

# MAE314 - Análise estatística

12) O arquivo Peru do Minitab contém medidas associadas a 39 índios peruanos que migraram das montanhas para um ambiente urbano. Considerando as variáveis: Idade(X1),Peso(X2),Altura(X3) e Pulsação(X4).

a) Obtenha o vetor de médias e a matriz de covariância amostral para essas variáveis.

## Resolução

```
library(readxl)
peru <- read_excel("/home/gui/peru.xlsx")
attach(peru)
Matriz <- matrix(c(Age,Weight,Height,Pulse),39,4)
colnames(Matriz) <- c("Age","Weight","Height","Pulse")
Medias <- apply(Matriz,2,mean)
Covariancia <- cov(Matriz)
rownames(Covariancia) <- c("Age","Weight","Height","Pulse")
colnames(Covariancia) <- c("Age","Weight","Height","Pulse")
Medias
```

```
##      Age      Weight      Height      Pulse
## 36.53846  63.15897 1578.92308   70.30769
```

## Covariancia

```
##      Age      Weight      Height      Pulse
## Age    59.097166  23.55688   22.595142  6.619433
## Weight 23.556883  50.39406   168.457287 21.023482
## Height 22.595142 168.45729  2776.757085  3.919028
## Pulse   6.619433  21.02348    3.919028 90.218623
```

b) Obtenha a matriz de correlação

## Resolução

```
Correlacao <- cor(Matriz)
rownames(Correlacao) <- c("Age","Weight","Height","Pulse")
colnames(Correlacao) <- c("Age","Weight","Height","Pulse")
Correlacao
```

```
##      Age      Weight      Height      Pulse
## Age    1.00000000  0.4316630  0.055777982  0.090654502
## Weight 0.43166298  1.0000000  0.450330307  0.311793359
## Height 0.05577798  0.4503303  1.000000000  0.007829993
## Pulse   0.09065450  0.3117934  0.007829993  1.000000000
```

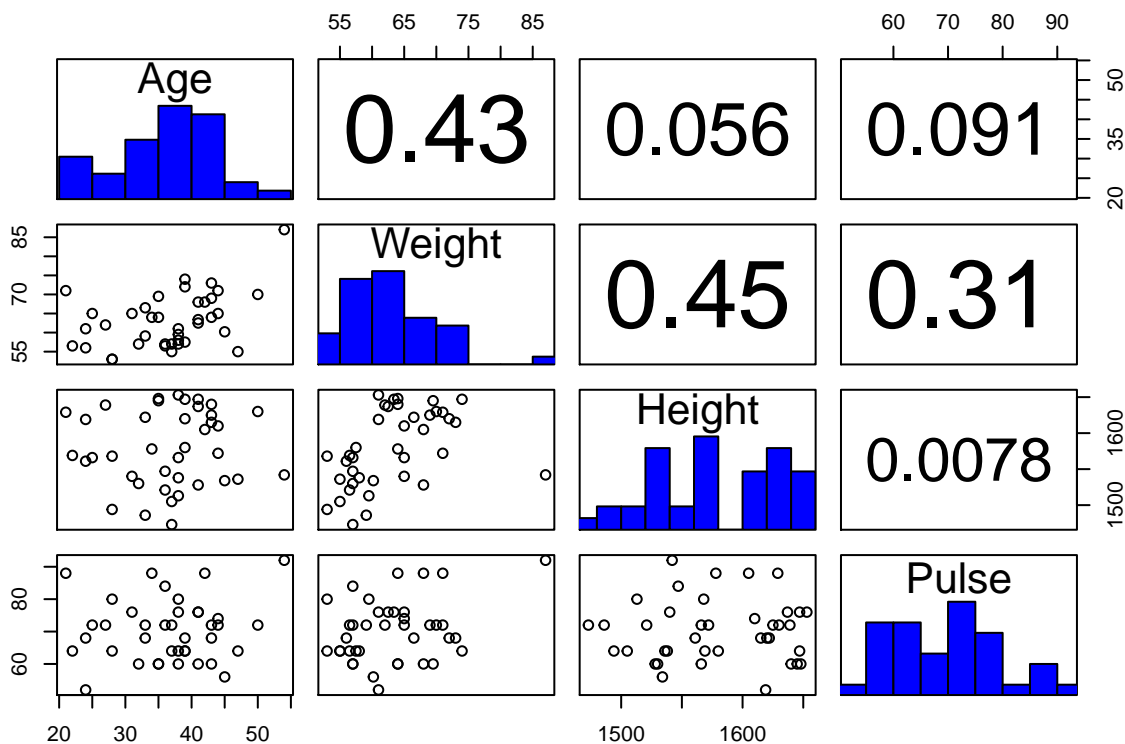
```

panel.hist <- function(x, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5) )
  h <- hist(x, plot = FALSE)
  breaks <- h$breaks; nB <- length(breaks)
  y <- h$counts; y <- y/max(y)
  rect(breaks[-nB], 0, breaks[-1], y, col = "blue", ...)
}

#função retirada do help(pairs)
panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- abs(cor(x, y))
  txt <- format(c(r, 0.123456789), digits = digits)[1]
  txt <- paste0(prefix, txt)
  if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
  text(0.5, 0.5, txt, cex = cex.cor * 1)
}

pairs(Matriz, diag.panel = panel.hist, upper.panel = panel.cor)

```



- 13) O sanduíche Big Mac, presente em todos os países onde a rede McDonald's opera, foi utilizado como um índice econômico. Em 1986, a revista britânica The Economist criou índice Big Mac, com o objetivo de

comparar o valor do Big Mac em diversos países do mundo. Os dados presentes no arquivo bigmac.xls apresentam variáveis relativas a essa análise para uma amostra de 45 países, em um determinado instante de tempo.

Para as variáveis:

BigMac - Número médio de minutos de trabalho necessários para comprar um BigMac

Bread - Número médio de minutos de trabalho necessários para comprar 1 kg de pão,

EngSal - Salário médio anual de um engenheiro elétrico, em milhares de dólares,

TeachSal-Salário médio anual do professor do ensino fundamental em milhares de dólares,

Service-Custo anual de 19 serviços primariamente relevantes para Europa e América do Norte

a) Obtenha o vetor de médias e a matriz de covariância amostral.

## Resolução

```
Mac <- read_excel("/home/gui/BigMac.xlsx")
attach(Mac)
Matriz <- matrix(c(BigMac,Bread,as.numeric(EngSal),as.numeric(TeachSal),Service),45,5)
colnames(Matriz) <- c("BigMac","Bread","EngSal","TeachSal","Service")
Media <- apply(Matriz,2,mean)
Covariancia <- cov(Matriz)
rownames(Covariancia) <- c("BigMac","Bread","EngSal","TeachSal","Service")
colnames(Covariancia) <- c("BigMac","Bread","EngSal","TeachSal","Service")
Media
```

```
##      BigMac      Bread      EngSal      TeachSal      Service
## 53.28889 25.35556 31.01556 19.92889 245.55556
```

### Covariancia

```
##           BigMac      Bread      EngSal      TeachSal      Service
## BigMac    2032.3465   837.6677 -608.0228 -446.0744 -2152.3232
## Bread      837.6677  1129.3253 -314.0488 -213.8901 -1135.6566
## EngSal    -608.0228 -314.0488   389.6445   263.9261  1116.3889
## TeachSal  -446.0744 -213.8901   263.9261   217.4894   932.6313
## Service  -2152.3232 -1135.6566  1116.3889   932.6313  8279.7980
```

b) Obtenha a matriz de correlação.(correlação linear de Pearson)

## Resolução

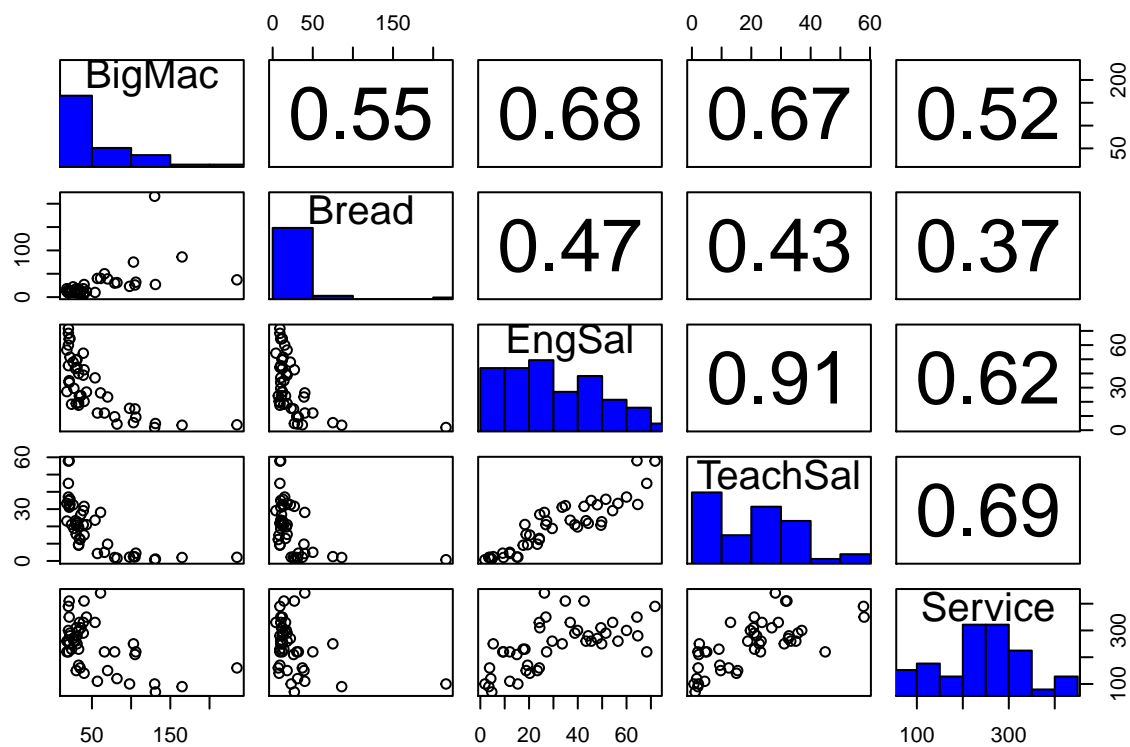
```
Correlacao <- cor(Matriz)
rownames(Correlacao) <- c("BigMac","Bread","EngSal","TeachSal","Service")
colnames(Correlacao) <- c("BigMac","Bread","EngSal","TeachSal","Service")
Correlacao
```

```
##           BigMac      Bread      EngSal      TeachSal      Service
## BigMac      1.0000000  0.5529214 -0.6832611 -0.6709489 -0.5246851
## Bread       0.5529214  1.0000000 -0.4734274 -0.4315807 -0.3713878
## EngSal      -0.6832611 -0.4734274  1.0000000  0.9066279  0.6215432
## TeachSal    -0.6709489 -0.4315807  0.9066279  1.0000000  0.6949941
## Service     -0.5246851 -0.3713878  0.6215432  0.6949941  1.0000000
```

```
panel.hist <- function(x, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5) )
  h <- hist(x, plot = FALSE)
  breaks <- h$breaks; nB <- length(breaks)
  y <- h$counts; y <- y/max(y)
  rect(breaks[-nB], 0, breaks[-1], y, col = "blue", ...)
}

#função retirada do help(pairs)
panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- abs(cor(x, y))
  txt <- format(c(r, 0.123456789), digits = digits)[1]
  txt <- paste0(prefix, txt)
  if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
  text(0.5, 0.5, txt, cex = cex.cor * 1)
}

pairs(Matriz, diag.panel = panel.hist, upper.panel = panel.cor)
```



c) Determine o par de variáveis com maior coeficiente de correlação linear

### Resolução

O par de variáveis com o maior coeficiente de correlação é (BigMac-Bread)