

Aula 1: Aula Prática 03

Cynthia Tojeiro

2025-02-06

Análise de Correlação linear

Exemplo 2: Vamos usar dados reais disponíveis no R no pacote car sob o nome Davis. As variáveis utilizadas são Altura e peso de homens e mulheres (espaços em branco) missing (dados faltantes). Estes dados são referentes aos pesos (em kg) e alturas (em cm) medidos e informados, de 200 indivíduos. O sexo de cada um também foi coletado, sendo 112 mulheres e 88 homens. O interesse é verificar se existe alguma relação linear entre as variáveis.

Inicialmente vamos carregar os dados no R, utilizando os códigos a seguir.

Pacote necessario para leitura dos dados

```
library(car)
library(xtable)
library(corrplot)
Davis
xtable(Davis)
dados<-na.exclude(Davis)
```

Impressão dos dados

```
print(dados)
genero <- Davis[,1]
amed <- Davis[,3]
pmed <- Davis[,2]
```

Vamos primeiro fazer uma análise descritiva dos dados para ver o comportamento das variáveis.

Usando a função summary verificamos o mínimo, o máximo, a mediana e os quartis para cada variável em estudo.

```
summary(dados)
```

Box-plots das variáveis

```
boxplot(amed, xlab = 'altura', col = "#FFCCCC")
text(locator(1), "12")
boxplot(pmed, xlab = 'peso', col = "#FFCCCC")
text(locator(1), "12") #Retirando a observação discrepante
generor <- genero[-12]
amedr <- amed[-12]
pmedr <- pmed[-12]
table(genero)
```

Box-plots das variáveis

```
boxplot(amedr, xlab = 'altura', col = "#FFCCCC")
boxplot(pmedr, xlab = 'peso', col = "#FFCCCC")
```

Histogramas

```
hist(amedr, probability = TRUE, xlab = 'altura', ylab = 'Densidade', main = '', col = "#FFCCCC")
lines(density(amedr), col = "#990000", lwd = 4)
hist(pmedr, probability = TRUE, xlab = 'peso', ylab = 'Densidade', main = '', col = "#FFCCCC")
lines(density(pmedr), col = "#990000", lwd = 4)
```

Gráfico de dispersão

```
plot(amed, pmed, cex=1.2, cex.lab=1.2, cex.main=1.2, xlab="altura", ylab="peso", pch=19)
```

por grupo (sexo)

```
plot(amed, pmed, cex=1.2, cex.lab=1.2, cex.main=1.2, xlab="altura", ylab="peso", pch=19, col=as.numeric(genero))
```

Plot sem a observação discrepante

```
plot(amedr, pmedr, cex=1.2, cex.lab=1.2, cex.main=1.2, xlab="altura", ylab="peso", pch=19)
```

Obtendo o coeficiente de correlação linear de Pearson

```
cor(amed[-12], pmed[-12])
#Gráfico de correlações entre as variáveis
corrplot(cor(dados[,-1]), method = c("number"), type = c("upper"))
```