta "no" è attribuito il valore 10; contemplano le sfumature intermedie da 2 a 9. Se l'intervistato non sa rispondere o non vuole, viene attribuito rispettivamente il valore -1 e -2.*

```
v119 = wvs$v119
```

```
#Distribuzione per frequenze
```

```
#Distribuzione per frequenze assolute
```

```
FreAss = table(v119)
```

```
#Distribuzione per frequenze relative
```

```
FreRel = prop.table(table(v119))
```

```
#Distribuzione per frequenza percentuale
```

```
FrePer = prop.table(table(v119))*100
```

```
#Creo un dataset contenente la variabile in questione e le sue distribuzioni di frequenza, calcolate sopra
```

```
cbind(FreAss,FreRel,FrePer)
```

```
##      FreAss      FreRel      FrePer
## 1      112 0.11558308 11.558308
## 2      121 0.12487100 12.487100
## 3      132 0.13622291 13.622291
## 4      122 0.12590299 12.590299
## 5      205 0.21155831 21.155831
## 6      112 0.11558308 11.558308
## 7       65 0.06707946  6.707946
## 8       42 0.04334365  4.334365
## 9       19 0.01960784  1.960784
## 10      39 0.04024768  4.024768
```

```

#Media con R e con RMF
(miR = mean(v119, na.rm=TRUE)) #è necessario specificare di non considerare i dati
mancanti

## [1] 4.424149

(miRMF = media(v119)) #in automatico non considera i dati mancanti

## Osservazioni mancanti: 43

## [1] 4.424149

#Mediana con R e con RMF (discorso analogo alla media)
median(v119, na.rm=TRUE)

## [1] 4

mediana(v119)

## Osservazioni mancanti: 43

## [1] 4

#Calcolo della varianza
#1. Calcolo della devianza, escludendo i dati mancanti
devianza = sum((v119-miR)^2, na.rm=TRUE)
#2. Calcolo della varianza con conseguente stampa a video del risultato
N = sum(FreAss) #dimensione della popolazione, al netto dei dati mancanti
(varianza = devianza/N)

## [1] 5.406269

#Con R
var(v119, na.rm=TRUE)*((N-1)/N)

## [1] 5.406269

#Calcolo della deviazione standard
sqrt(varianza)

## [1] 2.325139

#con R
sd(v119, na.rm=TRUE)*((N-1)/N)

## [1] 2.323938

#Si può concludere che il 50% degli intervistati hanno espresso un giudizio minore
o uguale a 4, mostrando quindi una leggera propensione per la competizione.

#ESERCIZIO 2
#L'analisi dovrà essere condotta sulla variabile v119 e sulla variabile area, che
raccoglie le zone d'Italia di provenienza degli intervistati (nord-ovest,
nord-est, centro, sud e isole)

```

```
CompetitionIsGood = wvs$v119
Area = wvs$area
```

*#gli output prodotti sono i medesimi, solo che tapply è più sintetica*  
`by(CompetitionIsGood, Area, mean, na.rm=TRUE)`

```
## Area: C
## [1] 4.434615
## -----
## Area: I
## [1] 4.689655
## -----
## Area: NE
## [1] 3.949153
## -----
## Area: NO
## [1] 4.441406
## -----
## Area: S
## [1] 4.506849
```

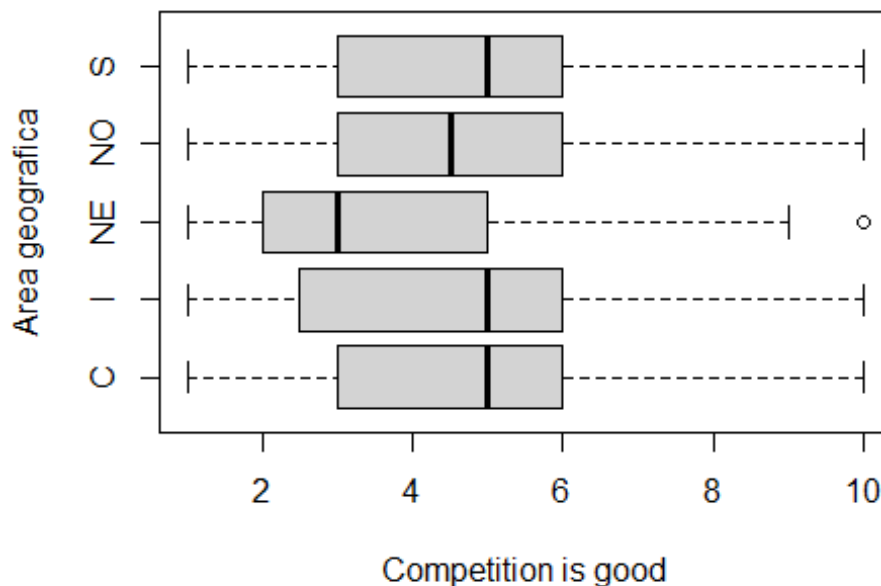
```
tapply(CompetitionIsGood,Area,mean,na.rm=TRUE)
```

```
##      C      I      NE      NO      S
## 4.434615 4.689655 3.949153 4.441406 4.506849
```

*#I dati in questione mostrano che nel nord-est d'Italia c'è una leggera propensione per la competizione rispetto alle altre aree d'Italia, seppur non netta in quanto il valore medio è approssimabile a 4. Le altre regioni invece sono più conformi alla media nazionale.*

### *#ESERCIZIO 3*

```
CompetitionIsGood = wvs$v119
Area = wvs$area
boxplot(CompetitionIsGood~Area, horizontal=TRUE, xlab="Competition is
good", ylab="Area geografica")
```



#Il box plot rappresenta in modo esplicito la situazione descritta sopra: il nord-est si distacca dall'andamento mediano nazionale, mentre il centro, il sud e le isole hanno un comportamento simile.

#### #ESERCIZIO 4

#L'analisi dovrà essere condotta sulle variabili v120 e v207, che contengono rispettivamente le risposte alle domande "lavorare duramente nel lungo periodo conduce di solito al successo?" e "non è mai giustificato suicidarsi?". Con 1 viene rappresentata la risposta affermativa e con 10 la risposta negativa. Al solito, si associa -1 e -2 alle risposte "non so" e "non voglio rispondere".

```
c3 = wvs$v120
```

```
c4 = wvs$v207
```

```
x = c3[complete.cases(c3,c4)]
```

```
cov(c3,c4,use="complete.obs")*(length(x)-1)/length(x)
```

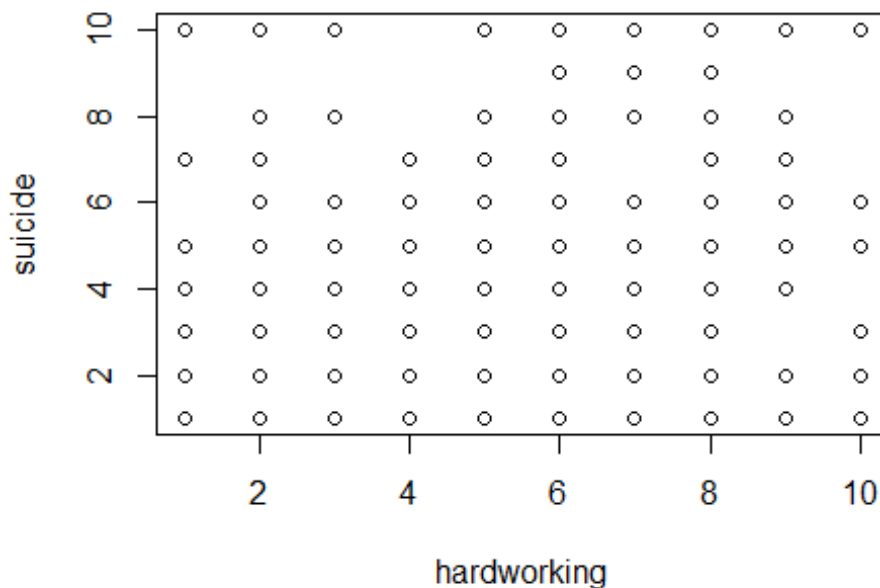
```
## [1] 0.3990293
```

```
cor(c3,c4,use="complete.obs")
```

```
## [1] 0.07817852
```

```
hardworking=c3
```

```
plot(hardworking,c4,ylab="suicide")
```



*#Si può concludere che non c'è nessuna correlazione tra la convinzione o meno che il duro lavoro porti al successo e la giustificazione del suicidio. A sostegno di ciò, oltre al coefficiente di correlazione lineare prossimo allo zero, la rappresentazione grafica mostra come le varie osservazioni sono posizionate in modo caotico sul piano cartesiano.*

#### *#ESERCIZIO 5*

```
dado = c(1:20) #dado con 20 facce
lanci = 6 #numero di lanci
c7 = 18 #valore della colonna 7
```

```
#Lancio il dado 100000 volte
```

```
nrep = 100000
ris = numeric(nrep)
for (i in 1:nrep)
{
  x = max(sample(dado,size=lanci,replace=TRUE))
  ris[i] = x
}
```

```
#output della prova
```

```
frq(ris)
```

```
## ris
##      x          n          f
## +---+ +-----+ +-----+
## | 4 |         7 |    0.007 |
## | 5 |        29 |    0.029 |
## | 6 |        44 |    0.044 |
## | 7 |       112 |    0.112 |
## | 8 |       232 |    0.232 |
## | 9 |       385 |    0.385 |
## |10 |       763 |    0.763 |
## |11 |      1199 |    1.199 |
## |12 |      1939 |    1.939 |
## |13 |      2888 |    2.888 |
## |14 |      4230 |    4.230 |
## |15 |      6005 |    6.005 |
## |16 |      8426 |    8.426 |
## |17 |     11516 |   11.516 |
## |18 |     15631 |   15.631 |
## |19 |     20208 |   20.208 |
## |20 |     26386 |   26.386 |
## +---+ +-----+ +-----+
##           100000   100.000
## Osservazioni mancanti: 0
```

*#maggiore o uguale al valore della colonna 7*

```
ok = ris>=c7
out = ris[ok]
frq(out)
```

```
## out
##      x          n          f
## +---+ +-----+ +-----+
## | 18 |     15631 |   25.12013 |
## | 19 |     20208 |   32.47569 |
## | 20 |     26386 |   42.40418 |
## +---+ +-----+ +-----+
##           62225   100.00000
## Osservazioni mancanti: 0
```

*#Probabilità stimata con il metodo Monte Carlo (# successi/# prove) che esca un numero maggiore della colonna 7*

```
sum(ok)/nrep
```

```
## [1] 0.62225
```

*#ESERCIZIO 6*

```
pins = 0.10
crediti = 256
c10 = 27
```

*#valore atteso di insolvenza*  
(vadi = pins\*crediti)

```
## [1] 25.6

#P(X>E) con R e con RMF
pbinom(vadi,crediti,pins,lower.tail=FALSE)

## [1] 0.4971694

Binomiale(crediti,pins,da=round(vadi,0),dettaglio=FALSE,grafico=FALSE)

## -----
## Distribuzione Binomiale
## -----
## Numero delle prove      : 256
## Probabilita' di successo: 0.1
## Valore atteso (media)   : 25.6
## Varianza                : 23.04
## Somma delle probabilita': 0.4971694

#P(X>c10) con R e con RMF
pbinom(c10,crediti,pins,lower.tail=FALSE)

## [1] 0.3375722

Binomiale(crediti,pins,da=c10+1,dettaglio=FALSE,grafico=FALSE)

## -----
## Distribuzione Binomiale
## -----
## Numero delle prove      : 256
## Probabilita' di successo: 0.1
## Valore atteso (media)   : 25.6
## Varianza                : 23.04
## Somma delle probabilita': 0.3375722

#Il numero di insolvenze attese a due anni è 25.6, ma la probabilità che se ne
verifichino di più è di 0.49. Allo stesso tempo, la probabilità che il numero di
insolvenze a due anni sia maggiore di 27 è di 0.34

#ESERCIZIO 7
mi = 6.1
sigma = 9.7
c13 = 16.8

#standardizzazione dei due estremi
(estremo1 = ((mi-sigma)-mi)/sigma)

## [1] -1

(estremo2 = (c13-mi)/sigma)

## [1] 1.103093
```



*#calcolo della probabilità*

```
pnorm(estremo2)-pnorm(estremo1)
```

```
## [1] 0.7063513
```

*#ESERCIZIO 8*

*#generazione del vettore*

```
set.seed(226091); x<- sample(-3:5,size=300,replace=TRUE); set.seed(NULL)
```

*#a) calcolo della mediana*

*frequenze(x,cumul=TRUE) #cerco la modalità che lascia prima di sé il 50% e dopo di sé il 50%*

```
## x
```

```
##      x      n      f      N      F
## +---+---+-----+---+-----+
## | -3 | 36 | 12.000000 | 36 | 12.00000 |
## | -2 | 37 | 12.333333 | 73 | 24.33333 |
## | -1 | 23 | 7.666667 | 96 | 32.00000 |
## |  0 | 33 | 11.000000 | 129 | 43.00000 |
## |  1 | 35 | 11.666667 | 164 | 54.66667 |
## |  2 | 39 | 13.000000 | 203 | 67.66667 |
## |  3 | 30 | 10.000000 | 233 | 77.66667 |
## |  4 | 37 | 12.333333 | 270 | 90.00000 |
## |  5 | 30 | 10.000000 | 300 | 100.00000 |
## +---+---+-----+---+-----+
```

```
## Osservazioni mancanti: 0
```

*#La mediana sarà 1*

```
median(x,na.rm=TRUE) #mediana con i dati grezzi
```

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

*#b) dimostro che la media degli scarti in valore assoluto dalla mediana è minore della media degli scarti in valore assoluto da un qualsiasi altro valore*

*ValoreCasuale = runif(1,-200,200) #generazione del numero casuale*

```
oggetto1 = sum(abs(x-median(x,na.rm=TRUE)))/length(x)
```

```
oggetto2 = sum(abs(x-ValoreCasuale))/length(x)
```

```
0 > oggetto1-oggetto2 #se la dimostrazione è avvenuta correttamente deve risultare "TRUE"
```

```
## [1] TRUE
```

*#ESERCIZIO 9*

*#vettori dei contagiati*

```
contagiatiM =
```

```
c(rep(5,444638),rep(15,667351),rep(25,697747),rep(35,675651),rep(45,791996),rep(55,778178),rep(65,469786),rep(75,305004),rep(85,163360),rep(95,32024))
```

```
(length(contagiatiM)) #devono essere 5025735
```

```
## [1] 5025735
```

```

contagiatiF=
c(rep(5,413787),rep(15,644080),rep(25,694093),rep(35,743312),rep(45,903908),rep(55,827714),rep(65,466019),rep(75,309141),rep(85,229827),rep(95,93094))
(length(contagiatiF)) #devono essere 5324975

## [1] 5324975

contagiatiT =
c(rep(5,444638+413787),rep(15,667351+644080),rep(25,697747+694093),rep(35,675651+743312),rep(45,791996+903908),rep(55,778178+827714),rep(65,469786+466019),rep(75,305004+309141),rep(85,163360+229827),rep(95,32024+93094))
length(contagiatiT) #devono essere 10350710

## [1] 10350710

#vettori dei morti
mortiM =
c(rep(5,7),rep(15,15),rep(25,61),rep(35,222),rep(45,948),rep(55,3808),rep(65,10785),rep(75,24300),rep(85,31274),rep(95,9870))
(length(mortiM)) #devono essere 81290

## [1] 81290

mortiF =
c(rep(5,10),rep(15,13),rep(25,34),rep(35,128),rep(45,429),rep(55,1535),rep(65,4278),rep(75,11978),rep(85,26480),rep(95,18144))
(length(mortiF)) #devono essere 63029

## [1] 63029

mortiT =
c(rep(5,7+10),rep(15,15+13),rep(25,61+34),rep(35,222+128),rep(45,948+429),rep(55,3808+1535),rep(65,10785+4278),rep(75,24300+11978),rep(85,31274+26480),rep(95,9870+18144))
(length(mortiT)) #devono essere 144319

## [1] 144319

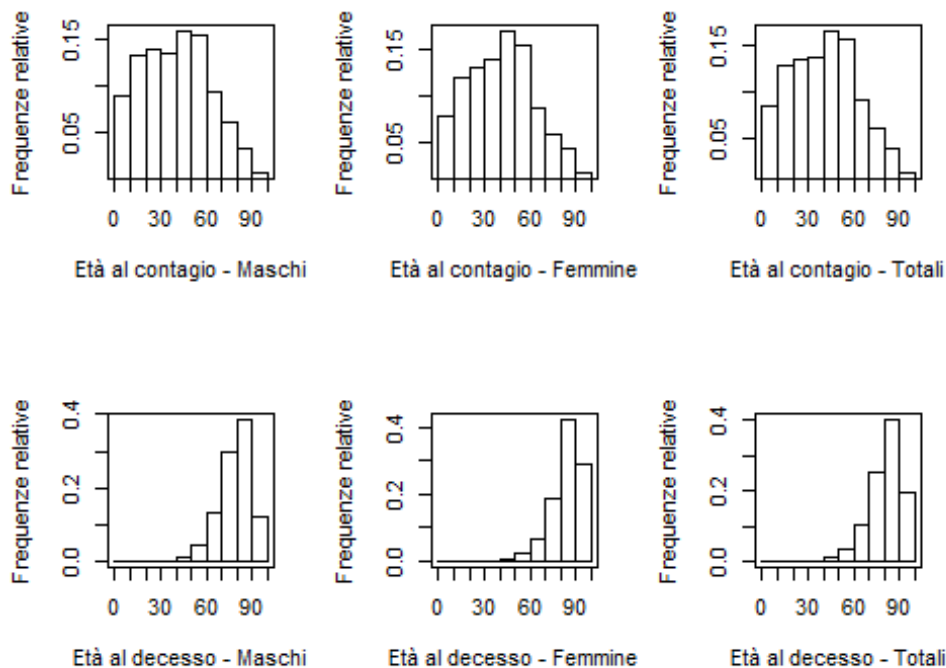
#per avere tutti i grafici in un'unica schermata per comodità:
par(mfrow=c(2,3))

#istogramma per la variabile "età al contagio"
istogramma(contagiatiM,da=0,a=100,nclassi=10, stampa=FALSE, nome="Età al contagio - Maschi")
istogramma(contagiatiF,da=0,a=100,nclassi=10, stampa=FALSE, nome="Età al contagio - Femmine")
istogramma(contagiatiT,da=0,a=100,nclassi=10, stampa=FALSE, nome="Età al contagio - Totali")

#istogramma per la variabile età al decesso"
istogramma(mortiM,da=0,a=100,nclassi=10, stampa=FALSE, nome="Età al decesso - Maschi")
istogramma(mortiF,da=0,a=100,nclassi=10, stampa=FALSE, nome="Età al decesso - Femmine")

```

```
istogramma(mortiT,da=0,a=100,nclassi=10, stampa=FALSE, nome="Età al decesso - Totali")
```



*#Osservando i grafici è possibile notare come le classi d'età maggiormente contagiate sono quelle centrali, mentre gli anziani hanno probabilità maggiore di morire se incontrano il virus.*

```
#chiudo la finestra multipla dei grafici
par(mfrow=c(1,1))
```

```
#media, mediana e sd per i maschi per la variabile "età al contagio"
(miCM = media(contagiatiM))
```

```
## [1] 40.21378
```

```
frequenze(contagiatiM,cumul=TRUE) #la classe mediana è 40-49
```

```
## x
## x n f N F
## +---+---+---+---+
## | 5 | 444638 | 8.8472233 | 444638 | 8.847223 |
## | 15 | 667351 | 13.2786747 | 1111989 | 22.125898 |
## | 25 | 697747 | 13.8834817 | 1809736 | 36.009380 |
## | 35 | 675651 | 13.4438246 | 2485387 | 49.453204 |
## | 45 | 791996 | 15.7588094 | 3277383 | 65.212014 |
## | 55 | 778178 | 15.4838645 | 4055561 | 80.695878 |
## | 65 | 469786 | 9.3476079 | 4525347 | 90.043486 |
## | 75 | 305004 | 6.0688437 | 4830351 | 96.112330 |
```

```
## | 85 | 163360 | 3.2504698 | 4993711 | 99.362800 |
## | 95 | 32024 | 0.6372003 | 5025735 | 100.000000 |
## +---+-----+-----+-----+-----+
## Osservazioni mancanti: 0

(medCM = 40+((0.5-0.49453204)/(0.65212014-0.49453204))*10)

## [1] 40.34698

(sqmCM = sd(contagiatiM, na.rm=TRUE))

## [1] 21.96736

#media, mediana e sd per i maschi per la variabile "età al decesso"
(miMM = media(mortiM))

## [1] 78.49822

frequenze(mortiM,cumul=TRUE) #la classe mediana è 70-79

## x
## x n f N F
## +---+-----+-----+-----+-----+
## | 5 | 7 | 8.611145e-03 | 7 | 8.611145e-03 |
## | 15 | 15 | 1.845245e-02 | 22 | 2.706360e-02 |
## | 25 | 61 | 7.503998e-02 | 83 | 1.021036e-01 |
## | 35 | 222 | 2.730963e-01 | 305 | 3.751999e-01 |
## | 45 | 948 | 1.166195e+00 | 1253 | 1.541395e+00 |
## | 55 | 3808 | 4.684463e+00 | 5061 | 6.225858e+00 |
## | 65 | 10785 | 1.326731e+01 | 15846 | 1.949317e+01 |
## | 75 | 24300 | 2.989298e+01 | 40146 | 4.938615e+01 |
## | 85 | 31274 | 3.847214e+01 | 71420 | 8.785829e+01 |
## | 95 | 9870 | 1.214171e+01 | 81290 | 1.000000e+02 |
## +---+-----+-----+-----+-----+
## Osservazioni mancanti: 0

(medMM = 70+((0.5-0.1949317)/(0.4938615-0.1949317))*10)

## [1] 80.20535

(sqmMM = sd(mortiM, na.rm=TRUE))

## [1] 11.16404

#media, mediana e sd per Le femmine per la variabile "età al contagio"
(miCF = media(contagiatiF))

## [1] 41.90715

frequenze(contagiatiF,cumul=TRUE) #la classe mediana è 40-49

## x
## x n f N F
## +---+-----+-----+-----+-----+
## | 5 | 413787 | 7.770684 | 413787 | 7.770684 |
```

```
## | 15 | 644080 | 12.095456 | 1057867 | 19.866140 |
## | 25 | 694093 | 13.034672 | 1751960 | 32.900812 |
## | 35 | 743312 | 13.958976 | 2495272 | 46.859788 |
## | 45 | 903908 | 16.974878 | 3399180 | 63.834666 |
## | 55 | 827714 | 15.543998 | 4226894 | 79.378664 |
## | 65 | 466019 | 8.751572 | 4692913 | 88.130235 |
## | 75 | 309141 | 5.805492 | 5002054 | 93.935727 |
## | 85 | 229827 | 4.316020 | 5231881 | 98.251748 |
## | 95 | 93094 | 1.748252 | 5324975 | 100.000000 |
## +---+-----+-----+-----+-----+-----+
## Osservazioni mancanti: 0
```

```
(medCF = 40+((0.5-0.46859788)/(0.63834666-0.46859788))*10)
```

```
## [1] 41.84992
```

```
(sqmCF = sd(contagiatiF, na.rm=TRUE))
```

```
## [1] 22.47853
```

```
#media, mediana e sd per le femmine per la variabile "età al decesso"
```

```
(miMF = media(mortiF))
```

```
## [1] 83.4569
```

```
frequenze(mortiF,cumul=TRUE) #la classe mediana è 80-89
```

```
## x
```

```
##      x      n      f      N      F
## +---+-----+-----+-----+-----+
## | 5 | 10 | 0.01586571 | 10 | 0.01586571 |
## | 15 | 13 | 0.02062543 | 23 | 0.03649114 |
## | 25 | 34 | 0.05394342 | 57 | 0.09043456 |
## | 35 | 128 | 0.20308112 | 185 | 0.29351568 |
## | 45 | 429 | 0.68063907 | 614 | 0.97415475 |
## | 55 | 1535 | 2.43538689 | 2149 | 3.40954164 |
## | 65 | 4278 | 6.78735185 | 6427 | 10.19689349 |
## | 75 | 11978 | 19.00395056 | 18405 | 29.20084406 |
## | 85 | 26480 | 42.01240699 | 44885 | 71.21325104 |
## | 95 | 18144 | 28.78674896 | 63029 | 100.00000000 |
## +---+-----+-----+-----+-----+
## Osservazioni mancanti: 0
```

```
(medMF = 80+((0.5-0.2920084406)/(0.7121325104-0.2920084406))*10)
```

```
## [1] 84.95072
```

```
(sqmMF = sd(mortiF, na.rm=TRUE))
```

```
## [1] 10.69647
```

```
#media, mediana e sd per la totalità dei dati per la variabile "età al contagio"
```

```
(miCT = media(contagiatiT))
```

```
## [1] 41.08494
```

```
frequenze(contagiatiT,cumul=TRUE) #la classe mediana è 40-49
```

```
## x
```

| ## | x  | n       | f         | N        | F          |
|----|----|---------|-----------|----------|------------|
| ## | 5  | 858425  | 8.293392  | 858425   | 8.293392   |
| ## | 15 | 1311431 | 12.669962 | 2169856  | 20.963354  |
| ## | 25 | 1391840 | 13.446807 | 3561696  | 34.410161  |
| ## | 35 | 1418963 | 13.708847 | 4980659  | 48.119008  |
| ## | 45 | 1695904 | 16.384422 | 6676563  | 64.503430  |
| ## | 55 | 1605892 | 15.514800 | 8282455  | 80.018231  |
| ## | 65 | 935805  | 9.040974  | 9218260  | 89.059205  |
| ## | 75 | 614145  | 5.933361  | 9832405  | 94.992566  |
| ## | 85 | 393187  | 3.798648  | 10225592 | 98.791213  |
| ## | 95 | 125118  | 1.208787  | 10350710 | 100.000000 |

```
## Osservazioni mancanti: 0
```

```
(medCT = 40+((0.5-0.48119008)/(0.64503430-0.48119008))*10)
```

```
## [1] 41.14804
```

```
(sqmCT = sd(contagiatiT, na.rm=TRUE))
```

```
## [1] 22.24791
```

```
#media, mediana e sd per la totalità dei dati per la variabile "età al decesso"  
(miMT = media(mortiT))
```

```
## [1] 80.66384
```

```
frequenze(mortiT,cumul=TRUE) #la classe mediana è 80-89
```

```
## x
```

| ## | x  | n     | f           | N      | F            |
|----|----|-------|-------------|--------|--------------|
| ## | 5  | 17    | 0.01177946  | 17     | 0.01177946   |
| ## | 15 | 28    | 0.01940146  | 45     | 0.03118093   |
| ## | 25 | 95    | 0.06582640  | 140    | 0.09700732   |
| ## | 35 | 350   | 0.24251831  | 490    | 0.33952563   |
| ## | 45 | 1377  | 0.95413632  | 1867   | 1.29366196   |
| ## | 55 | 5343  | 3.70221523  | 7210   | 4.99587719   |
| ## | 65 | 15063 | 10.43729516 | 22273  | 15.43317235  |
| ## | 75 | 36278 | 25.13736930 | 58551  | 40.57054165  |
| ## | 85 | 57754 | 40.01829281 | 116305 | 80.58883446  |
| ## | 95 | 28014 | 19.41116554 | 144319 | 100.00000000 |

```
## Osservazioni mancanti: 0
```

```
(medMT = 80+((0.5-0.4057054165)/(0.8058883446-0.4057054165))*10)
```

```
## [1] 82.35629
```

```
(sqmMT = sd(mortiT, na.rm=TRUE))
```

```
## [1] 11.23475
```

*#riassunto per medie e mediane per le variabili "età al contagio" e "età al decesso"*

```
MisuraDiSintesi=c("media - maschi","mediana - maschi","media - femmine","mediana - femmine","media - totale","mediana - totale")
```

```
EtàAlContagio=round(c(miCM,medCM,miCF,medCF,miCT,medCT),4)
```

```
EtàAlDecesso=round(c(miMM,medMM,miMF,medMF,miMT,medMT),4)
```

```
cbind(MisuraDiSintesi,EtàAlContagio,EtàAlDecesso)
```

| ##      | MisuraDiSintesi     | EtàAlContagio | EtàAlDecesso |
|---------|---------------------|---------------|--------------|
| ## [1,] | "media - maschi"    | "40.2138"     | "78.4982"    |
| ## [2,] | "mediana - maschi"  | "40.347"      | "80.2053"    |
| ## [3,] | "media - femmine"   | "41.9072"     | "83.4569"    |
| ## [4,] | "mediana - femmine" | "41.8499"     | "84.9507"    |
| ## [5,] | "media - totale"    | "41.0849"     | "80.6638"    |
| ## [6,] | "mediana - totale"  | "41.148"      | "82.3563"    |

*#Come si evince dal prospetto di sintesi sopra riportato media e mediana, per la variabile "età al contagio" sono molto simili, a riprova del fatto che assume una forma tendenzialmente campanulare con una asimmetria destra*

*#Discorso simile anche per la variabile "età al decesso", che presenta anch'essa media e mediana appartenenti alla stessa classe. L'istogramma evidenzia però un'asimmetria destra pronunciata, mettendo in luce che con un numero basso di contagiati, in proporzione alle altre classi, gli anziani hanno un tasso di mortalità nettamente più elevato*

*#Perché l'età mediana al contagio ottenuta eseguendo lo script (41.15) è diversa da quella fornita dall'Istituto Superiore di Sanità (47)? Questa discrepanza tra dati è dovuta al fatto che l'ISS utilizza i dati grezzi per calcolare le misure di sintesi, mentre lo script sopra eseguito fa riferimento a una distribuzione per classi (quindi minore contenuto informativo). Entrambi i dati ricadono infatti nella classe medesima mediana 40-49.*