

Atividade 1 - Correlação

Disciplina: Regressão 1

Schênia Taynna Medeiros Silva - 20190156798

2023-07-22

##Pacotes necessários

```
#install.packages("ggplot2") # Gráficos
```

Questão 1

Considere os dados sobre massa muscular e da idade de mulheres adultas.

#Dados

```
idade<-c(71, 64, 43, 67, 56, 73, 68, 56, 76, 65, 45, 58, 45, 53, 49, 78, 73, 68)

massamuscular<-c(82, 91, 100, 68, 87, 73, 78, 80, 65, 84, 116, 76, 97, 100, 105,
77, 73, 78)

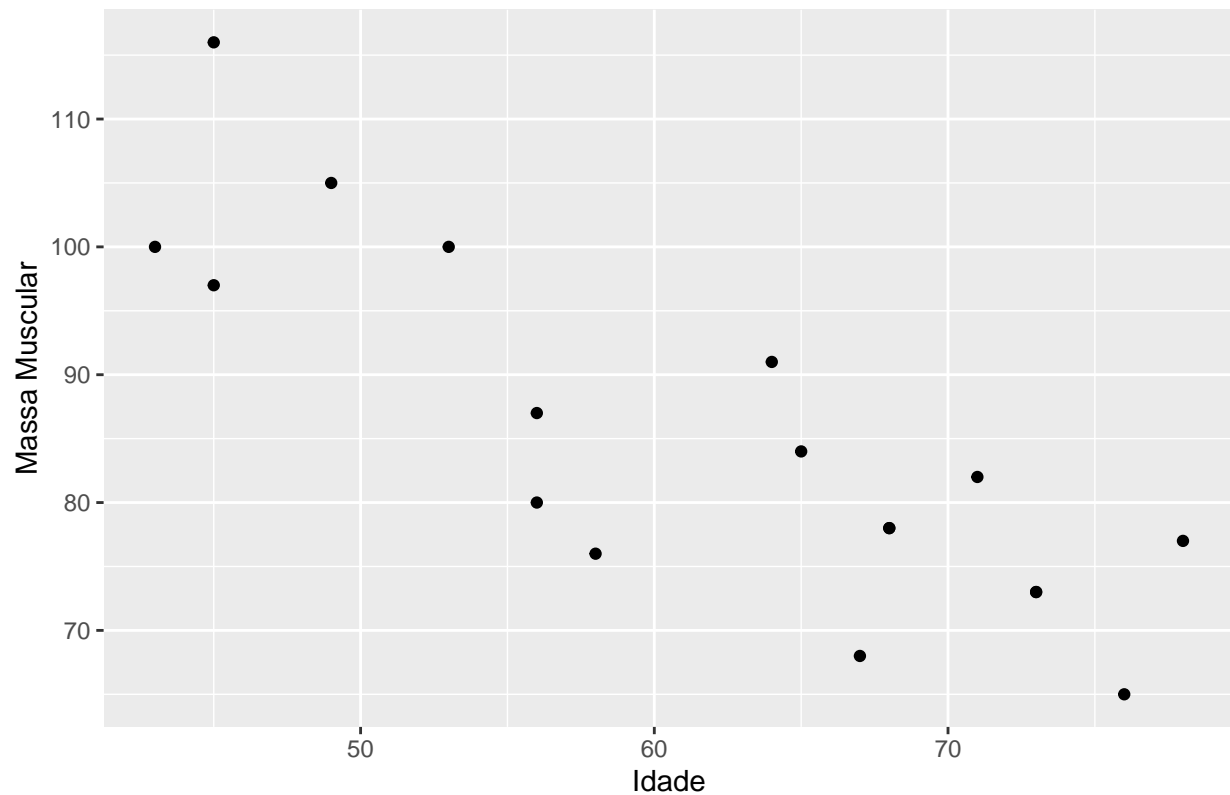
# Criando um dataframe
dados_Q1 <- data.frame(idade, massamuscular)
```

a) Faça um gráfico de dispersão entre as variáveis massa muscular e da idade de mulheres adultas.

```
library(ggplot2) # Carregamento do pacote

# Gráfico de dispersão com ggplot2
ggplot(data = dados_Q1, aes(x = idade, y = massamuscular)) +
  geom_point() +
  labs(x = "Idade", y = "Massa Muscular",
       title = "Gráfico de Dispersão: Massa Muscular x Idade")
```

Gráfico de Dispersão: Massa Muscular x Idade



b) Calcule o coeficiente de correlação de Pearson entre as variáveis consideradas.

```
# Calculando o coeficiente de correlação de Pearson
correlacao <- cor(idade, massamuscular, method = "pearson")

correlacao
```

```
## [1] -0.8366766
```

c) Realize o teste de hipóteses para verificar se esta correlação é estatisticamente significativa.

```
# Teste de hipóteses para a correlação
resultado_teste <- cor.test(idade, massamuscular)

resultado_teste

##
## Pearson's product-moment correlation
##
## data: idade and massamuscular
## t = -6.1105, df = 16, p-value = 1.504e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.9373858 -0.6068563
```

```
## sample estimates:
##      cor
## -0.8366766
```

d) Escreva uma análise estatística sobre os resultados obtidos.

O gráfico de dispersão entre a massa muscular e a idade de mulheres adultas mostra uma clara tendência negativa, assim conforme a idade aumenta, é possível observar uma diminuição na massa muscular. Os pontos formam uma linha decrescente, isso significa que mulheres mais velhas tendem a ter uma massa muscular menor em comparação com mulheres mais jovens nesta amostra.

Uma correlação negativa próxima de -1 sugere que à medida que a idade das mulheres aumenta, a massa muscular tende a diminuir. Então, com a correlação igual a -0.84 indica que essa relação é relativamente forte e que, em geral, há uma tendência linear negativa entre a idade e a massa muscular das mulheres analisadas.

O resultado do teste de hipóteses para a correlação de Pearson entre as variáveis é :

- Coeficiente de correlação de Pearson (correlação): * -0.8366766
- Estatística t: -6.1105
- Graus de liberdade (df): 16
- Valor-p: 1.504e-05 (ou seja, 0.00001504)
- Intervalo de confiança de 95%: -0.9373858 a -0.6068563

Portanto, considerando os dados amostrais com o nível de significância de 5% devemos rejeitar a hipótese nula (H_0), e assim afirmar que ao realizar o teste confirmamos que há correlação negativa entre a massa muscular e a idade é estatisticamente significativa.

Questão 2

Considere os dados sobre idade de morte e do comprimento de linha da mão.

```
# Dados
idade_morte<-c(40, 42, 42, 47, 49, 50, 54, 56, 56, 57, 57, 58, 61, 62, 62, 65,
               65, 65, 66, 66, 66, 67, 68, 68, 68, 69, 69, 70, 71, 71, 71, 72,
               73, 74, 74, 75, 75, 75, 75, 76, 77, 80, 82, 82, 82, 83, 85, 86,
               88, 88, 94)

comprimento<-c(9.00, 9.60, 9.75, 11.25, 9.45, 11.25, 9.00, 7.95, 12.00, 8.10, 10.20, 8.55,
               7.20, 7.95, 8.85, 8.25, 8.85, 9.75, 8.85, 9.15, 10.20, 9.15, 7.95, 8.85,
               9.00, 7.80, 10.05, 10.50, 9.15, 9.45, 9.45, 9.45, 8.10, 8.85, 9.60, 6.45,
               9.75, 10.20, 12.00, 6.00, 8.85, 9.00, 9.75, 10.65, 13.20, 7.95, 12.50,
               7.95, 9.15, 9.75, 9.00)

# Criando um dataframe
dados_Q2 <- data.frame(idade_morte, comprimento)
```

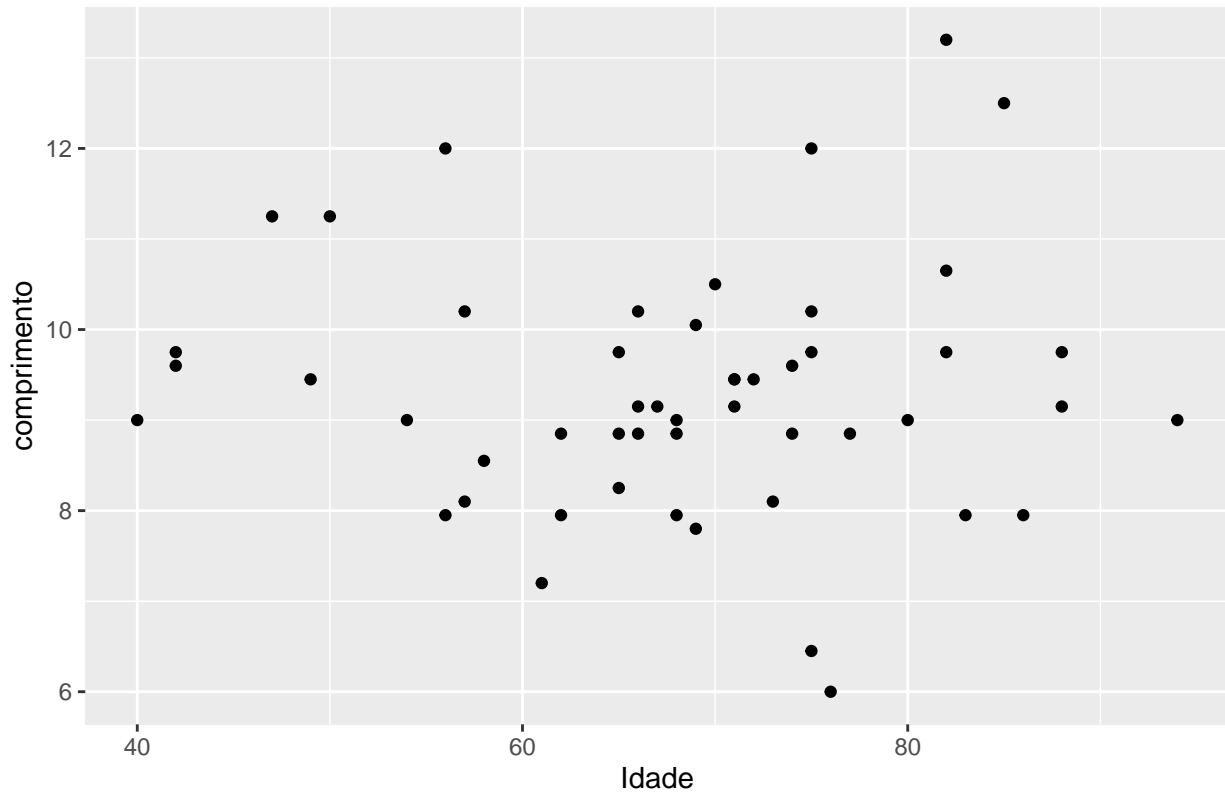
a) Faça um gráfico de dispersão entre as variáveis idade de morte e do comprimento de linha da mão.

```
library(ggplot2) # Carregamento do pacote

# Gráfico de dispersão com ggplot2
ggplot(data = dados_Q2, aes(x = idade_morte, y = comprimento)) +
```

```
geom_point() +
labs(x = "Idade", y = "comprimento",
     title = "Gráfico de Dispersão: comprimento x Idade")
```

Gráfico de Dispersão: comprimento x Idade



b) Calcule o coeficiente de correlação de Pearson entre as variáveis consideradas.

```
# Calculando o coeficiente de correlação de Pearson
correlacao_Q2 <- cor(idade_morte, comprimento, method = "pearson")

correlacao_Q2
```

```
## [1] -0.005824976
```

c) Realize o teste de hipóteses para verificar se esta correlação é estatisticamente significativa.

```
# Teste de hipóteses para a correlação
resultado_teste_Q2 <- cor.test(idade_morte, comprimento)

resultado_teste_Q2
```

```
##
## Pearson's product-moment correlation
##
```

```
## data:  idade_morte and comprimento
## t = -0.040776, df = 49, p-value = 0.9676
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2809576  0.2701924
## sample estimates:
##          cor
## -0.005824976
```

d) Escreva uma análise estatística sobre os resultados obtidos.

O gráfico de dispersão entre o comprimento de linha da mão e a idade de morte mostrou os dados de forma aparentemente aleatória e que não há uma relação clara entre as variáveis no gráfico, é possível que a correlação entre as variáveis seja fraca ou próxima de zero. E isso ocorrer porque temos pontos no gráfico que estão espalhados de forma dispersa, sem mostrar uma tendência definida.

Quando temos o coeficiente de correlação de Pearson próximo de zero, isso significa que as variáveis não estão relacionadas linearmente. Logo, ao calcularmos o a correlação obtemos -0.0058 indicando uma correlação extremamente próxima de zero entre as variáveis, assim não há uma relação linear significativa entre essas variáveis idade de morte e comprimento da linha da mão.

O resultado do teste de hipóteses para a correlação de Pearson entre as variáveis é :

- Coeficiente de correlação de Pearson (correlação): * -0.005824976
- Estatística t: -0.040776
- Graus de liberdade (df): 49
- Valor-p: 0.9676
- Intervalo de confiança de 95%: -0.2809576 a 0.2701924

Portanto, considerando os dados amostrais com o nível de significância de 5% não devemos rejeitar a hipótese nula (H_0), e assim não há evidências suficientes para afirmar que existe uma correlação significativa entre as variáveis idade de morte e o comprimento da linha da mão.

Questão 3

Considere os dados sobre salário e anos de experiência de executivos.

```
# Dados
salario<-c(19307,31769,22769,31307,27769,30923,26538,22230,28538,32307,28230,
          19076,25384,25692,42230,40923,36000,47076,31461,29923,47461,41153,
          23615,40923,45076,29076,44846)

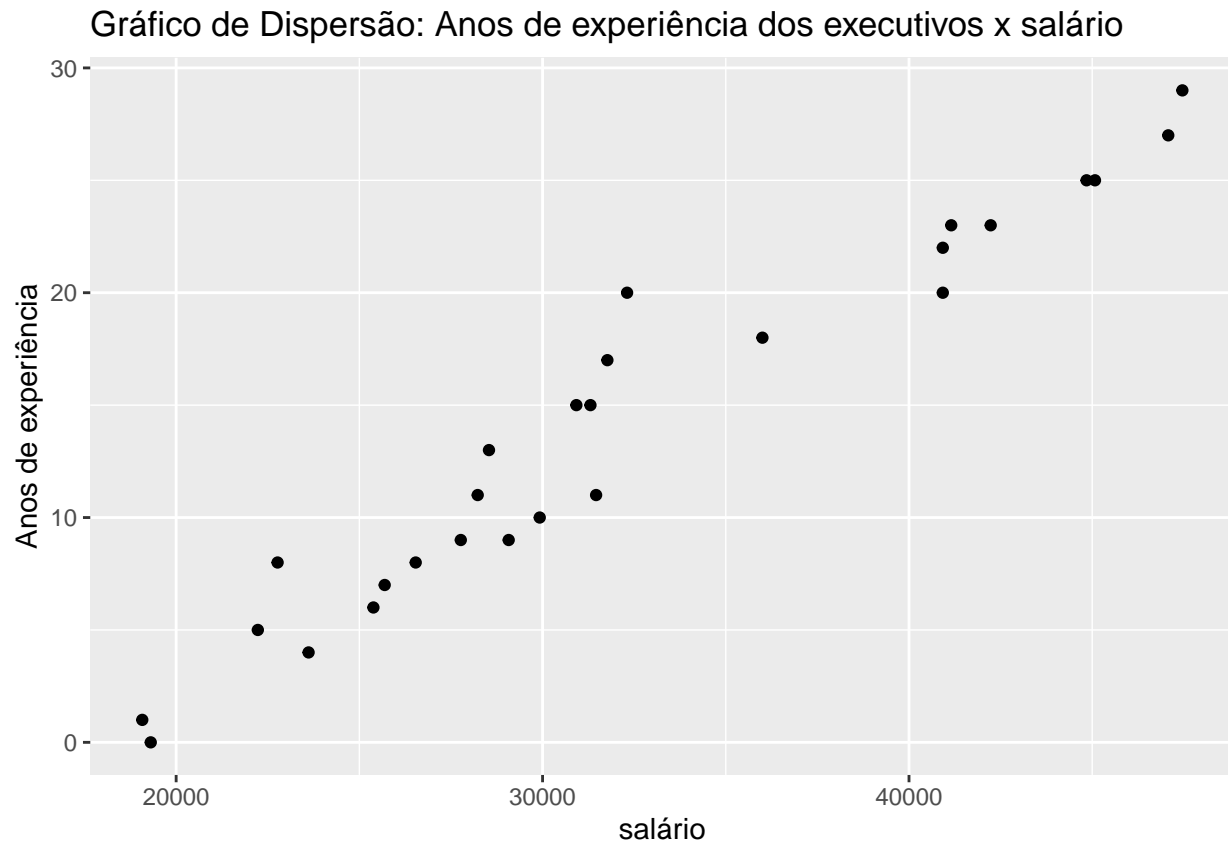
experiencia<-c(0, 17, 8, 15, 9, 15, 8, 5, 13, 20, 11, 1, 6, 7, 23, 20, 18, 27,
              11, 10, 29, 23, 4, 22, 25, 9, 25)

# Criando um dataframe
dados_Q3 <- data.frame(salario, experiencia)
```

a) Faça um gráfico de dispersão entre as variáveis salário e anos de experiência.

```
library(ggplot2) # Carregamento do pacote

# Gráfico de dispersão com ggplot2
ggplot(data = dados_Q3, aes(x = salario, y = experiencia)) +
  geom_point() +
  labs(x = "salário", y = "Anos de experiência",
       title = "Gráfico de Dispersão: Anos de experiência dos executivos x salário")
```



b) Calcule o coeficiente de correlação de Pearson entre as variáveis consideradas.

```
# Calculando o coeficiente de correlação de Pearson
correlacao <- cor(salario, experiencia, method = "pearson")

correlacao
```

```
## [1] 0.9704137
```

c) Realize o teste de hipóteses para verificar se esta correlação é estatisticamente significativa.

```
# Teste de hipóteses para a correlação
resultado_teste_Q3 <- cor.test(salario, experiencia)

resultado_teste_Q3
```

```
##
## Pearson's product-moment correlation
##
## data:  salario and experiencia
## t = 20.096, df = 25, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9353175 0.9865989
## sample estimates:
##      cor
## 0.9704137
```

d) Escreva uma análise estatística sobre os resultados obtidos.

O gráfico de dispersão entre salário e anos de experiência de executivos amostra uma clara tendência positiva, assim conforme o ano de experiência aumenta, é possível observar um aumento no salário. Os pontos formam uma linha crescente.

Uma correlação positiva próxima de 1 sugere que à medida que os anos aumentam, o salário também aumenta. Então, com a correlação igual a 0.97 indica que essa relação é relativamente forte e que, em geral, há uma tendência linear positiva entre as variáveis analisadas.

O resultado do teste de hipóteses para a correlação de Pearson entre as variáveis é :

- Coeficiente de correlação de Pearson (correlação): * 0.9704137
- Estatística t: 20.096
- Graus de liberdade (df): 25
- Valor-p: < 2.2e-16 (ou seja, extremamente próximo de zero)
- Intervalo de confiança de 95%: 0.9353175 a 0.9865989

Portanto, considerando os dados amostrais com o nível de significância de 5% devemos rejeitar a hipótese nula (H_0), e assim não há evidências suficientes para afirmar que existe uma correlação significativa entre as variáveis. Dessa forma, podemos concluir que a correlação observada entre salário e anos de experiência de executivos.