Matrice dei dati centrati e standardizzati

Si consideri la seguente matrice `X` di dimensioni 10×2

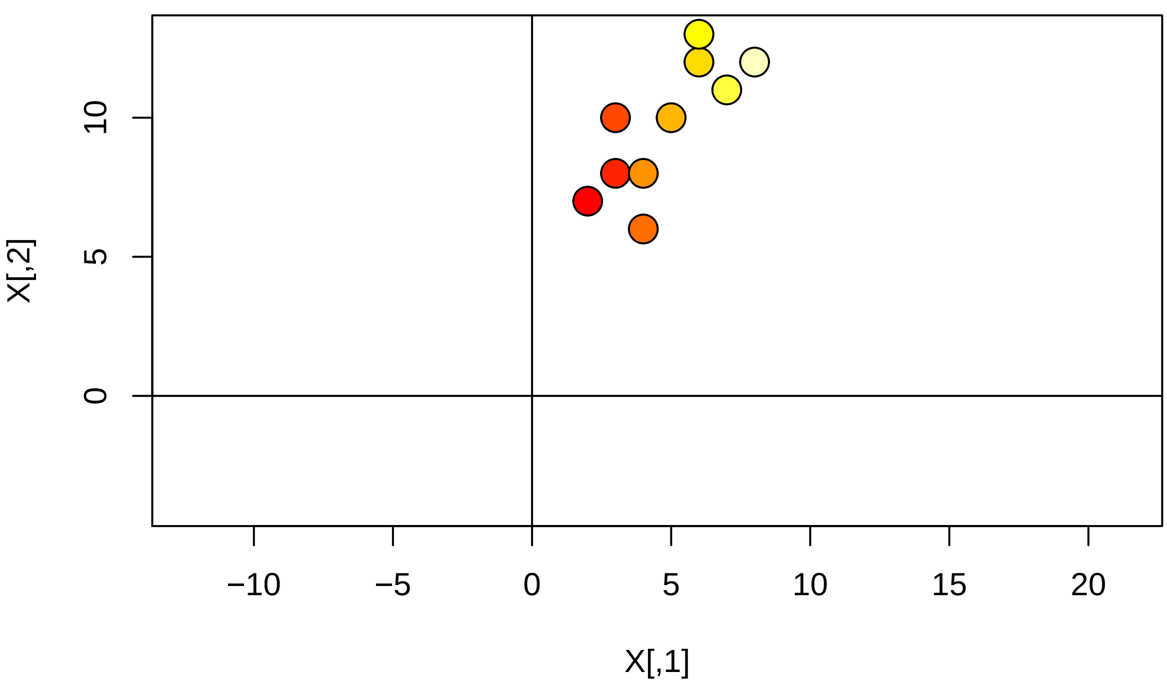
```
n = 10
p = 2
X <- matrix(c(2,3,3,4,4,5,6,6,7,8,7,8,10,6,8,10,12,13,11,12),nrow=n,ncol=p)
X
```

```
##       [,1] [,2]
##  [1,]    2    7
##  [2,]    3    8
##  [3,]    3   10
##  [4,]    4    6
##  [5,]    4    8
##  [6,]    5   10
##  [7,]    6   12
##  [8,]    6   13
##  [9,]    7   11
## [10,]    8   12
```

1. Costruire il diagramma di dispersione per X

- limitare valori sull'asse delle x e delle y all'intervallo [-4,13], mantenendo le proporzioni dei due assi (argomento `asp=1`)
- colorare le unità statistiche e ingrandirle (`bg=heat.colors(n), pch=21, cex=2`)
- aggiungere gli assi $x = 0$ e $y = 0$

```
# diagramma di dispersione per X
plot(X,xlim=c(-4,13),ylim=c(-4,13),asp=1, bg=heat.colors(n),pch=21,cex=2)
abline(h=0)
abline(v=0)
```



2. Ottenere il vettore delle medie $\bar{x} = \frac{1}{n}X'1$:

```r
# vettore di 1
one.n <- matrix(rep(1,n),ncol=1)
one.n
```

```
##       [,1]
##  [1,]    1
##  [2,]    1
##  [3,]    1
##  [4,]    1
##  [5,]    1
##  [6,]    1
##  [7,]    1
##  [8,]    1
##  [9,]    1
## [10,]    1
```

```r
# vettore delle medie
xbar <- (1/n) * t(X) %*% one.n
xbar
```

```
##      [,1]
## [1,]  4.8
## [2,]  9.7
```

3. Ottenere la matrice di centramento $H = I - \frac{1}{n}11'$, verificandone la simmetria e la proprietà di idempotenza

```r
# matrice identità
I.n <- diag(rep(1,n))
I.n
```

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
##  [1,]    1    0    0    0    0    0    0    0    0     0
##  [2,]    0    1    0    0    0    0    0    0    0     0
##  [3,]    0    0    1    0    0    0    0    0    0     0
##  [4,]    0    0    0    1    0    0    0    0    0     0
##  [5,]    0    0    0    0    1    0    0    0    0     0
##  [6,]    0    0    0    0    0    1    0    0    0     0
##  [7,]    0    0    0    0    0    0    1    0    0     0
##  [8,]    0    0    0    0    0    0    0    1    0     0
##  [9,]    0    0    0    0    0    0    0    0    1     0
## [10,]    0    0    0    0    0    0    0    0    0     1
```

```r
# matrice di centramento
H <- I.n - (1/n) * one.n %*% t(one.n)
H
```

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
##  [1,]  0.9 -0.1 -0.1 -0.1 -0.1 -0.1 -0.1 -0.1 -0.1  -0.1
##  [2,] -0.1  0.9 -0.1 -0.1 -0.1 -0.1 -0.1 -0.1 -0.1  -0.1
##  [3,] -0.1 -0.1  0.9 -0.1 -0.1 -0.1 -0.1 -0.1 -0.1  -0.1
##  [4,] -0.1 -0.1 -0.1  0.9 -0.1 -0.1 -0.1 -0.1 -0.1  -0.1
##  [5,] -0.1 -0.1 -0.1 -0.1  0.9 -0.1 -0.1 -0.1 -0.1  -0.1
##  [6,] -0.1 -0.1 -0.1 -0.1 -0.1  0.9 -0.1 -0.1 -0.1  -0.1
##  [7,] -0.1 -0.1 -0.1 -0.1 -0.1 -0.1  0.9 -0.1 -0.1  -0.1
##  [8,] -0.1 -0.1 -0.1 -0.1 -0.1 -0.1 -0.1  0.9 -0.1  -0.1
##  [9,] -0.1 -0.1 -0.1 -0.1 -0.1 -0.1 -0.1 -0.1  0.9  -0.1
```

```
## [10,] -0.1 -0.1 -0.1 -0.1 -0.1 -0.1 -0.1 -0.1 -0.1   0.9
```

```
# simmetria t(H) = H
sum( t(H) - H )
```

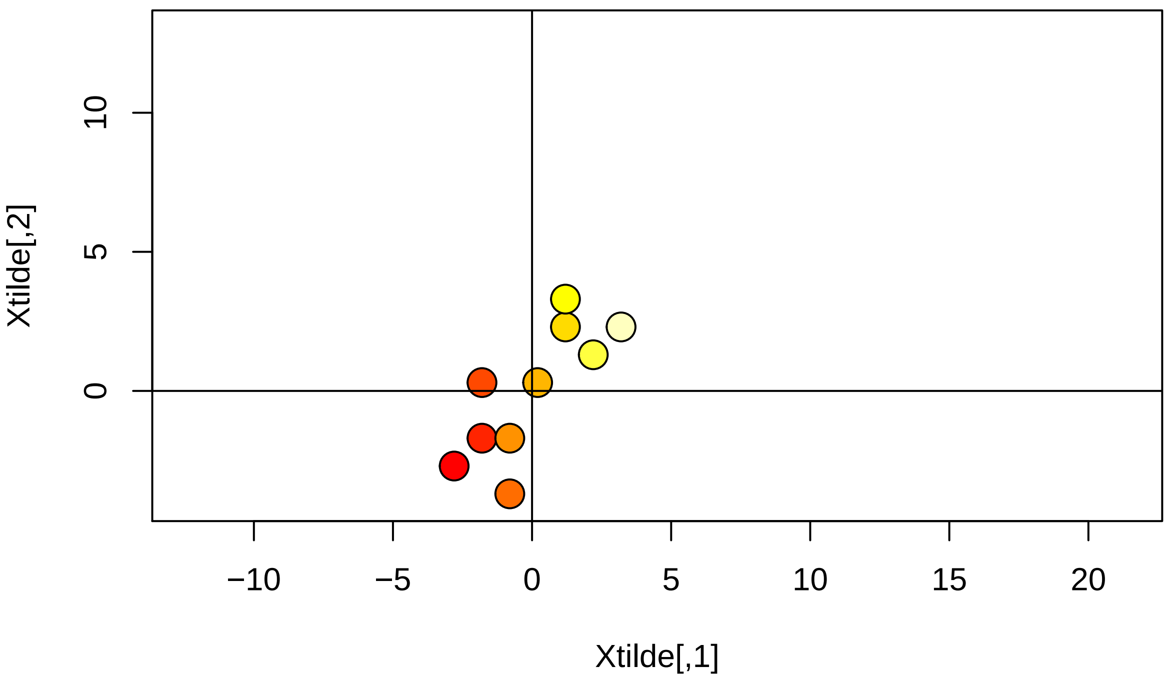
```
## [1] 0
```

```
# idempotenza HH = H
sum( H %*% H - H )
```

```
## [1] 2.331468e-15
```

4. Ottenere la matrice dei dati centrati $\tilde{X} = HX$, e costruire il diagramma di dispersione specificando gli argomenti richiesti al punto 1.

```
# matrice dei dati centrati
Xtilde <- H %*% X
# diagramma di dispersione per dati centrati
plot(Xtilde, xlim=c(-4,13),ylim=c(-4,13),
     bg=heat.colors(n),pch=21,cex=2,asp=1)
abline(h=0)
abline(v=0)
```



5. Verificare che centrare una matrice già centrata, non produce alcun effetto:

```
sum( H%*%Xtilde - Xtilde )
```

```
## [1] 5.329071e-15
```

6. Costruire la matrice di varianze/covarianze di X: $S = \frac{1}{n}(HX)'(HX)$

```
# matrice di varianze/covarianze S
S <- (1/n) * t(H%*%X) %*% (H%*%X)
S
```

```
##      [,1] [,2]
## [1,] 3.36 3.14
## [2,] 3.14 5.01
```

7. Costruire la matrice di correlazione di X: $R = D^{-1/2} S D^{-1/2}$ dove $D^{1/2} = \text{diag}(\sqrt{s_{11}}, \ldots, \sqrt{s_{pp}})$

```
# matrice diagonale
D <- diag(diag(S)^(-.5))
D
```

```
##           [,1]      [,2]
## [1,] 0.5455447 0.0000000
## [2,] 0.0000000 0.4467671
```

```
# matrice di correlazione
R <- D %*% S %*% D
R
```

```
##           [,1]      [,2]
## [1,] 1.0000000 0.7653166
## [2,] 0.7653166 1.0000000
```

8. Costruire la matrice di varianze covarianze di X come $S = D^{1/2} R D^{1/2}$

```
# matrice diagonale
D2 <- diag(diag(S)^(.5))
D2
```

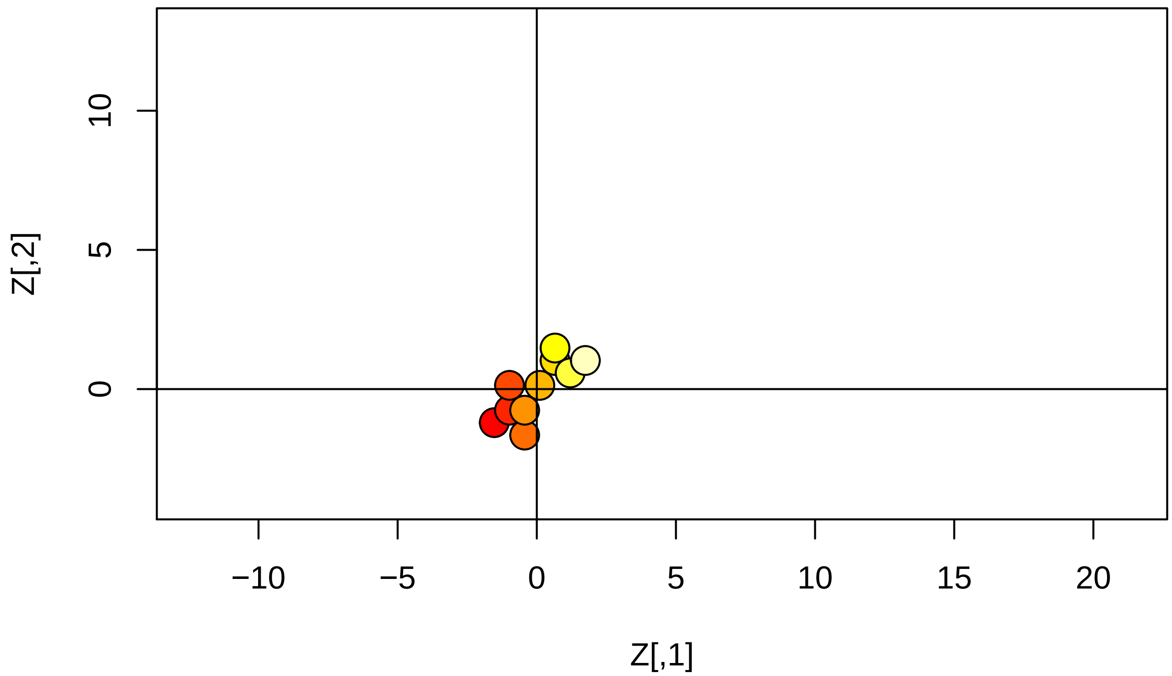
```
##         [,1]     [,2]
## [1,] 1.83303 0.000000
## [2,] 0.00000 2.238303
```

```
# matrice di varianze/covarianze S
S = D2 %*% R %*% D2
S
```

```
##      [,1] [,2]
## [1,] 3.36 3.14
## [2,] 3.14 5.01
```

9. Ottenere la matrice di dati standardizzati $Z = \tilde{X} D^{-1/2}$, e costruire il diagramma di dispersione specificando gli argomenti richiesti al punto 1.

```
# matrice dati standardizzati
Z = Xtilde %*% D
# diagramma dispersione dati standardizzati
plot(Z,xlim=c(-4,13),ylim=c(-4,13), bg=heat.colors(n),pch=21,cex=2,asp=1)
abline(h=0)
abline(v=0)
```



10. Ottenere la matrice di varianze/covarianze e di correlazione per i dati centrati $\tilde{X}$ e i dati standardizzati Z:

```
( S_Xtilde <- (1/n) * t(H%*%Xtilde) %*% (H%*%Xtilde) )
```

```
##      [,1] [,2]
## [1,] 3.36 3.14
## [2,] 3.14 5.01
```

```
( S_Z <- (1/n) * t(H%*%Z) %*% (H%*%Z) )
```

```
##           [,1]      [,2]
## [1,] 1.0000000 0.7653166
## [2,] 0.7653166 1.0000000
```

```
( R_Xtilde <-  diag(diag(S_Xtilde)^(-.5)) %*% S_Xtilde %*% diag(diag(S_Xtilde)^(-.5)) )
```

```
##           [,1]      [,2]
## [1,] 1.0000000 0.7653166
## [2,] 0.7653166 1.0000000
```

```
( S_Z <- diag(diag(S_Z)^(-.5)) %*% S_Z %*% diag(diag(S_Z)^(-.5)) )
```

```
##           [,1]      [,2]
## [1,] 1.0000000 0.7653166
## [2,] 0.7653166 1.0000000
```