



INSTITUTO FEDERAL

Rio Grande do Sul

Campus Erechim

ESTATÍSTICA

- CORRELAÇÃO LINEAR -

Profa. Claudia Turik de Oliveira



Objetivo

Estudar a relação entre duas variáveis quantitativas.

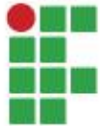
Exemplos:

- Idade e altura de crianças;
- Tempo de prática de esportes e ritmo cardíaco;
- Tempo de estudo e nota na prova;
- Taxa de desemprego e taxa de criminalidade;
- Expectativa de vida e taxa de analfabetismo.



Diagrama de Dispersão

É a representação gráfica para duas variáveis quantitativas.



Exemplo 1

- Nota da prova e tempo de estudo

X : tempo de estudo (em horas)

Y : nota da prova

Pares de observações para cada estudante:

| <i>Estudante</i> | <i>Tempo (X)</i> | <i>Nota (Y)</i> |
|------------------|------------------|-----------------|
| 1 | 3,0 | 4,5 |
| 2 | 7,0 | 6,5 |
| 3 | 2,0 | 3,7 |
| 4 | 1,5 | 4,0 |
| 5 | 12,0 | 9,3 |



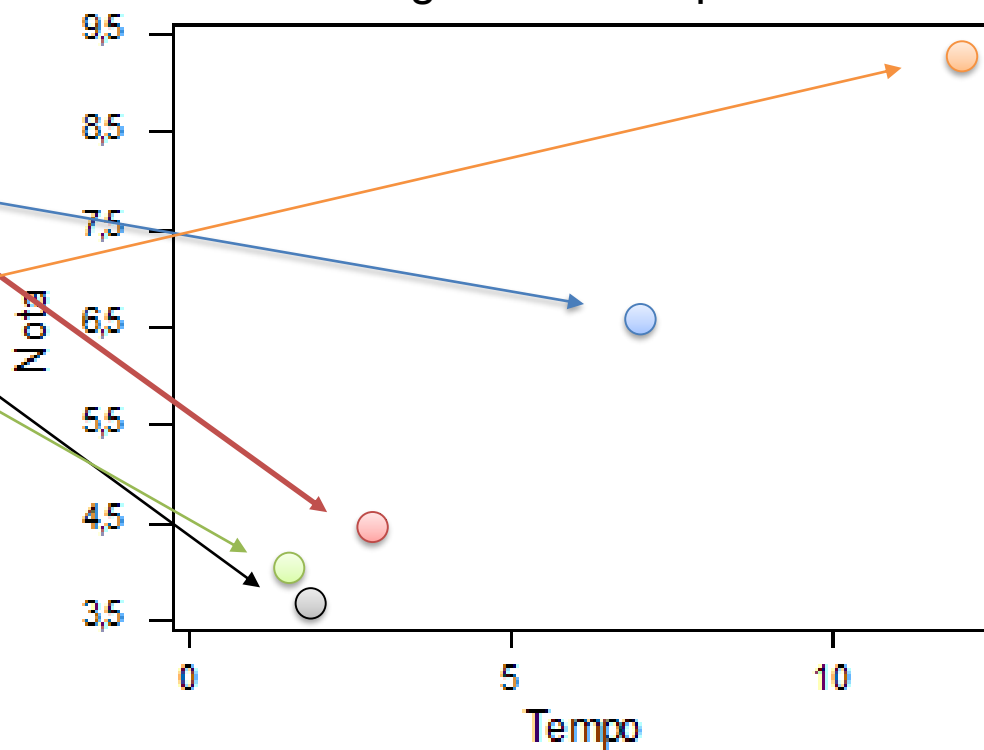
Exemplo 1

X : tempo de estudo (em horas)

Y : nota da prova

| Tempo (X) | Nota (Y) |
|-----------|----------|
| 3,0 | 4,5 |
| 7,0 | 6,5 |
| 2,0 | 3,7 |
| 1,5 | 4,0 |
| 12,0 | 9,3 |

Diagrama de Dispersão



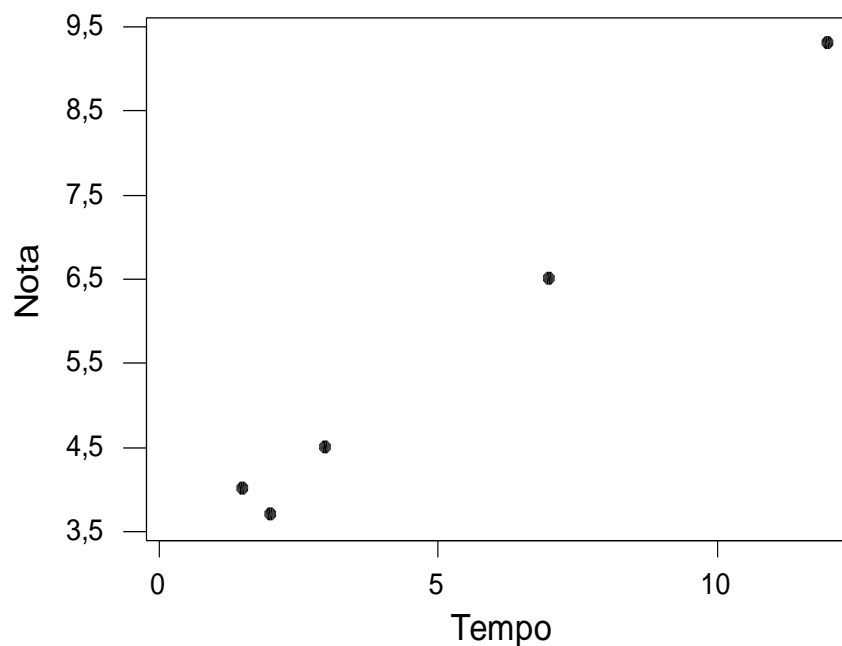
Exemplo 1

X : tempo de estudo (em horas)

Y : nota da prova

| Tempo (X) | Nota (Y) |
|-----------|----------|
| 3,0 | 4,5 |
| 7,0 | 6,5 |
| 2,0 | 3,7 |
| 1,5 | 4,0 |
| 12,0 | 9,3 |

Diagrama de Dispersão



Coeficiente de Correlação Linear

É uma medida que avalia o quanto a “nuvem de pontos” no diagrama de dispersão aproxima-se de uma reta.



Coeficiente de Correlação Linear

Propriedade: $-1 \leq r \leq 1$

Casos particulares:

$r = 1 \Rightarrow$ correlação linear positiva e perfeita

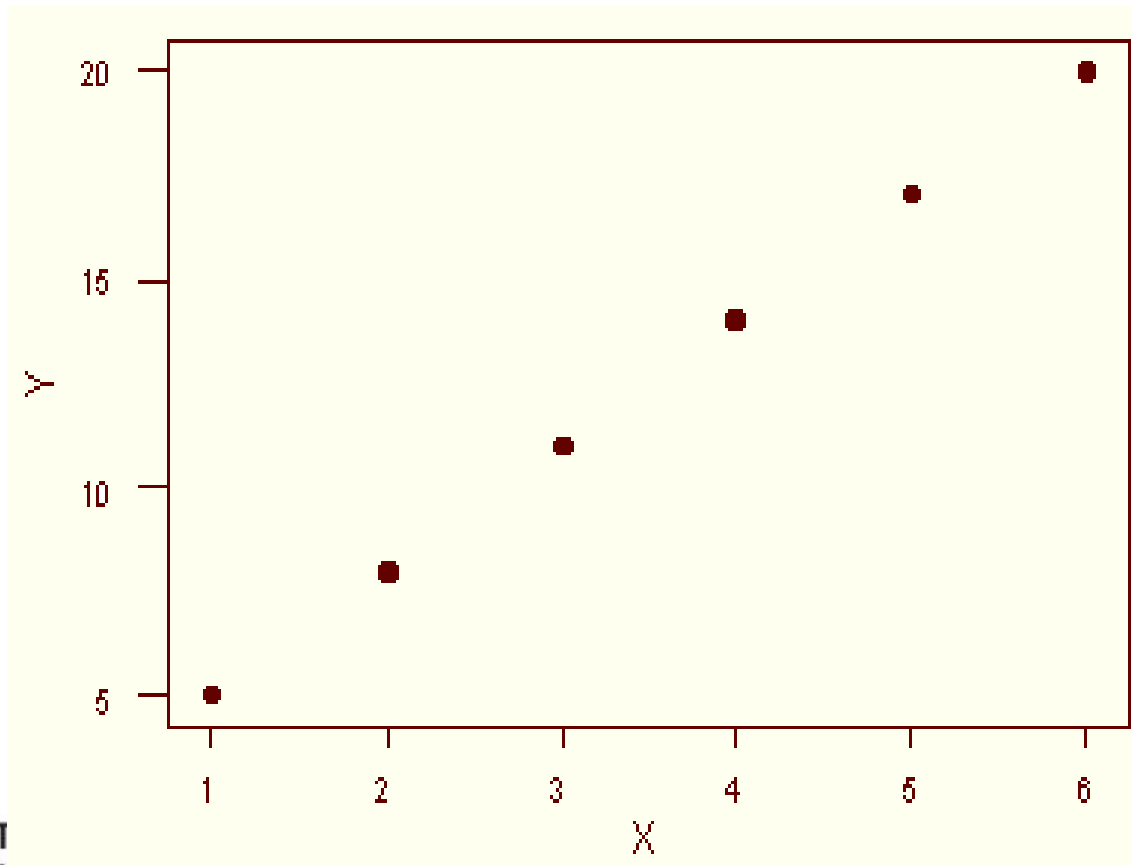
$r = -1 \Rightarrow$ correlação linear negativa e perfeita

$r = 0 \Rightarrow$ inexistência de correlação linear



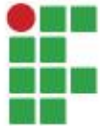
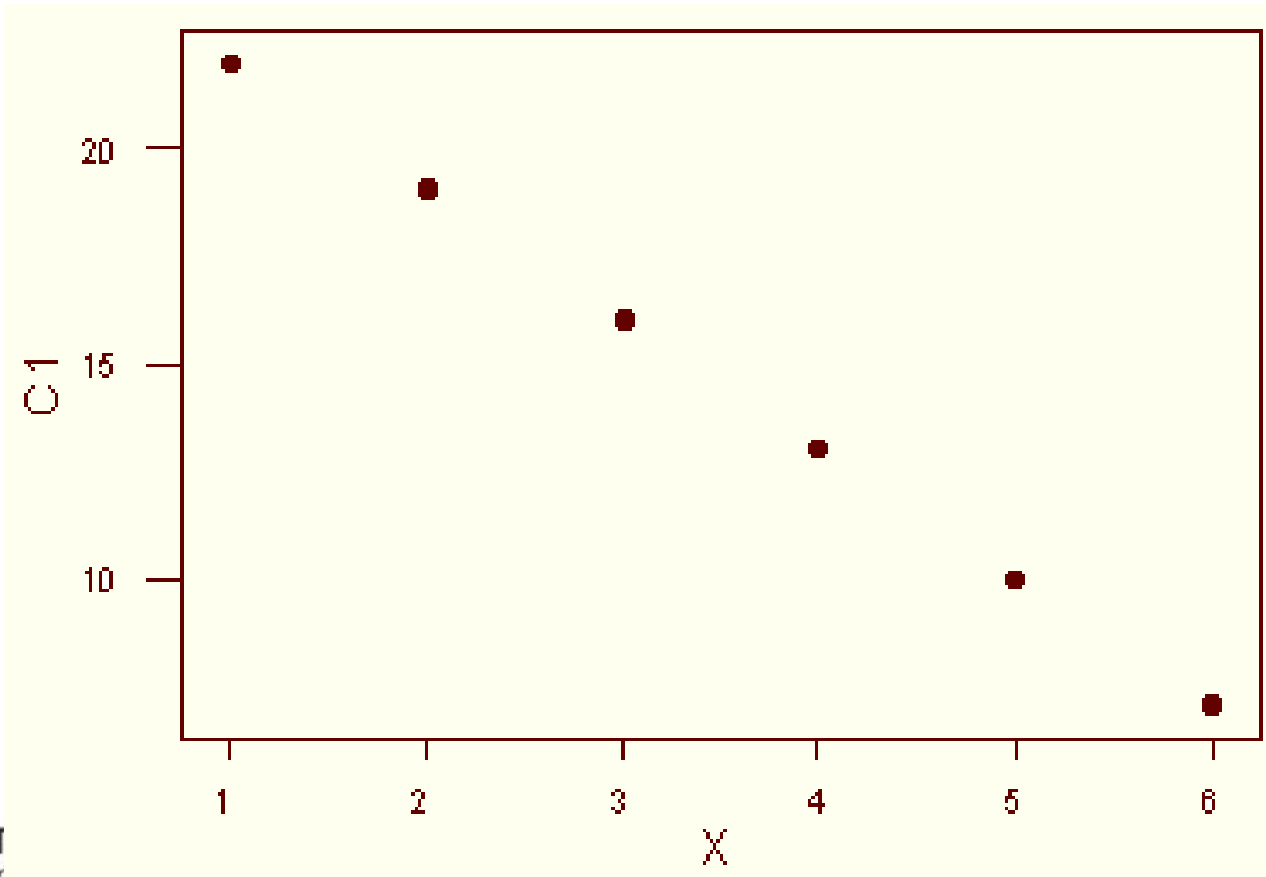
Coeficiente de Correlação Linear

$r = 1$, correlação linear positiva e perfeita



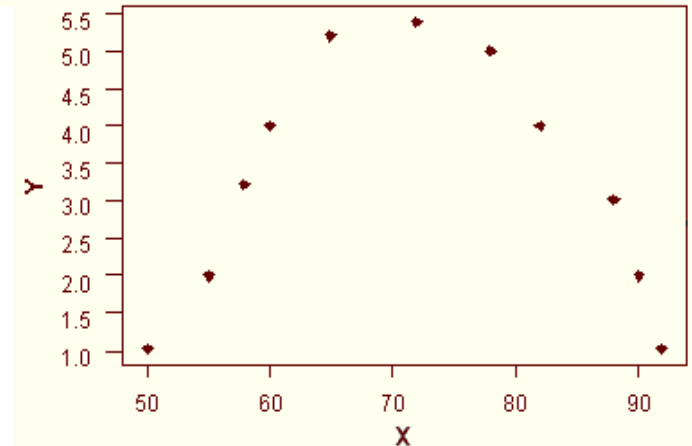
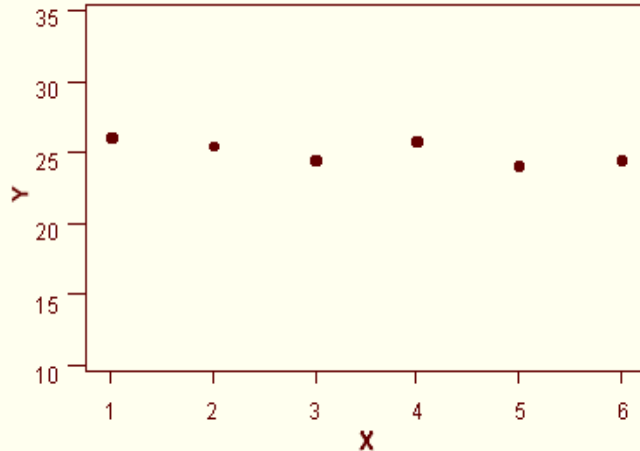
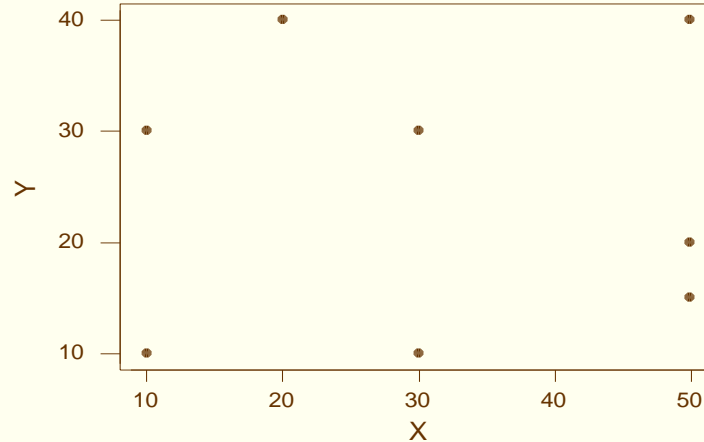
Coeficiente de Correlação Linear

$r = -1$, correlação linear negativa e perfeita



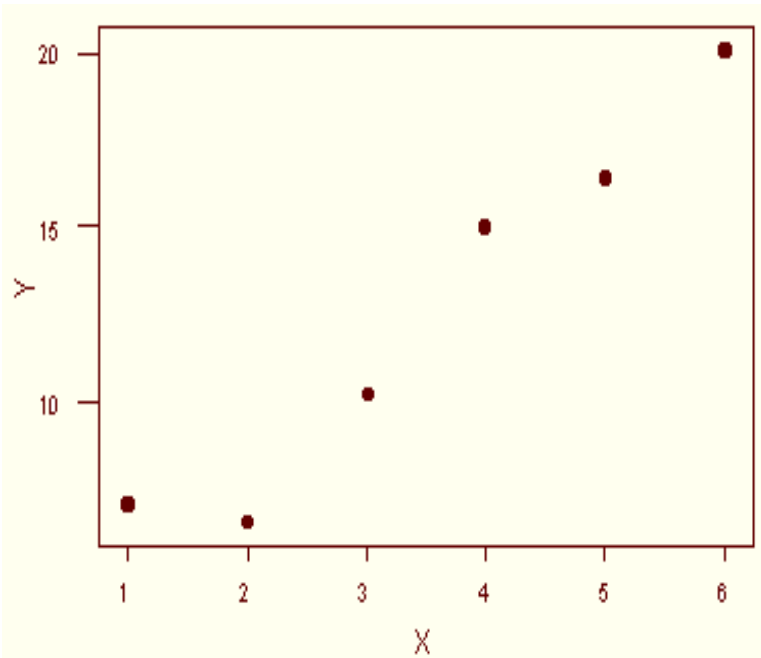
Coeficiente de Correlação Linear

$r \cong 0$, ausência de correlação linear

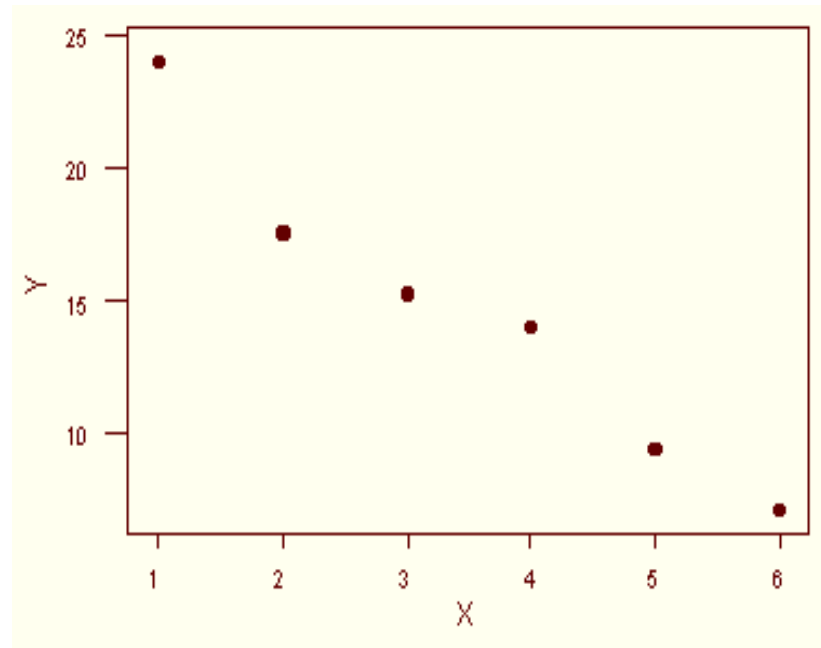


Coeficiente de Correlação Linear

$$r \cong 1$$



$$r \cong -1$$



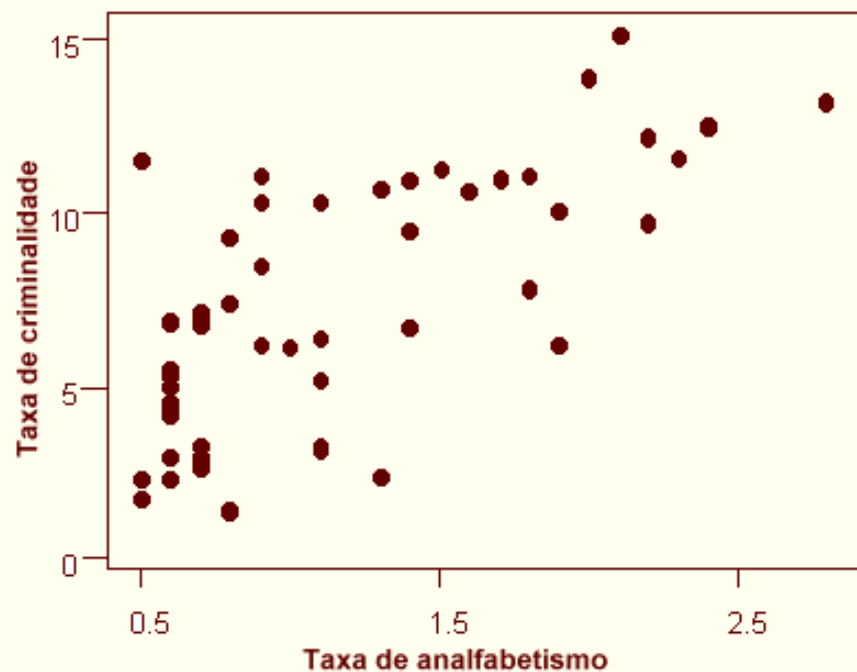
Exemplo 2

- Criminalidade e Analfabetismo

Considere duas variáveis observadas em 50 estados:

Y: taxa de criminalidade

X: taxa de analfabetismo



$$r = 0,702$$

Conforme aumenta a taxa de analfabetismo (X), a taxa de criminalidade (Y) tende a aumentar também.

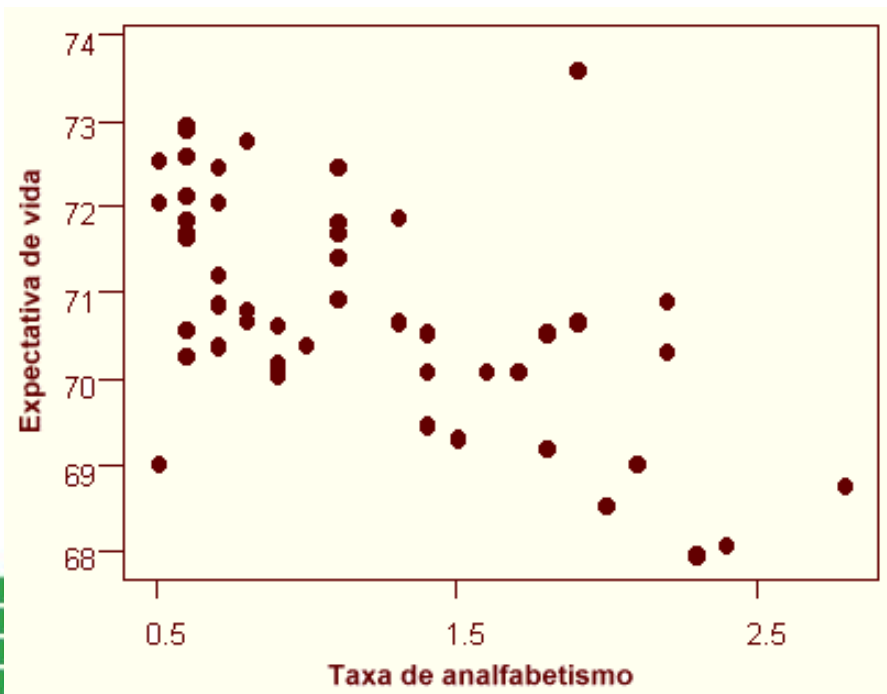
Exemplo 3

- Expectativa de vida e Analfabetismo

Considere duas variáveis observadas em 50 estados:

Y: expectativa de vida

X: taxa de analfabetismo



Coeficiente de Pearson

- O **Coeficiente de Correlação Linear de Pearson** é dado por:

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

- Sendo:

$$S_{xy} = \frac{\sum[(x - \bar{x})(y - \bar{y})]}{n - 1}$$



Covariância

$$s_x = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

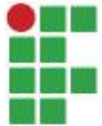


Desvio-padrão de X

$$s_y = \sqrt{\frac{\sum(y - \bar{y})^2}{n - 1}}$$



Desvio-padrão de Y



Coeficiente de Correlação Linear

Exemplo 1:

| Tempo (X) | Nota (Y) |
|-----------|----------|
| 3,0 | 4,5 |
| 7,0 | 6,5 |
| 2,0 | 3,7 |
| 1,5 | 4,0 |
| 12,0 | 9,3 |
| 25,5 | 28,0 |

1) Calcular as médias:

$$\bar{X} = 5,1$$

$$\bar{Y} = 5,6$$

2) Calcular os desvios-padrão:

$$S_x = 4,42$$

$$S_y = 2,34$$

Coeficiente de Correlação Linear

3) Calcular a covariância

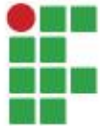
| Tempo (X) | Nota (Y) | $(X - \bar{X})$ | $(Y - \bar{Y})$ | $(X - \bar{X})(Y - \bar{Y})$ |
|-----------------|-----------------|-----------------|-----------------|------------------------------|
| 3,0 | 4,5 | -2,1 | -1,1 | 2,31 |
| 7,0 | 6,5 | 1,9 | 0,9 | 1,71 |
| 2,0 | 3,7 | -3,1 | -1,9 | 5,89 |
| 1,5 | 4,0 | -3,6 | -1,6 | 5,76 |
| 12,0 | 9,3 | 6,9 | 3,7 | 25,53 |
| 25,5 | 28,0 | 0 | 0 | 41,2 |
| $\bar{X} = 5,1$ | $\bar{Y} = 5,6$ | | | |

$$S_{xy} = \frac{2,31 + \dots + 25,53}{4} = \frac{41,2}{4} = 10,3$$

Coeficiente de Correlação Linear

4) Calcular o Coeficiente de Pearson:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{10,3}{4,42 * 2,34} = 0,9959$$



Correlação e Causalidade

- Quando duas variáveis são correlacionadas, é possível prever valores de uma delas com base no conhecimento da outra.
- Isso leva frequentemente à conclusão errônea de que uma variável é *causa* da outra.
- Entretanto, o fato de haver um relacionamento matemático entre duas variáveis nada nos diz quanto a causa e efeito.



Correlação e Causalidade

- Há três explicações possíveis para a obtenção de uma correlação:
 1. Existe uma relação de causa e efeito;
 2. Ambas as variáveis se acham relacionadas com uma terceira;
 3. A correlação é devida ao acaso.

