Exercício 1: Correlações simples

Você tem esse conjunto de dados:

x=[1 2 3 4 5]

y=[2 4 6 8 10]

a) **Calcule o coeficiente de correlação entre x e y.**

$$r = \frac{Cov(X,Y)}{desp(x) * desp(y)} = \frac{4}{1{,}41 * 2{,}83} = 1$$

b) **Interprete o resultado. O que isso diz sobre a relação de x e y?**

Que há uma correlação linear positiva perfeita entre x e y

Agora você tem esse conjunto de dados:

x=[1 2 3 4 5]

y=[10 8 6 4 2]

c) **Calcule o coeficiente de correlação entre x e y.**

$$r = \frac{Cov(X,Y)}{desp(x) * desp(y)} = \frac{-4}{1{,}41 * 2{,}83} = -1$$

d) **Interprete o resultado. O que isso diz sobre a relação de x e y?**

Ainda há uma correlação linear perfeita mas neste caso negativa, ou seja, para cada ponto de X que se incrementa corresponde a um ponto de Y que decresce

Agora com esse conjunto de dados:

x=[1 2 3 4 5]

y=[13 15 17 19 21]

e) **Calcule o coeficiente de correlação entre x e y.**

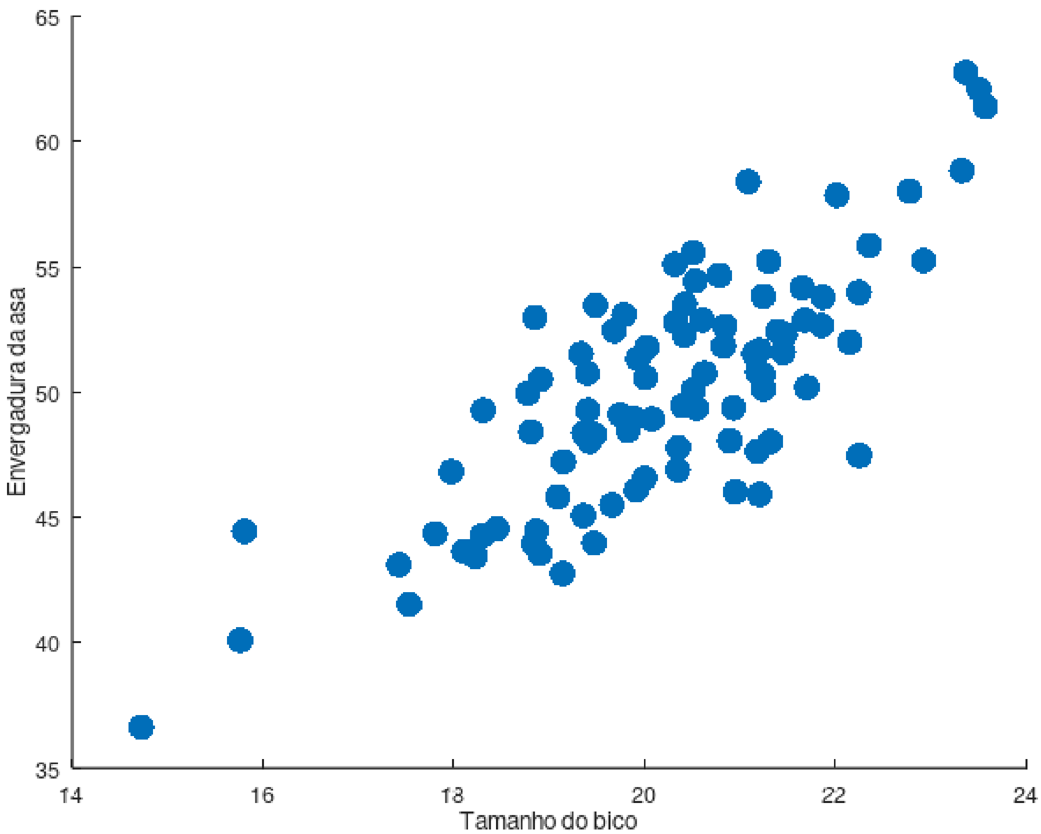
$$r = \frac{Cov(X,Y)}{desp(x) * desp(y)} = \frac{4}{1{,}41 * 2{,}83} = 1$$

f) **O offset dos dados em y causa algum efeito?**

Não no coeficiente de correlação, isso significa que a correlação mensurada existe ainda que valores sejam escalados para cima

Exercício 2: Visualização de correlação

Considere o seguinte gráfico:



a) Existe alguma correlação entre o tamanho do bico e a envergadura da asa?

Sim, existe

b) Se sim, ela é positiva ou negativa?
Correlação positiva

c) O que isso quer dizer da relação entre tamanho do bico e envergadura da asa?
Que quanto maior o bico, maior a envergadura da asa

Exercício 3: Outliers

Considere os seguintes dados:

x= [2,4,6,8,10,12,14,16,18,20,22,24,26,28,30]

y= [3,5,7,9,11,13,15,17,19,21,23,25,27,29,31]

Calcule o coeficiente de correlação.

$$r = \frac{Cov(X,Y)}{desp(x) * desp(y)} = \frac{74,67}{8,64 * 10,51} = 0,8214$$

Veja como o coeficiente de correlação é afetado quando adicionamos alguns pontos ao conjunto:

x= [2,4,6,8,10,12,14,16,18,20,22,24,26,28,30, 15, 25]

y= [3,5,7,9,11,13,15,17,19,21,23,25,27,29,31, 50, 0]

Calcule o coeficiente de correlação dos novos dados.

$$r = \frac{Cov(X,Y)}{desp(x) * desp(y)} = \frac{61,76}{8,94 * 8,69} = 0,7950$$

Exercício 4: Correlação não linear

Considere os dados:

X=[1,2,3,4,5,6,7,8,9,10]

Y=[3,9,27,81,243,729,2187,6561,19683,59049]

a) **Calcule a correlação entre as variáveis X e Y**

$$r = \frac{Cov(X,Y)}{desp(x) * desp(y)} = \frac{61,76}{2.872 * 17714.775} = 0,6963$$

b) Transforme a variável Y com o logaritmo e calcule a correlação entre X e log(Y)

$$r = \frac{Cov(X,Y)}{desp(x) * desp(y)} = \frac{3.94}{2.872 * 1.370} = 1$$

c) Você consegue explicar a mudança do coeficiente de correlação após a transformação do dado em Y?

O logaritmo do número “escalona” seu valor para uma escala menor, como os dados do Y possui uma progressão geométrica exponencial, este escalonamento passa a ser linear. Desta forma houve um ajuste no coeficiente de correlação com o conjunto de dados do primeiro conjunto de dados