Partiamo con una panoramica rapida del dataset.

library(ISwR)  
source("./../functions.R", local = knitr::knit\_global())  
data("bp.obese")  
  
bp.obese$sex <- factor(bp.obese$sex, labels = c("male", "female"))  
str(bp.obese)

## 'data.frame': 102 obs. of 3 variables:  
## $ sex : Factor w/ 2 levels "male","female": 1 1 1 1 1 1 1 1 1 1 ...  
## $ obese: num 1.31 1.31 1.19 1.11 1.34 1.17 1.56 1.18 1.04 1.03 ...  
## $ bp : int 130 148 146 122 140 146 132 110 124 150 ...

summary(bp.obese)

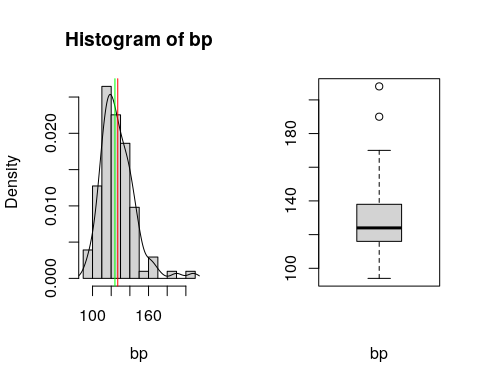
## sex obese bp   
## male :44 Min. :0.810 Min. : 94.0   
## female:58 1st Qu.:1.143 1st Qu.:116.0   
## Median :1.285 Median :124.0   
## Mean :1.313 Mean :127.0   
## 3rd Qu.:1.430 3rd Qu.:137.5   
## Max. :2.390 Max. :208.0

Il dataset è costituito da tre variabili, due numeriche continue (obese e bp) e una categoriale non ordinale (sex).

La variabile risposta è bp e lo scopo dell’analisi è fornire un modello di regressione per stabilire come il valore della pressione sistolica delle persone vari a seconda del loro indice di obesità e del loro sesso.

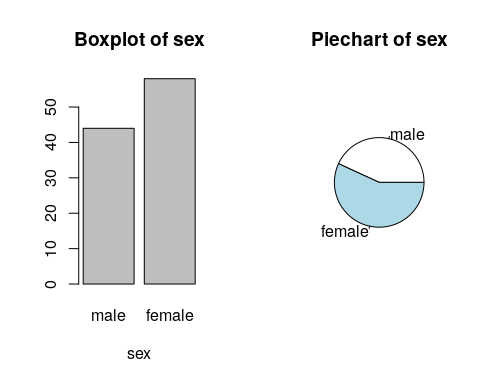
### Distribuzione delle variabili

numeric.graph(bp.obese, c(2, 3))



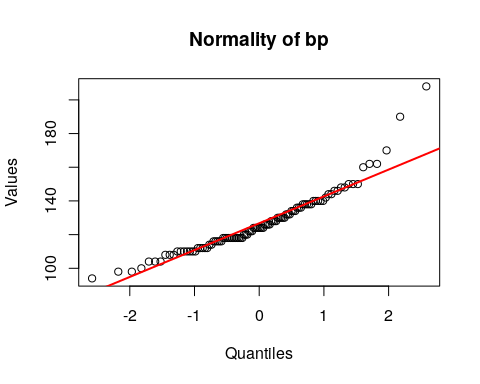
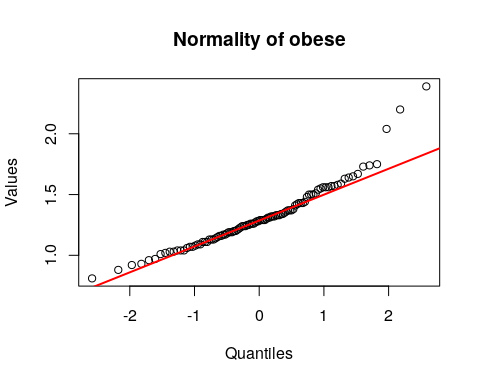
## [1] "Skewness indexes"  
## obese bp   
## 1.260221 1.406322   
## [1] "Kurtosis indexes - 3"  
## obese bp   
## 3.085016 3.740396

categorical.graph(bp.obese, c(1))

 #### Osservazioni

* C’è una marcata assimmetria verso destra delle variabili bp e obese
* Le curve di bp e obese sono fortemente ipernormali, avendo entrambi delle curve destre molto pesanti, come è possibile vedere dai qqplot delle due variabili:

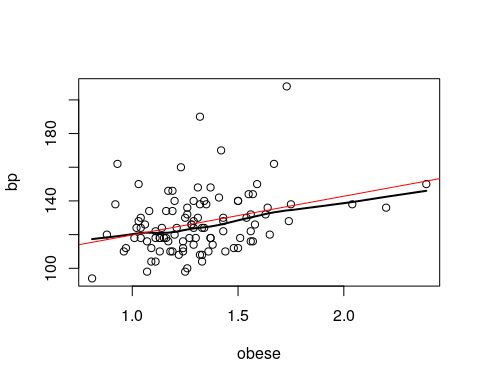
qqplot.graph(bp.obese, c(2,3))



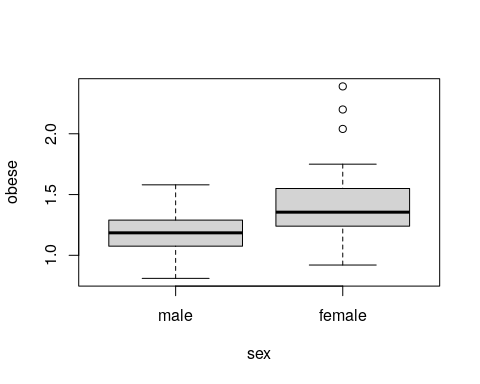
* Per questi motivi, a meno di trasformazioni delle due variabili, le ipotesi di normalità sono violate.
* In bp e in obese sono presenti degli outliers sui valori massimi.
* Le osservazioni sembrano essere ben distribuite tra uomini e donne

### Relazioni tra variabili

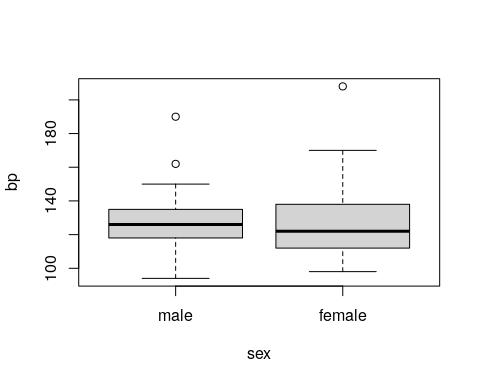
scatterplots.graph(response.var = 3, esplicative.vars = c(2), data = bp.obese)



boxplots.graph(response.var = 2, esplicative.vars = c(1), data = bp.obese)



boxplots.graph(response.var = 3, esplicative.vars = c(1), data = bp.obese)

 #### Osservazioni

* Tra bp e obese sembra esserci una debole correlazione lineare.
* Sembra esserci correlazione tra obese e sex, seppur molto debole.
* I tre ouliers di obese sono tutti riferiti a soggetti di sesso femminile, come confermato dal secondo boxplot. Considerata anche la distribuzione tra soggetti maschili e femminili nel dataset, potrebbe essere una buona idea rimuovere i tre outlier dalle osservazioni.
* Non sembra esserci una correlazione significativa tra bp e sex.