

# Coeficiente de Correlação Linear de Pearson

# Introdução

Tem-se uma variável estatística **bidimensional** quando, relativamente a cada elemento da população, se observa e estuda duas características distintas.

Para as variáveis estatísticas  $X$  e  $Y$ , a variável estatística bidimensional é representada por  $(X, Y)$ .

**Diagrama de dispersão** – “nuvem “ de pontos – é o conjunto dos pontos do tipo  $(x, y)$  representados num referencial, onde  $x$  e  $y$  são os valores observados das variáveis  $X$  e  $Y$ , respectivamente.

Quando tomamos as variáveis duas a duas podemos verificar o que sucede a uma variável,  $X$ , quando outra variável,  $Y$ , varia.

Existe correlação linear quando é possível ajustar à “nuvem” de pontos uma recta.

# Coeficiente de Correlação de Pearson:

A intensidade da associação linear existente entre as variáveis pode ser quantificada através do chamado coeficiente de correlação linear de Pearson:

$$r = \frac{C_{X,Y}}{S_X S_Y}, \quad r \in [-1, 1]$$

onde:

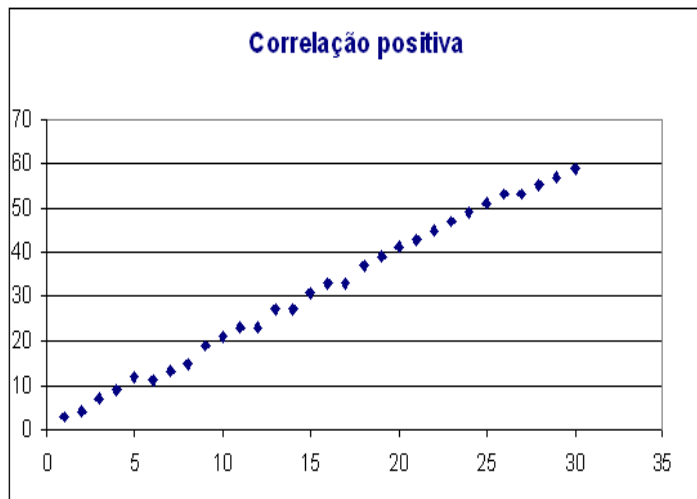
- $C_{XY}$ - Covariância ou variância conjunta das variáveis X e Y;
- $S_X$ - desvio padrão da variável X;
- $S_Y$ - desvio padrão da variável Y.

- **Caso de Dados não agrupados**

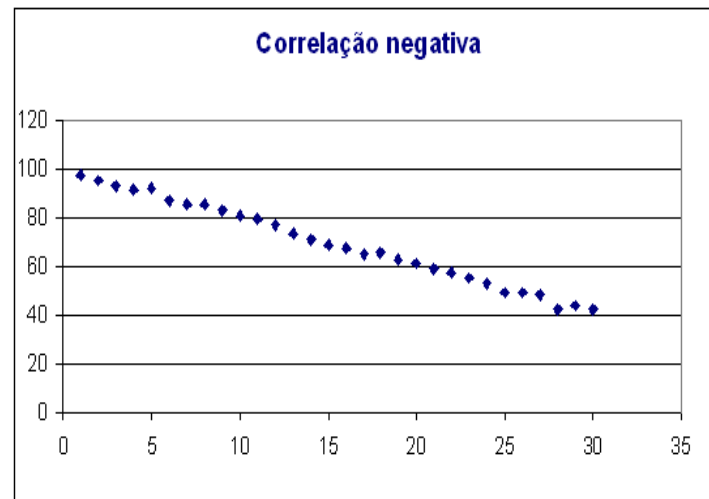
$$C_{X,Y} = \frac{\sum_{i=1}^n x_i y_i}{n} - \bar{X}\bar{Y}$$

- **Caso de Dados agrupados**

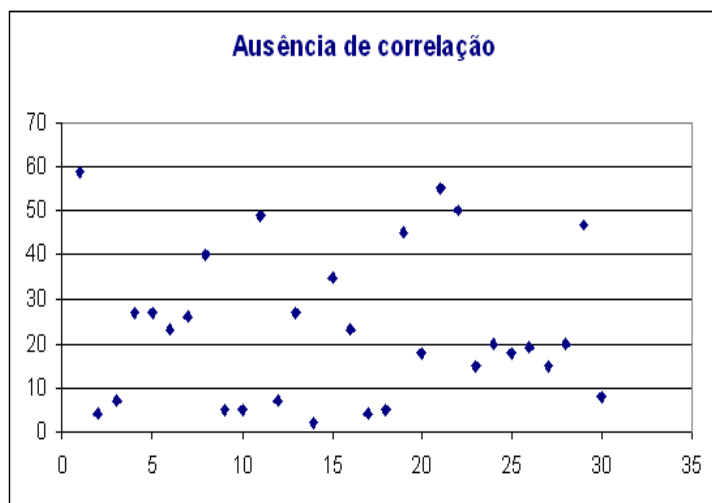
$$C_{X,Y} = \frac{\sum_{i=1}^m x_i y_i F_i}{n} - \bar{X}\bar{Y}$$



a)



b)



c)

a) Variáveis **positivamente correlacionadas**. No limite, isto é, se a correlação for "perfeita" - como é o caso se considerarmos a correlação da variável  $x$  consigo própria - o coeficiente de correlação será igual a 1.

b) as variáveis estão **negativamente correlacionadas**. No limite, isto é, se a correlação for "perfeita" o coeficiente de correlação será igual a -1.

c) as variáveis **não estão correlacionadas**. No limite, isto é, em caso de "absoluta independência" o coeficiente de correlação será igual a 0.

Coeficiente de correlação	Correlação
$r = 1$	Perfeita positiva
$0,8 \leq r < 1$	Forte positiva
$0,5 \leq r < 0,8$	Moderada positiva
$0,1 \leq r < 0,5$	Fraca positiva
$0 < r < 0,1$	Ínfima positiva
0	Nula
$-0,1 < r < 0$	Ínfima negativa
$-0,5 < r \leq -0,1$	Fraca negativa
$-0,8 < r \leq -0,5$	Moderada negativa
$-1 < r \leq -0,8$	Forte negativa
$r = -1$	Perfeita negativa

Observação 1: Não se verificar correlação linear, **não significa que não se verifique outro tipo de correlação,** por exemplo, exponencial.

Observação 2: Qualquer que seja a correlação verificada, **correlação não significa causalidade.**



**Exemplo:** Calcular o valor do coeficiente de correlação linear de Pearson entre o peso e a altura:

<i>Altura (X)</i>	<i>Peso (Y)</i>
1,75	76
1,69	71
1,85	90
1,8	81
1,7	70
1,74	73
1,87	76
1,68	65
1,76	70
1,95	92
1,96	90
1,62	55

# Diagrama de Dispersão

