C12-Lab

filippo

May 3, 2018

Contents

(a) Si estraggano tutte le rilevazioni giornaliere rispetto al mese di aprile 1
(b) Si analizzino i dati per evidenziare se vi sono dati anomali. 1
(c) Si adatti una distribuzione parametrica ai dati chiarendo le ipotesi sottostanti. 2
(d) Si confronti la distribuzione parametrica stimata con la distribuzione empirica. 2
(e) Si considerino ora le medie mensili e si ripeta l'analisi 4
(f) Quale teorema della probabilità entra in gioco ? 6


```
## Loading required package: MASS
```

```
## Loading required package: survival
```

12. Il file KNMI_20160831.txt contiene le precipitazioni giornaliere rilevate dal 1901 al 2016 nella stazione di De Bilt (Olanda)

(a) Si estraggano tutte le rilevazioni giornaliere rispetto al mese di aprile

```
head(de_bilt)
```

```
##   station yyyymmdd rainfall
## 1     260 19010101       NA
## 2     260 19010102       NA
## 3     260 19010103       NA
## 4     260 19010104       NA
## 5     260 19010105       NA
## 6     260 19010106       NA
```

```
month<-substr(de_bilt$yyyymmdd, 5, 6)
sel<-month=="04"
y<-de_bilt$rainfall[sel]

summary(y)
```

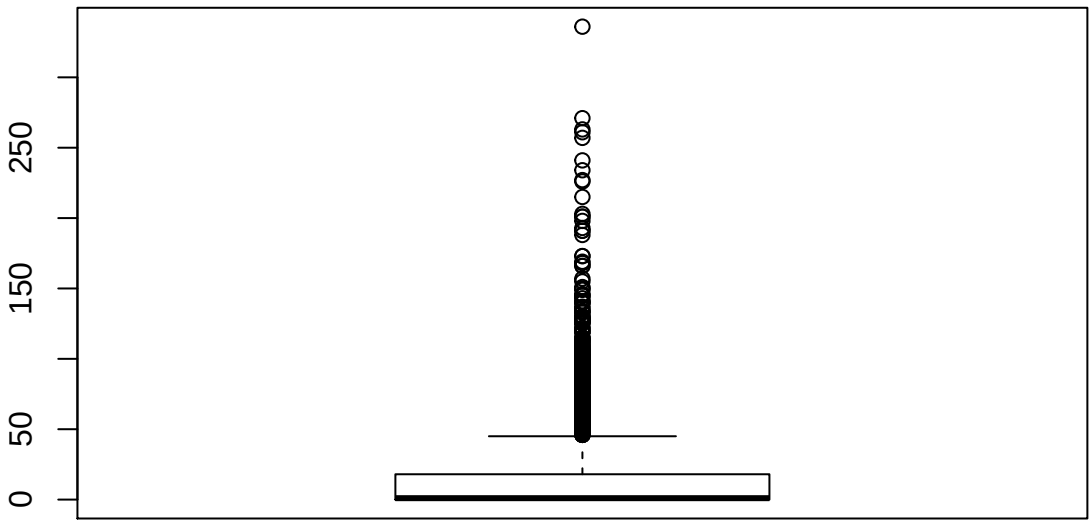
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   -1.00    0.00    1.00   15.87   18.00  336.00     180
```

```
y[y < 0]<-0
summary(y)
```

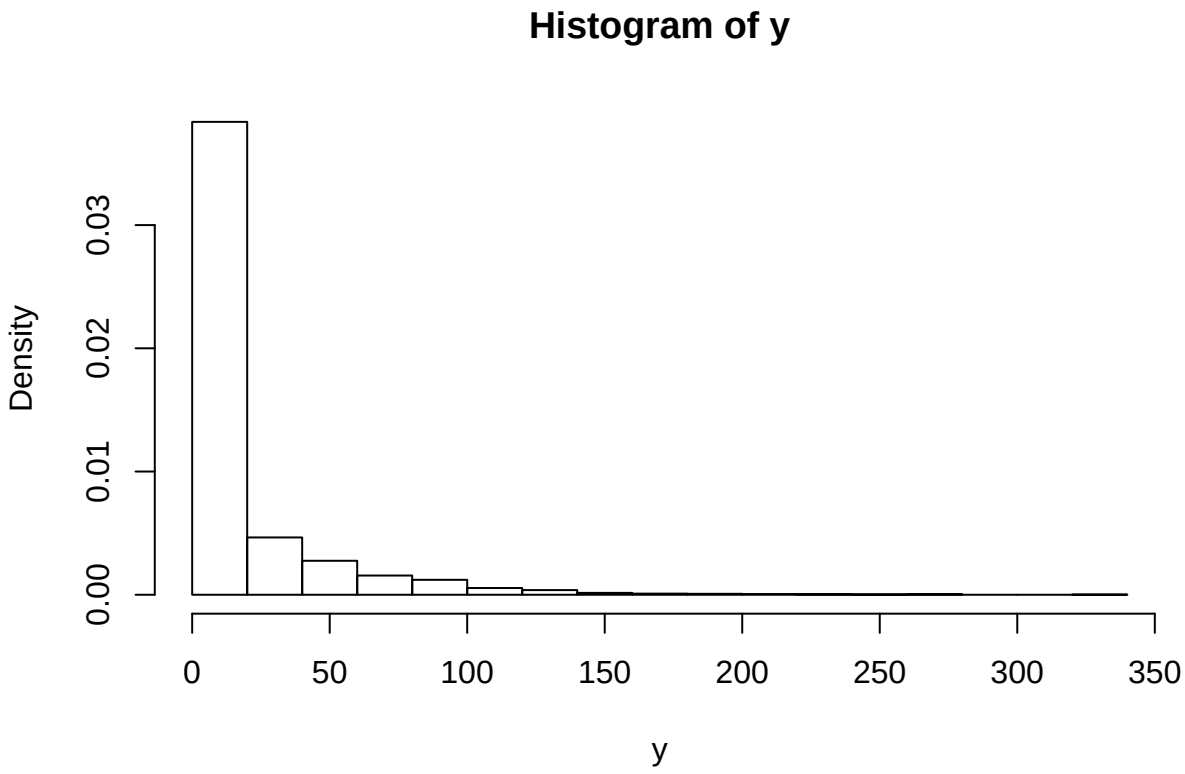
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00    0.00    1.00   15.95   18.00  336.00     180
```

(b) Si analizzino i dati per evidenziare se vi sono dati anomali.

```
boxplot(y)
```




```
hist(y, freq = FALSE)
```

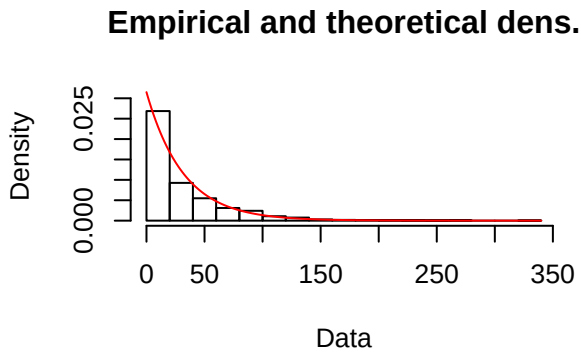


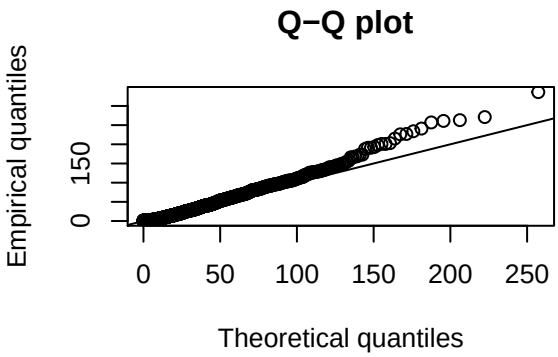

```
x<-y[y>0]
```

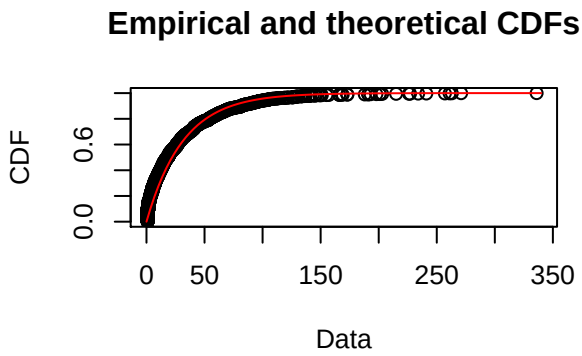
(c) Si adatti una distribuzione parametrica ai dati chiarendo le ipotesi sottostanti.

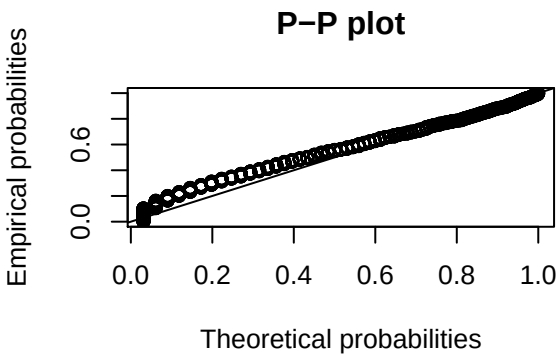
(d) Si confronti la distribuzione parametrica stimata con la distribuzione empirica.

```
fitmle<-fitdist(as.numeric(na.omit(x)), distr = "exp")
plot(fitmle)
```

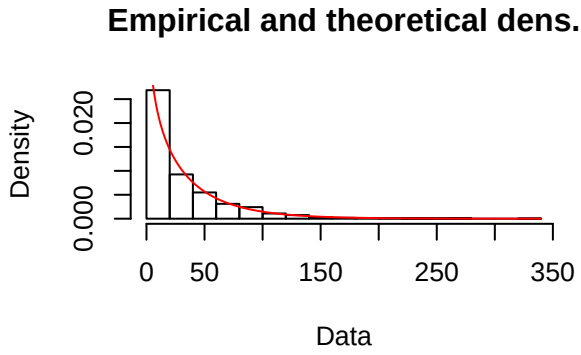


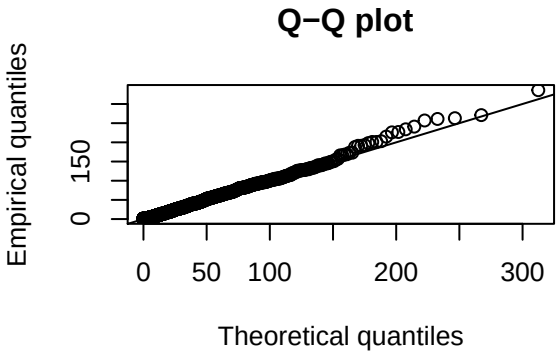


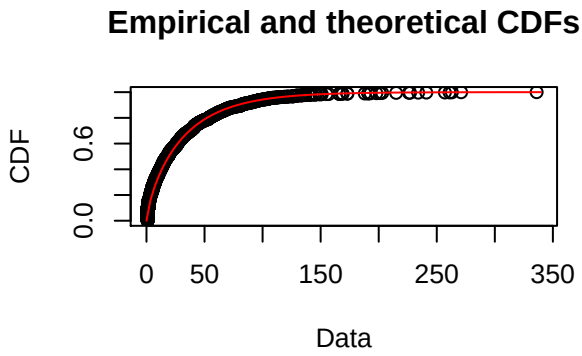


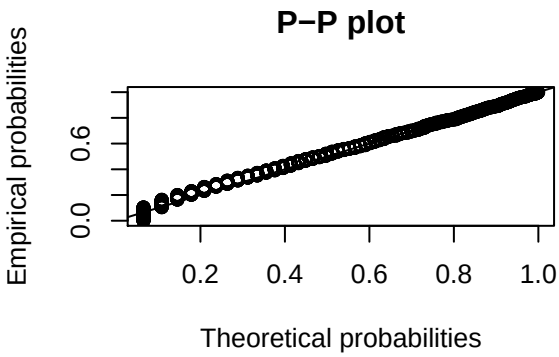



```
fitmlg<-fitdist(as.numeric(na.omit(x)), distr = "gamma")
plot(fitmlg)
```









(e) Si considerino ora le medie mensili e si ripeta l'analisi

```
mat <- matrix(y,nrow=30)
length(y)/30
```

```
## [1] 116
```

```
apply(mat,2,mean)
```

```
##   [1]        NA        NA        NA        NA        NA  9.833333 13.800000
##   [8] 11.900000 31.100000 23.400000  9.866667 12.966667  6.466667 13.533333
##  [15] 14.366667 26.800000 15.133333  9.366667 22.500000 29.700000  9.766667
##  [22] 22.966667 10.733333 17.300000 19.200000 12.466667 27.533333 15.366667
##  [29] 16.000000  7.833333 19.766667 28.333333  6.366667 14.466667 30.933333
##  [36] 17.400000 19.933333 12.100000 22.900000 22.300000 11.333333  4.600000
##  [43] 10.500000 10.933333        NA  5.533333 16.300000 19.700000 22.300000
##  [50] 25.200000 26.333333 10.533333 18.566667  5.166667 11.766667 11.733333
##  [57]  5.700000 17.766667 25.866667  9.900000 22.033333 26.566667 15.733333
##  [64] 16.933333 35.933333 24.066667 15.366667 11.533333 21.500000 30.700000
##  [71]  6.800000 20.633333 23.700000  3.333333 20.466667  3.266667 20.566667
##  [78] 12.733333 26.600000 17.866667  5.366667  8.566667 28.333333  3.666667
##  [85] 22.633333  8.933333 15.266667  4.233333 24.400000 17.433333  9.666667
##  [92] 17.333333 14.966667 22.466667  6.400000  2.700000 10.300000 32.500000
##  [99] 20.666667 13.066667 29.133333 16.400000 15.400000 11.033333 20.966667
## [106] 13.266667  0.100000 11.300000  6.633333 10.266667  2.733333 15.833333
## [113]  7.666667 19.466667  6.966667 27.533333
```

```
#colMeans(mat)

#lines(colMeans(mat))
```

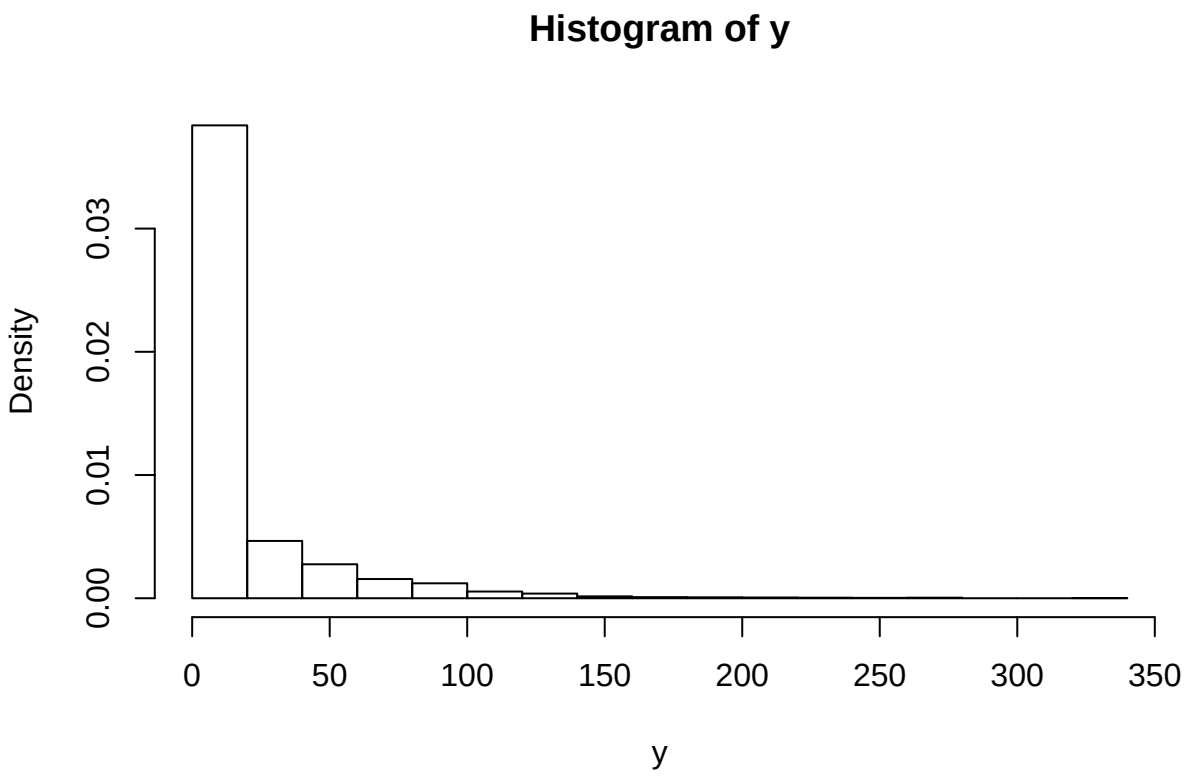
Analizziamo i dati mancanti

```
apply(is.na(mat),2,sum)
```

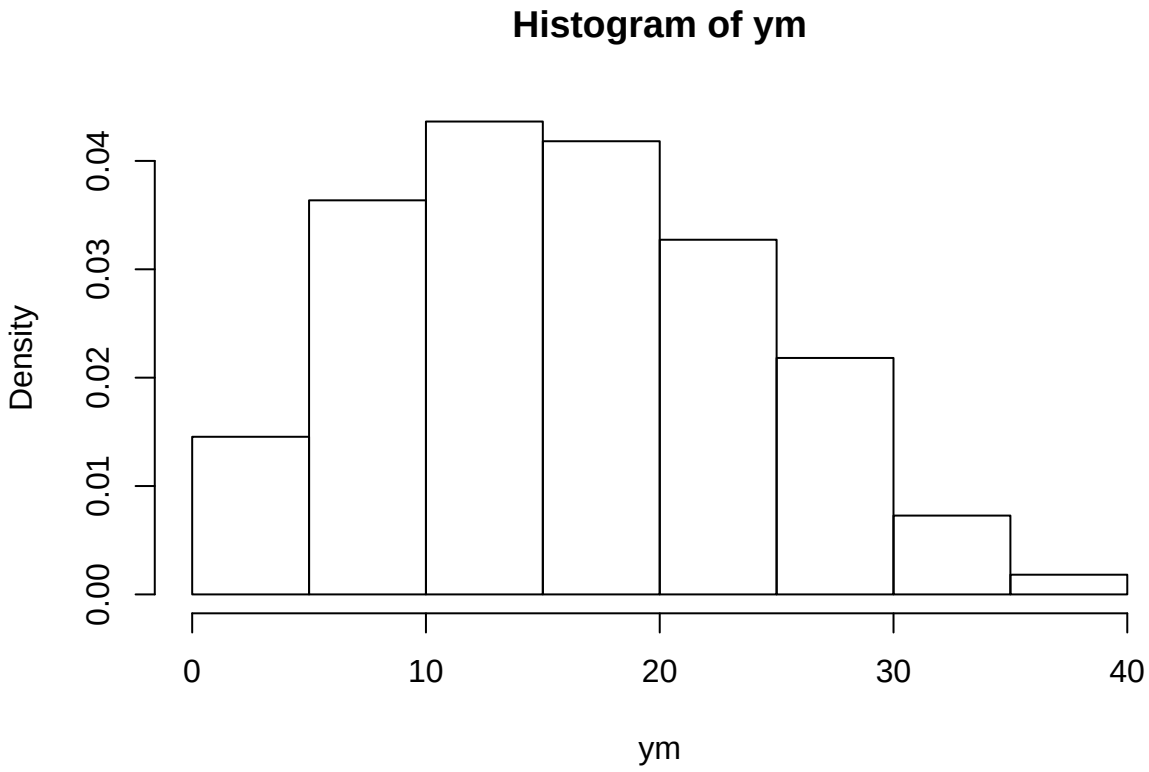
```
##   [1] 30 30 30 30 30  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  [24]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 30  0
##  [47]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  [70]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  [93]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [116]  0
```

```
ym <- apply(mat,2,mean)

hist(y,prob=TRUE)
```



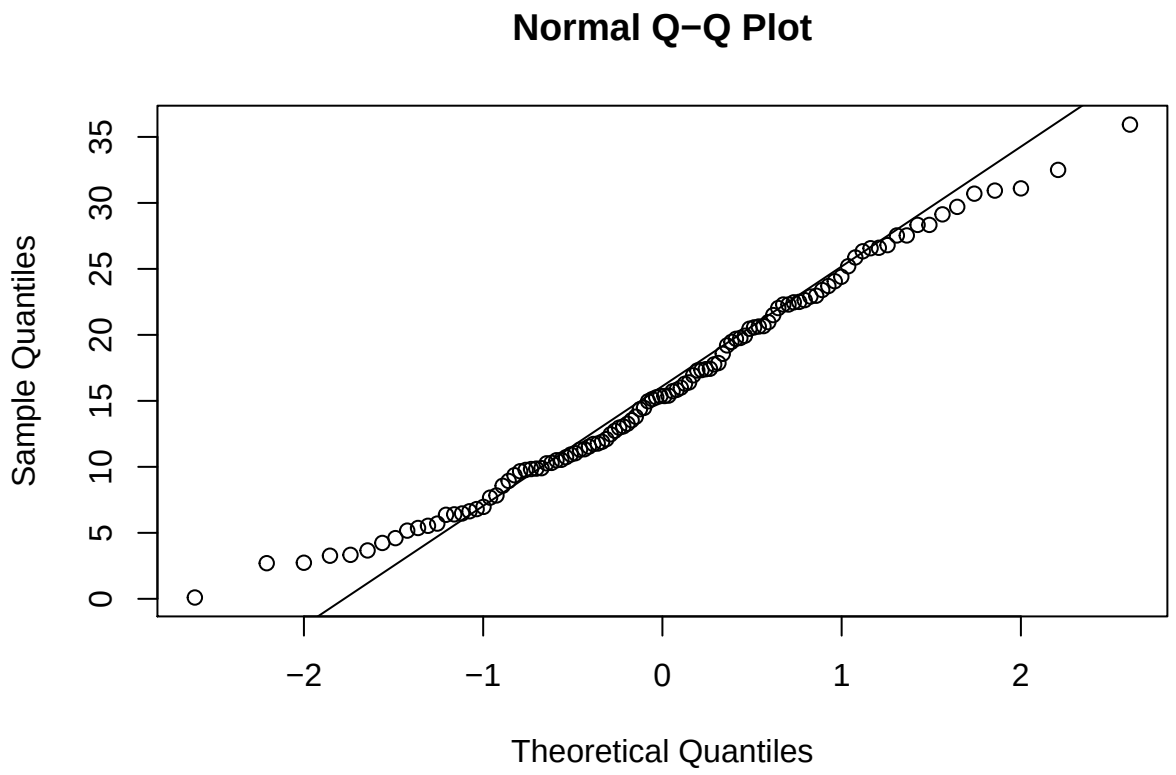

```
hist(ym,prob=TRUE)
```



(f) Quale teorema della probabilità entra in gioco ?

Il teorema che entra in gioco è quello del limite centrale

```
qqnorm(ym)
qqline(ym)
```

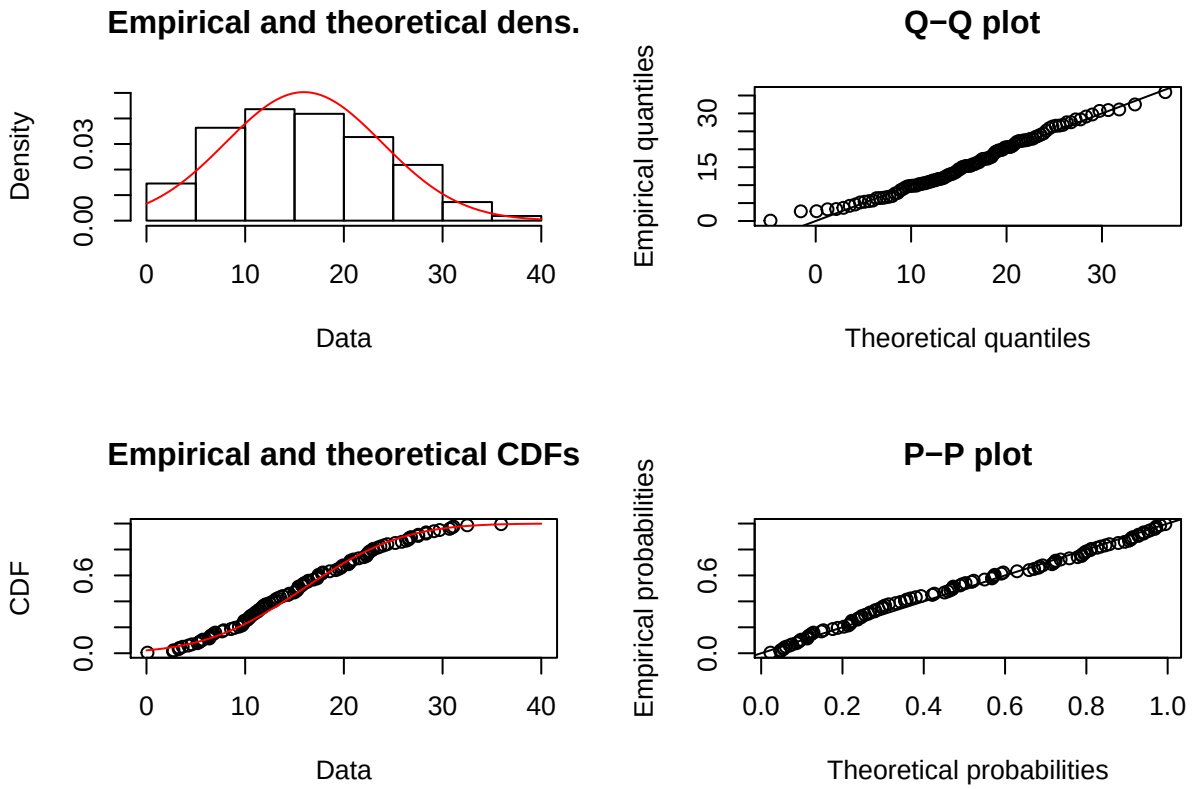


Fare lo stesso con ggplot

```
fitnrm<-fitdist(as.numeric(na.omit(ym)), distr = "norm")
fitnrm
```

```
## Fitting of the distribution ' norm ' by maximum likelihood
## Parameters:
##       estimate Std. Error
## mean 15.946061  0.7561135
## sd    7.930185  0.5346529
```

```
plot(fitnrm)
```



Mi creo lo stesso grafico in alto a sx

```
hist(ym,prob=TRUE,main="mio grafico",ylim = c(0,0.05))
curve(dnorm(x, mean=fitnrm$estimate[1], sd=fitnrm$estimate[2]),from = min(na.omit(ym)), to = max(na.omi
```