

Análise de Informação Económica e Empresarial

Aula 20: Associação e relação entre variáveis



Aula 20: Associação e relação entre variáveis

Conceitos Fundamentais:

1. Relação entre variáveis
2. Diagrama de dispersão
3. Nuvem de pontos
4. Covariância
5. Coeficiente de correlação linear

Tópicos:

1. A análise da relação entre variáveis: o caso da relação entre duas variáveis
2. O diagrama de dispersão
3. A covariância
4. O coeficiente de correlação linear

Exemplos a utilizar para consolidação dos conceitos: Exercício de relação entre variáveis

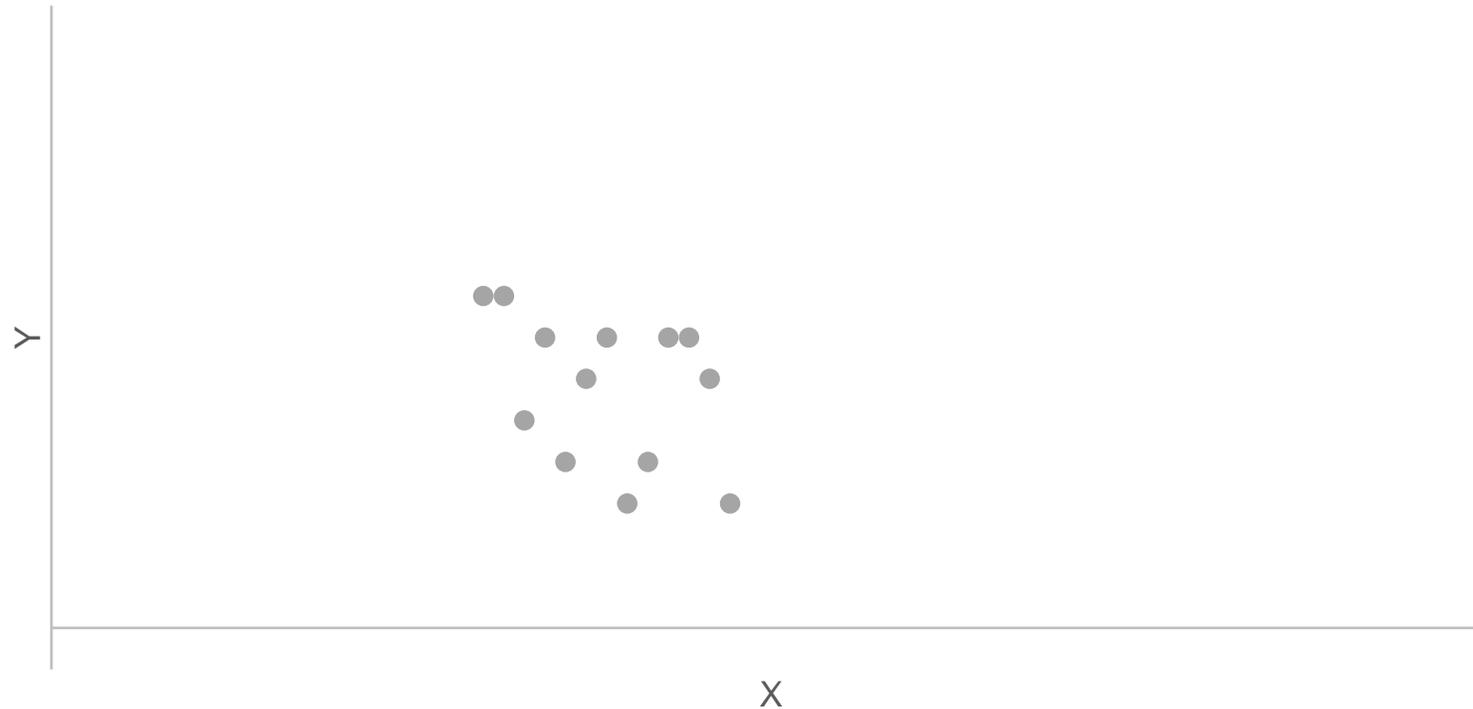
Exercício de aplicação:

Bibliografia: Reis, Elizabeth (2005) Estatística Descritiva, Lisboa: Edições Sílabo, 6ª edição: Cap 8. Associação entre duas variáveis: Correlação Linear e Regressão simples, Secção 8.1. e 8.3. pp 181-188

- ▶ Até aqui cada variável foi analisada de forma isolada
- ▶ No entanto, em muitas situações as variáveis de interesse mantêm uma relação com outras variáveis
- ▶ Porque é importante considerar a relação existente entre as variáveis?
 - Conhecer o comportamento de uma das variáveis pode ajudar a prever o comportamento de outra;
 - Se a relação for perfeita, conhecer uma variável garante o conhecimento da outra;
 - Na maior parte dos casos relação não é perfeita logo conhecer uma das variáveis permite ter indicações sobre o comportamento da outra;

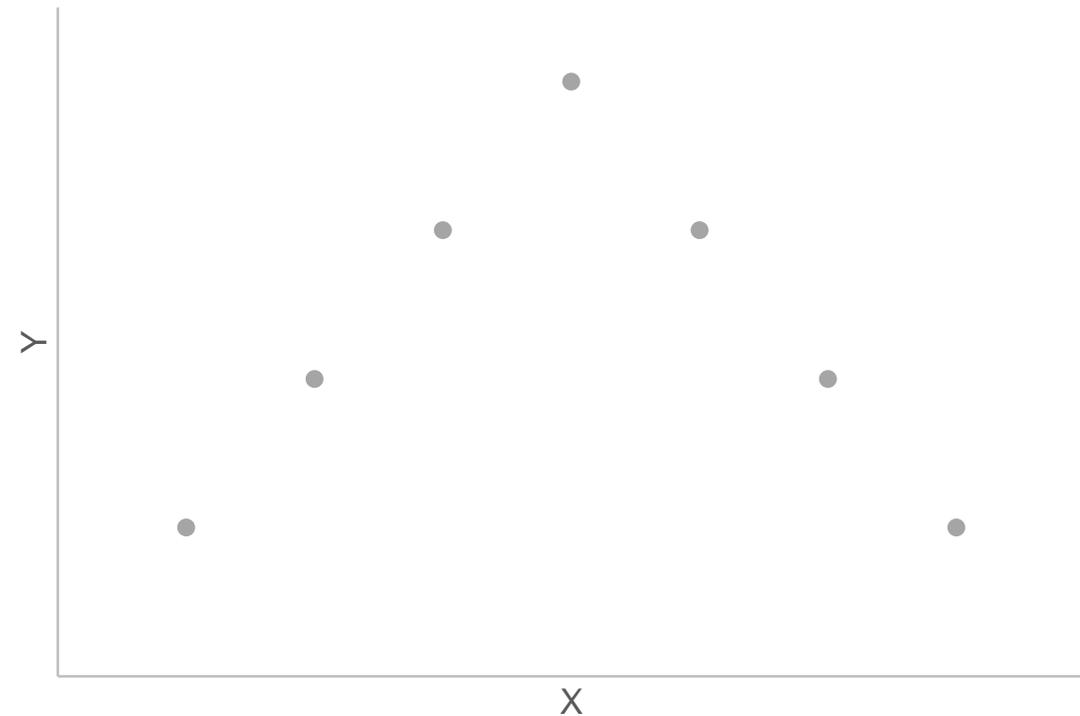
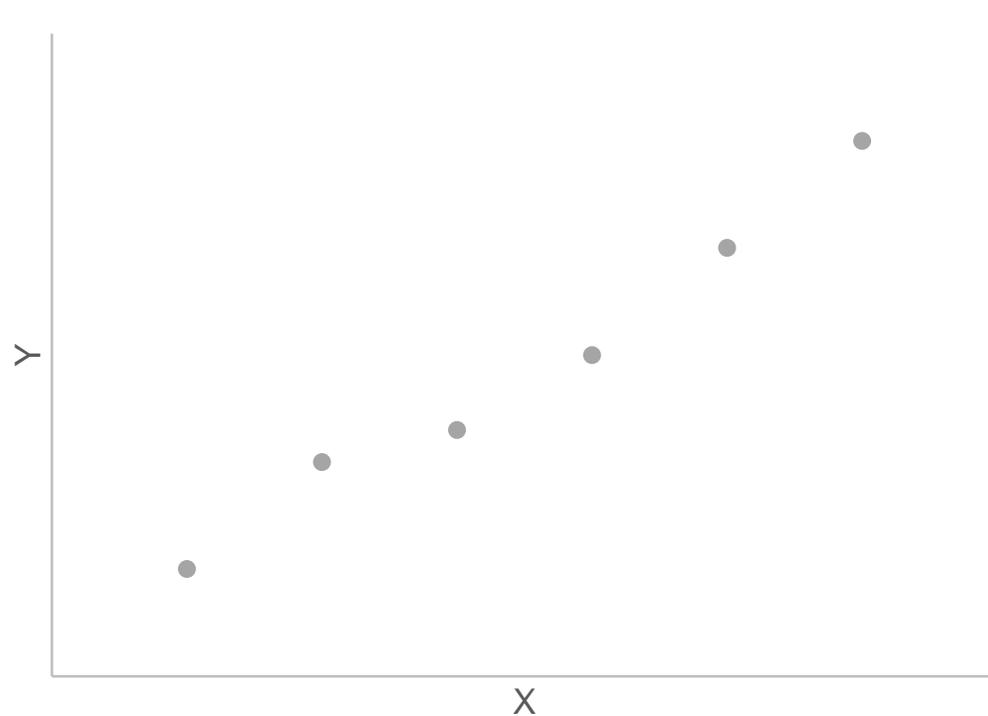
◆ Diagrama de dispersão e a nuvem de pontos

Representação gráfica dos pares (x,y)

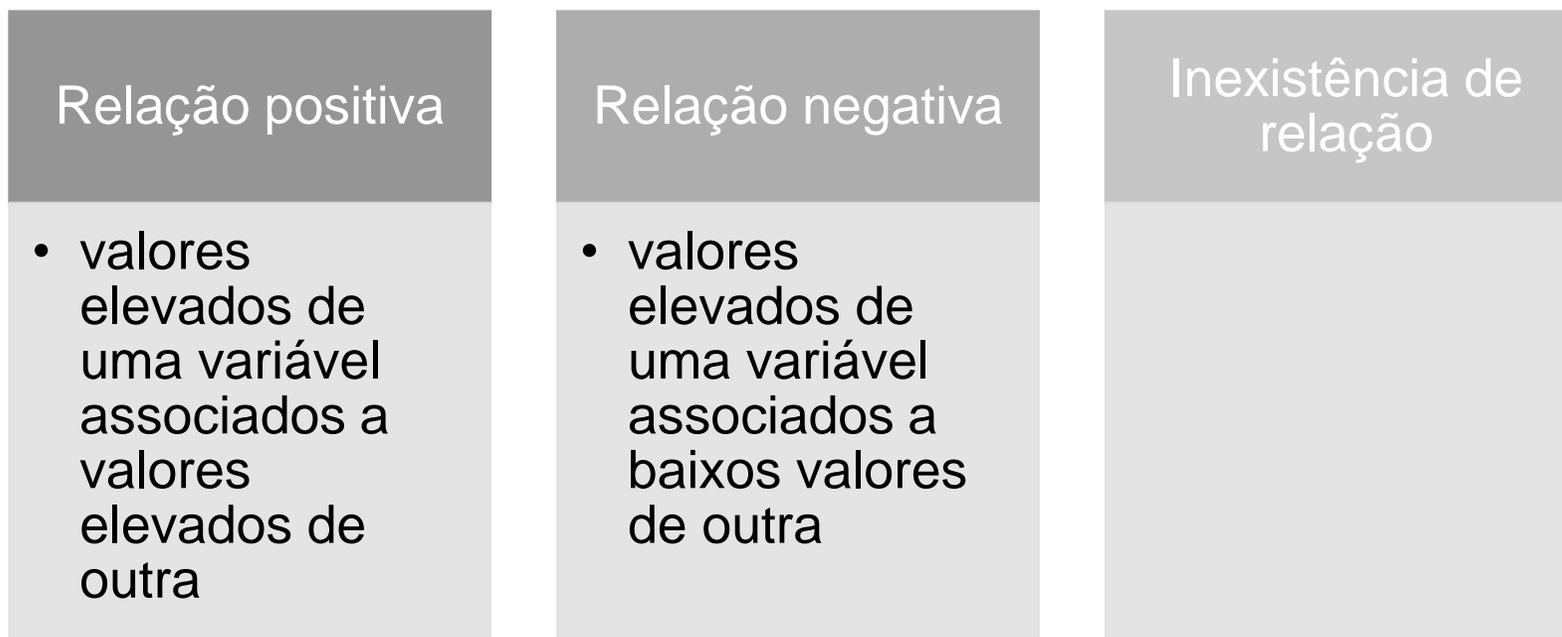


◆ Diagrama de dispersão e a nuvem de pontos

Representação gráfica dos pares (x,y)



- ◆ A partir da análise da nuvem de pontos podemos ter ideias sobre a relação entre as variáveis – até que ponto elas estão correlacionadas
- ◆ **Tipos de relações possíveis:**



◆ Como medir o nível da correlação?

- O conceito de **covariância** - as variáveis variam no mesmo sentido ou em sentido inverso

$$S_{YX} = \frac{\sum_{j=1}^N (x_j - \bar{X})(y_j - \bar{Y})}{n}$$

◆ Interpretação:

- Se quando uma variável se afasta da média a outra também se afasta, estão relacionadas
- Se x acima da sua média e y também acima média variam no mesmo sentido
- Se x acima da média e y abaixo da média variam inversamente

◆ Como medir o nível da correlação?

- Uma alternativa - o conceito de **correlação linear** - a covariância vem expressa na mesma unidade das variáveis – vantagem de obter uma medida normalizada

$$r_{YX} = \frac{S_{YX}}{S_X \times S_Y} = \frac{\sum_{j=1}^N (x_j - \bar{X})(y_j - \bar{Y})}{\sqrt{\sum_{j=1}^N (x_j - \bar{X})^2 \sum_{j=1}^N (y_j - \bar{Y})^2}}$$

◆ Interpretação:

- Se $r_{YX} = 0$ ou $S_{YX} = 0$ – não há relação linear
- Se $r_{YX} = 1$ – relação positiva perfeita
- Se $r_{YX} = -1$ – relação negativa perfeita
- Quanto mais perto de 1 em valor absoluto melhor a relação
- NOTA: isto é uma medida da relação linear – se for igual a 0 não quer dizer que não há relação, só quer dizer que a relação não é linear – em alguns casos as relações podem ser linearizadas

- ◆ **O cálculo das medidas no *Excel***
 - Funções *covar* e *correl*