

# Análise de Informação Económica e Empresarial

---

## Aula 7: Redução de Dados: Medidas de Dispersão e Concentração



# Aula 7: Redução de Dados: Medidas de Dispersão e Concentração

---

## Conceitos Fundamentais

- ◆ Dispersão
- ◆ Intervalo de Variação (Amplitude Total) e Intervalo de Variação Interquartil (Q3-Q1)
- ◆ Dispersão Absoluta: Desvio Absoluto Médio, Desvio Padrão e Variância
- ◆ Dispersão Relativa: Coeficiente de Variação
- ◆ Concentração
- ◆ Coeficiente de Gini e Curva de Lorenz
- ◆ Diagrama de extremos e quartis ou Gráfico caixa e bigodes

## Tópicos

1. Medidas de Dispersão e de Concentração
2. Intervalo de Variação (Amplitude Total) e Intervalo de Variação Interquartil (Q3-Q1)
3. Dispersão Absoluta: Desvio Absoluto Médio, Desvio Padrão e Variância
4. Dispersão Relativa: Coeficiente de Variação
5. Coeficiente de Gini
6. Curva de Lorenz

**Exercício de consolidação:** Utilizar a base de dados de notas e calcular média aritmética, máximo, mínimo, 3º quartil, 1º quartil, amplitude total, intervalo de variação interquartil, desvio absoluto médio, desvio padrão, variância e coeficiente de variação

**Exercício de aplicação:** exercício de aplicação de medidas de concentração.

**Bibliografia:** Reis, Elizabeth (2005) Estatística Descritiva, Lisboa: Edições Sílabo, 6ª edição - Cap 5. Medidas de Dispersão e Concentração, pp 97-117

➤ Temos os dados:

$$X_1, X_2, X_3, \dots, X_n$$

- Calculámos valores que representavam toda a distribuição – as medidas de localização...
- *Até que ponto as medidas de localização calculadas são uma boa representação dos dados? Quão distintos são os elementos da população?*
- Necessitamos de uma **medida da heterogeneidade** dos valores!

➤ Medidas de dispersão: indicadores da heterogeneidade

- ◆ As medidas de **dispersão absolutas** – na unidade da variável
  - Amplitude total e amplitude interquartil
  - Desvio absoluto médio, desvio padrão e variância
- ◆ As medidas de **dispersão relativas**
  - Intervalo interquartil relativo
  - Coeficiente de variação

## ➤ Medidas de dispersão com base em medidas de posição

### ◆ Amplitude total ou Intervalo de variação

$$R_x = \max(X_i) - \min(X_i) = X_{max} - X_{min}$$

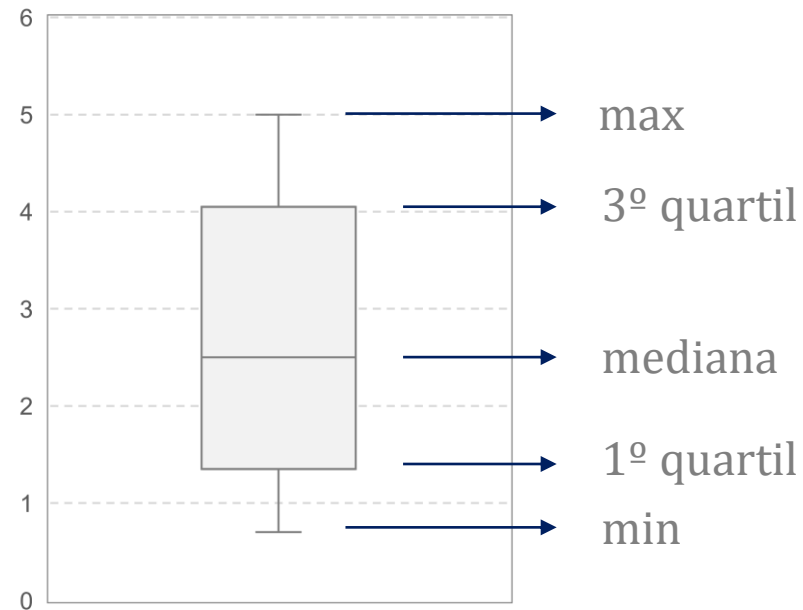
- Características do intervalo de variação
  - Facilidade de cálculo
  - Muito sensível aos valores extremos
  - Não tem em conta valores intermédios
  - Sensibilidade à dimensão da população ou da amostra

### ◆ Amplitude Interquartis ou Intervalo de variação interquartil

$$IQ = Q_3 - Q_1$$

- Características do intervalo de variação interquartil
  - Menor sensibilidade aos valores extremos
  - Não toma em consideração todos os valores da coleção

- ◆ **Diagrama de extremos e quartis**  
Avaliação da localização e dispersão a partir de medidas de posição



## ➤ Medidas de dispersão com base em todas as observações

Medir o afastamento entre todas as observações

### ◆ Desvio absoluto médio

- Dados não agrupados:

$$DM_x = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} = \frac{\sum_{i=1}^n |x_i - \mu|}{n}$$

- Dados agrupados/classificados:

$$DM_x = \frac{\sum_{i=1}^n n_i |v_i - \bar{x}|}{n} = \sum_{i=1}^n f_i |v_i - \bar{x}|$$

(ou  $C_i$  em vez  $v_i$ )

## ➤ Medidas de dispersão com base em todas as observações

Medem o afastamento entre todas as observações.

### ◆ Desvio-Padrão

- Dados não agrupados: 
$$s_x = \sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

- Dados agrupados/classificados: 
$$s_x = \sigma_x = \sqrt{\frac{\sum_{i=1}^n n_i (v_i - \bar{x})^2}{n}} = \sqrt{\sum_{i=1}^n f_i (v_i - \bar{x})^2}$$

### ◆ Variância

- Dados não agrupados: 
$$s_x^2 = \sigma_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

- Dados agrupados/classificados: 
$$s_x^2 = \sigma_x^2 = \frac{\sum_{i=1}^n n_i (v_i - \bar{x})^2}{n} = \sum_{i=1}^n f_i (v_i - \bar{x})^2$$



## ➤ Medidas de dispersão relativa

Permitem comparar distribuições com medidas de localização distintas.

### ◆ Intervalo interquartil relativo

$$IQR = \frac{Q_3 - Q_1}{Q_2}$$

### ◆ Coeficiente de variação

$$CV_x = \frac{s_x}{\bar{x}} = \frac{\sigma_x}{\mu}$$

## ➤ Medidas de concentração

Como se reparte, entre os elementos da população, o valor total de uma característica ordenável e somável (ex: vendas, rendimento, etc.)

### *Ideia:*

1. *Avaliar qual a proporção dessa característica que cada elemento da população detém.*
2. *Temos elevada concentração se poucos elementos da população detiverem uma elevada proporção do total.*

## ◆ Índice de Gini - uma forma de medir a concentração



$$IG = \frac{\sum_{i=1}^{n-1} (\text{cum } f_i - \text{cum } y_i)}{\sum_{i=1}^{n-1} \text{cum } f_i} = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

$$p_i = \text{cum } f_i \quad , \quad q_i = \text{cum } y_i$$

◆ **Exemplo:** rendimento de 8 trabalhadores

Rendimento ( $Y_i$ )	$F_i$	Rendimento Acumulado (cum $Y_i$ )	População Acumulada (cum $F_i$ )	% de rendimento (cum $y_i$ )	% de população (cum $f_i$ )
120	1	120	1	0,060	0,125
150	1	270	2	0,135	0,250
180	1	450	3	0,225	0,375
200	1	650	4	0,325	0,500
250	1	900	5	0,450	0,625
260	1	1160	6	0,580	0,750
300	1	1460	7	0,730	0,875
540	1	2000	8	1,000	1,000
<b>2000</b>	<b>8</b>				

- ◆ Representação gráfica: a curva de Lorenz

