Atividade 1 e 2

Correlação, Estimando os parâmetros

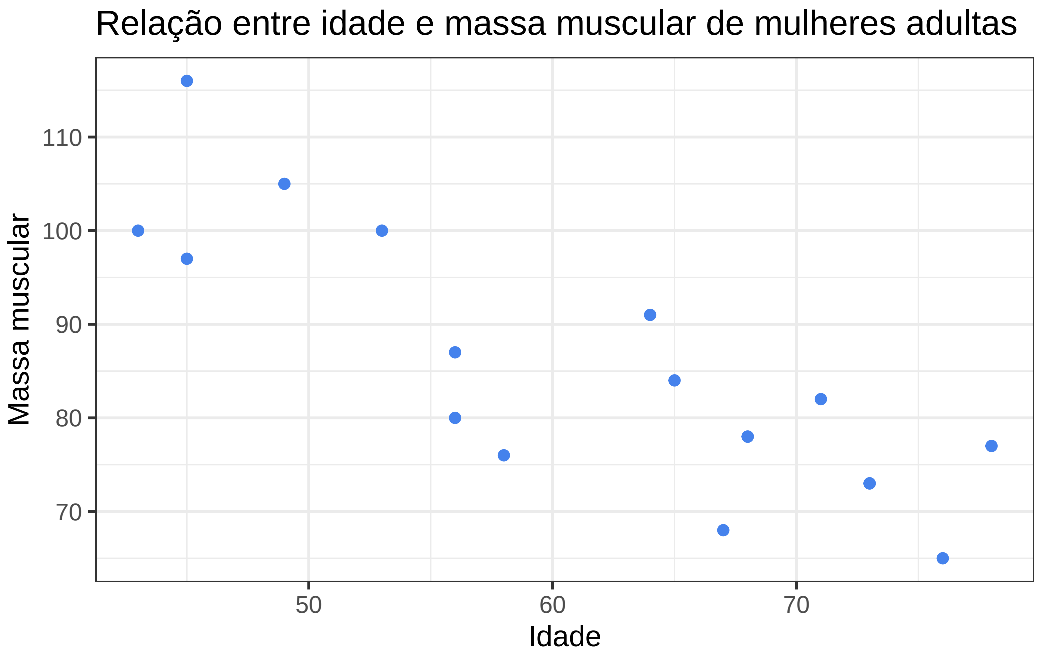
Paulo Ricardo Seganfredo Campana

23 de julho de 2023

Questão 1. Considere os dados sobre massa muscular e da idade de mulheres adultas.

```
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
)
```

a) Faça um gráfico de dispersão entre as variáveis massa muscular e da idade de mulheres adultas.
b) Calcule o coeficiente de correlação de Pearson entre as variáveis consideradas.
c) Realize o teste de hipóteses para verificar se esta correlação é estatísticamente significante.
d) Escreva uma análise estatística sobre os resultados obtidos.
e) Ajuste um modelo de regressão linear simples para explicar a massa muscular em função da idade de mulheres adultas.
f) Quais foram os valores estimados para os coeficientes de regressão? Quem é o intercepto e a inclinação da reta?
g) Expresse a reta estimada. E interprete os parâmetros.
h) Qual a estimativa do erro padrão para o modelo de regressão ajustado.



1: Teste de correlação de Pearson

Estimativa	Estatística	p-valor	IC inferior	IC superior
-0.837	-6.111	1.5e-05	-0.937	-0.607

Com uma correlação linear negativa significante de -0.837, vemos que, na faixa etária estudada, a massa muscular duminui quando a pessoa envelhece.

2: Parâmetros do modelo de regressão linear

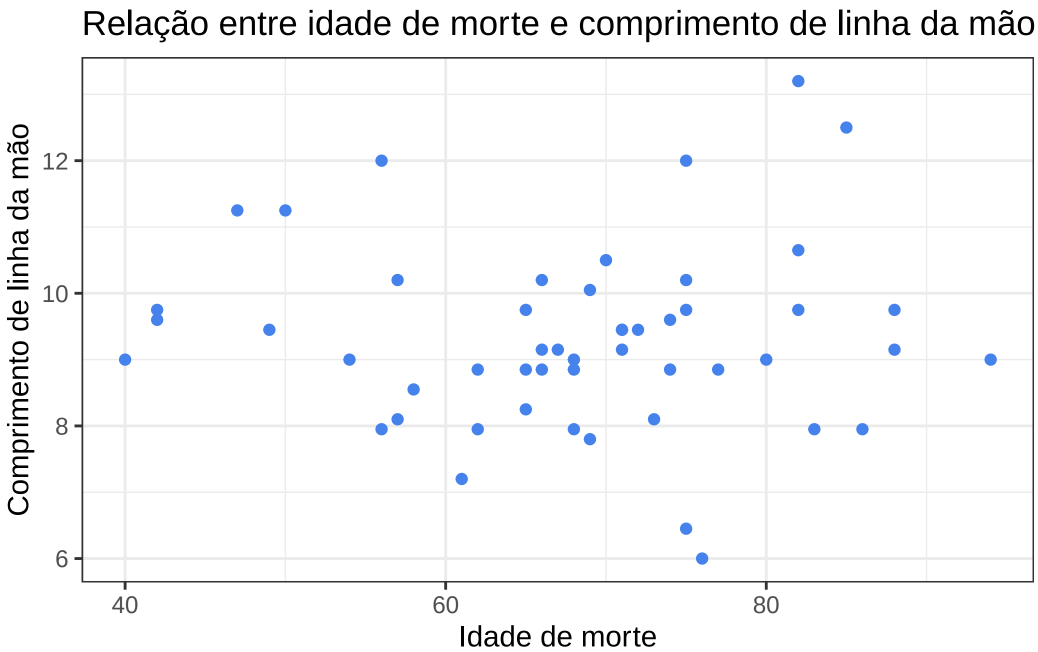
Parâmetro	Estimativa	Erro padrão	Estatística	p-valor
(Intercept)	148.197	10.505	14.108	0.0e+00
idade	-1.027	0.168	-6.111	1.5e-05

Temos intercepto de 148.197 e coeficiente angular de -1.027, que formam a reta $-1.027x + 148.197$, a inclinação da reta nos diz que, para cada ano que passa, as mulheres perdem em média 1.027 unidades de massa muscular. nesse exemplo o intercepto tem pouca interpretação, mas serve para poder ter uma reta que não passa pela origem.

Questão 2. Considere os dados sobre idade de morte e do comprimento de linha da mão.

```
idade <- 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
)
```

a) Faça um gráfico de dispersão entre as variáveis idade de morte e do comprimento de linha da mão.
b) Calcule o coeficiente de correlação de Pearson entre as variáveis consideradas.
c) Realize o teste de hipóteses para verificar se esta correlação é estatísticamente significante.
d) Escreva uma análise estatística sobre os resultados obtidos.
e) Ajuste um modelo de regressão linear simples para explicar idade de morte em função do comprimento de linha da mão.
f) Quais foram os valores estimados para os coeficientes de regressão? Quem é o intercepto e a inclinação da reta?
g) Expresse a reta estimada. E interprete os parâmetros do modelo.
h) Qual a estimativa do erro padrão para o modelo de regressão ajustado.



3: Teste de correlação de Pearson

Estimativa	Estatística	p-valor	IC inferior	IC superior
-0.006	-0.041	0.9676405	-0.281	0.27

A correlação linear estimada é de -0.006 e o teste de hipótese não mostra evidências que a mesma é diferente de 0, portanto esta correlação não é estatísticamente significante.

4: Parâmetros do modelo de regressão linear

Parâmetro	Estimativa	Erro padrão	Estatística	p-valor
(Intercept)	9.351	1.100	8.501	0.0000000
idade	-0.001	0.016	-0.041	0.9676405

O intercepto é de 9.351 e o coeficiente angular é de -0.001, que formam a reta $-0.001x + 9.351$, o parâmetro da idade de morte não pode ser interpretado pois não é significativo, enquanto o intercepto apenas diz que a média do comprimento de linha da mão é 9.351.

Questão 3. Considere os dados sobre salário e anos de experiência de executivos.

```
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
)
```

a) Faça um gráfico de dispersão entre as variáveis salário e anos de experiência.
b) Calcule o coeficiente de correlação de Pearson entre as variáveis consideradas.
c) Realize o teste de hipóteses para verificar se esta correlação é estatísticamente significante.
d) Escreva uma análise estatística sobre os resultados obtidos.
e) Ajuste um modelo de regressão linear simples para explicar o salário dos executivos em função dos anos de experiência.
f) Quais foram os valores estimados para os coeficientes de regressão? Quem é o intercepto e a inclinação da reta?
g) Expresse a reta estimada. E interprete os parâmetros do modelo.
h) Qual a estimativa do erro padrão para o modelo de regressão ajustado.

5: Teste de correlação de Pearson

Estimativa	Estatística	p-valor	IC inferior	IC superior
0.97	20.096	0	0.935	0.987

A correlação linear estimada de 0.97 é significativa, então vemos que para executivos, o salário está altamente correlacionado com os anos de experiência.

6: Parâmetros do modelo de regressão linear

Parâmetro	Estimativa	Erro padrão	Estatística	p-valor
(Intercept)	18063.303	816.101	22.134	0
experiencia	1007.593	50.140	20.096	0

Este modelo de regressão linear forma a reta $1007.59x + 18063.3$, isso significa que, para um executivo sem experiência no mercado, seu salário médio é R$ 18063,30 e para cada ano de experiência, é acrescido em média R$ 1007,59 no salário.