

Universidade Federal do Rio Grande do Norte
Instituto Metr pole Digital
IMD0601 - Bioestat stica

Covari ncia e Correla  o

Prof. Dr. Tetsu Sakamoto
Instituto Metr pole Digital - UFRN
Sala A224, ramal 182
Email: tetsu@imd.ufrn.br



Baixe a aula (e os arquivos)

- Para aqueles que não clonaram o repositório:

```
> git clone https://github.com/tetsufmbio/IMD0601.git
```

- Para aqueles que já tem o repositório local:

```
> cd /path/to/IMD0601
```

```
> git pull
```

Objetivos da aula

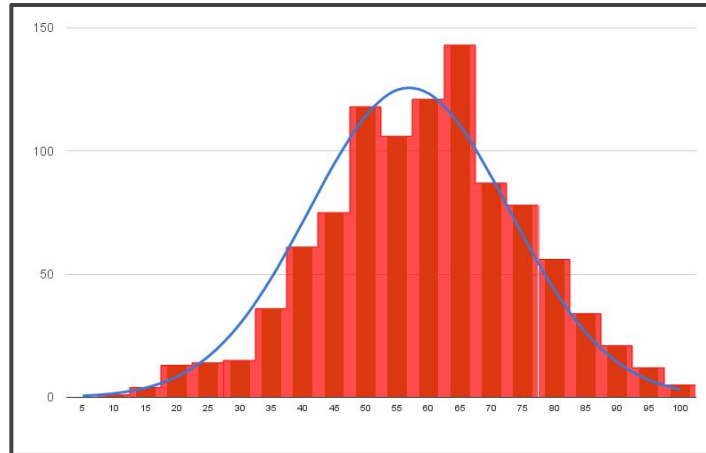
Obter noções básicas de:

- Covariância;
- Correlação.

Até o momento...

Aprendemos a descrever estatisticamente dados que envolve uma variável (**análise univariada**);

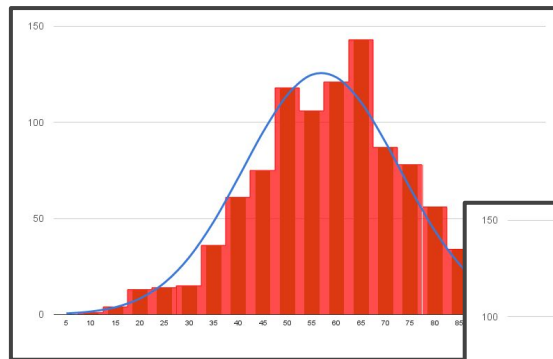
paciente	Força no quadríceps (N)
1	196
2	216
3	216
4	392
...	...



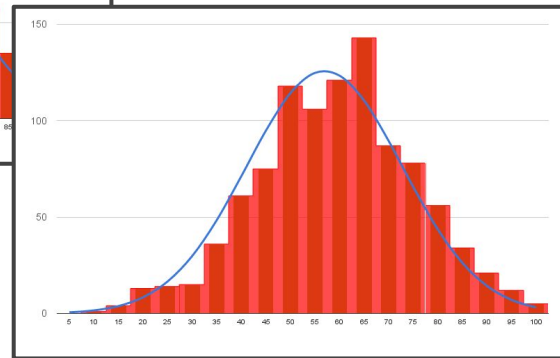
E se...

... além de termos os dados da força, temos dados da altura de cada paciente que fez o teste. Que tipo de pergunta podemos fazer?

paciente	Força no quadríceps (N)	Altura
1	196	155
2	216	159
3	216	159
4	392	160
...	...	



Força

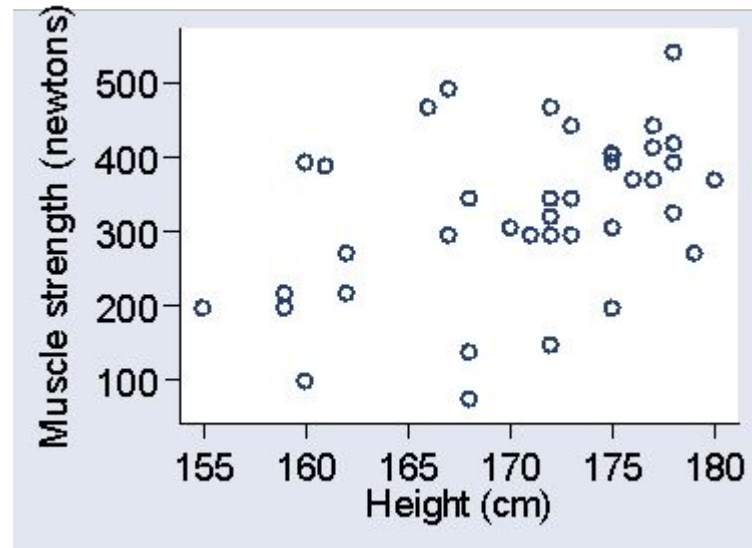


Altura

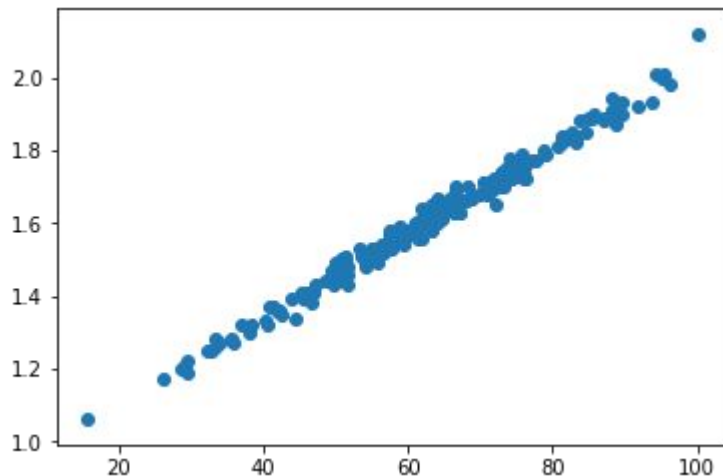
Será que estas
variáveis estão
associadas?

Será que estas variáveis se relacionam?

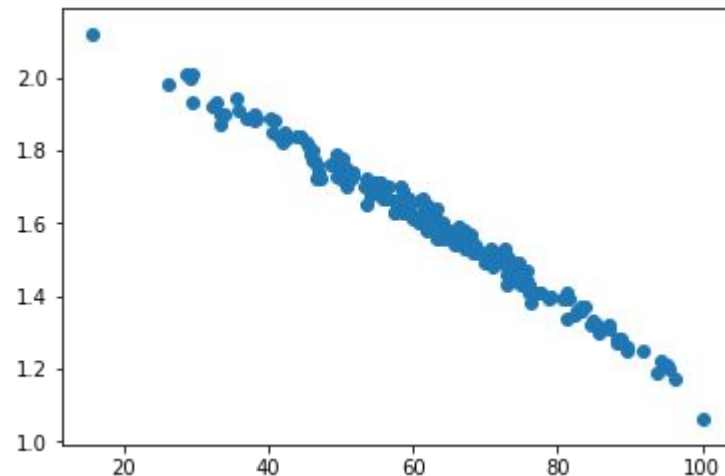
paciente	Força no quadríceps (N)	Altura
1	196	155
2	216	159
3	216	159
4	392	160
...	...	



Como se comportariam duas variáveis que estão associadas?

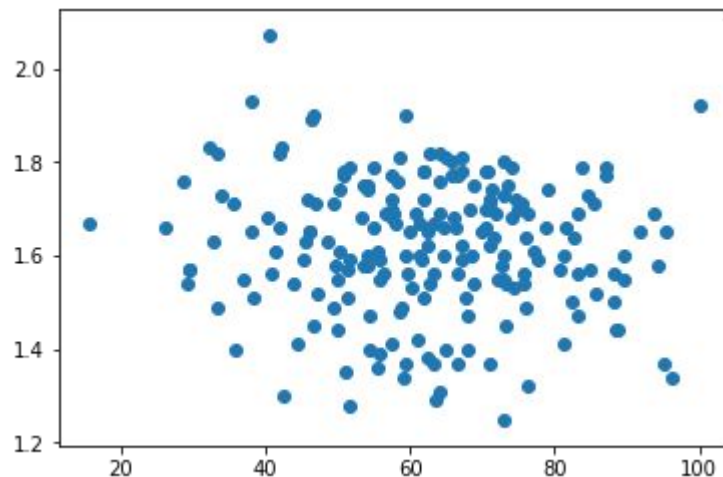


Associação positiva



Associação negativa

Como se comportariam duas variáveis que não estão associadas?

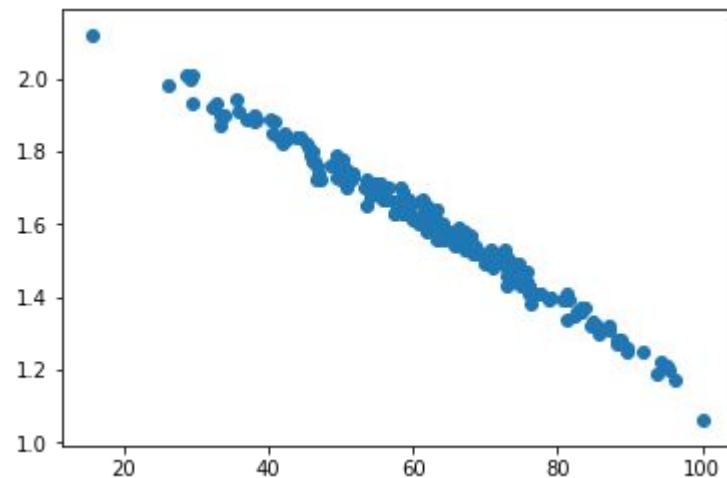
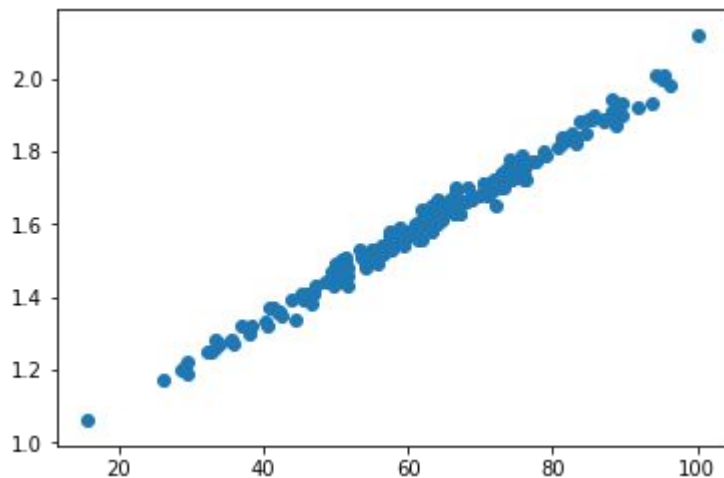


Como medir a relação entre as variáveis?

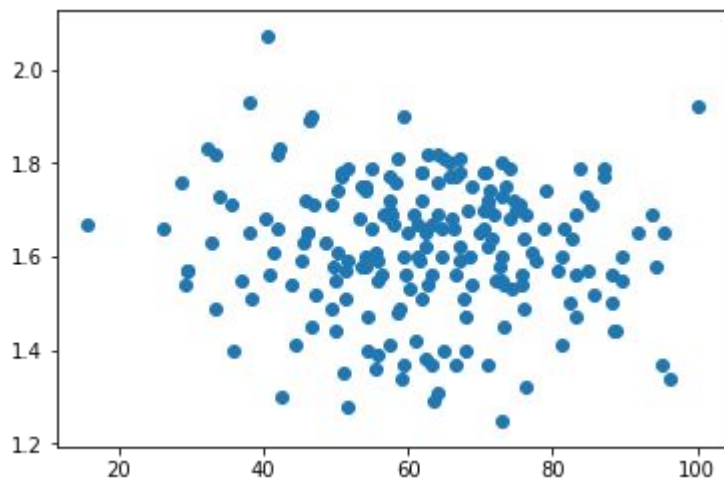
Covariância: média do produto do desvio da variável X e Y.

$$Cov(X, Y) = \frac{\sum (X_i - \bar{X}) * (Y_i - \bar{Y})}{n}$$

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X}) * (Y_i - \bar{Y})}{n}$$



$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X}) * (Y_i - \bar{Y})}{n}$$



Propriedades da Covariância

- $-\infty < \text{Cov}(X, Y) < +\infty$
- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, C) = 0$ se C é uma constante;

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X}) * (Y_i - \bar{Y})}{n}$$

Sobre o valor numérico da covariância

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X}) * (Y_i - \bar{Y})}{n}$$

A unidade da covariância seria: (unidade de X) * (unidade de Y);

Comparar covariância de diferentes pares de variável é difícil, pois se alterarmos a escala, a covariância muda também.

Como remover a escala dos dados?

Covariância em dados padronizados

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X}) * (Y_i - \bar{Y})}{n}$$

Correlação (r ou ρ) (Correlação de Pearson)

$$\textit{Correlation} = \frac{\textit{Cov}(x, y)}{\sigma x * \sigma y}$$

Propriedades da correlação

- ρ é a covariância dos dados padronizados de X e Y;
- Adimensional (lida com proporção);
- $-1 < \rho < 1$;

$$\textit{Correlation} = \frac{\textit{Cov}(x, y)}{\sigma x * \sigma y}$$