

# Estatística Descritiva

Catarina Marques

Estatística I - Gestão

Ano Lectivo 2015/2016

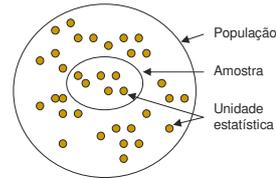
## Conceitos

- Estatística
  - Estatística é a ciência que se ocupa da obtenção de informação (...) com a finalidade de através de resultados probabilísticos adequados, inferir de uma amostra para a população(...), e eventualmente mesmo prever a evolução futura de um fenómeno. Por outras palavras, é um instrumento de leitura de informação, e da sua transformação em Conhecimento  
(Pestana, D. E Velosa, S.F., 2003, *Introdução à Probabilidade e à Estatística*, Lisboa, Fundação Calouste Gulbenkian).
- Utilidade da Estatística
  - Permite descrever e compreender relações entre variáveis de forma imediata: a informação é apresentada de modo a possibilitar uma rápida interpretação e identificação das relações mais importantes;
  - Permite a tomada de melhores e mais rápidas decisões: é possível controlar mais informação num espaço de tempo mais curto;
  - Facilita a tomada de decisões: o conhecimento de situações passadas e presentes, acompanhadas por uma previsão fundamentada da evolução futura, são as base para as tomadas de decisões.

Catarina Marques

## Conceitos: População, Unidade Estatística e Amostra

- **População** (ou Universo)
  - Conjunto de unidades com uma ou mais características comuns
  - **População alvo** - Conjunto de unidades que se pretendem estudar
  - **População inquirida** - Conjunto de unidades para as quais é operacionalmente possível recolher uma amostra
- **Unidade Estatística**
  - Um elemento da população
- **Amostra**
  - Subconjunto da População ( $n^{\circ}$  de elementos é  $n$ ).
  - Estudam-se apenas algumas características específicas, sendo cada uma delas representada por uma variável,  $X$ .
  - O processo de recolha de informação de toda a população é denominado de **censo** ou **recenseamento**. A consulta a um subconjunto da população é denominada de **amostragem** ou **sondagem**.



Catarina Marques

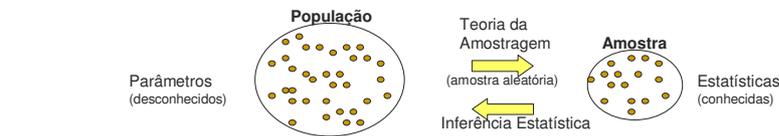
## Estatística Descritiva

- A Estatística Descritiva consiste na apresentação, análise e interpretação de um conjunto de dados, através da criação de instrumentos adequados:
  - utilização de tabelas para condensar os dados: **distribuição de frequências**
  - **representação gráfica**
  - cálculo de valores numéricos que caracterizam os dados de uma forma global: **medidas de estatística descritiva**.

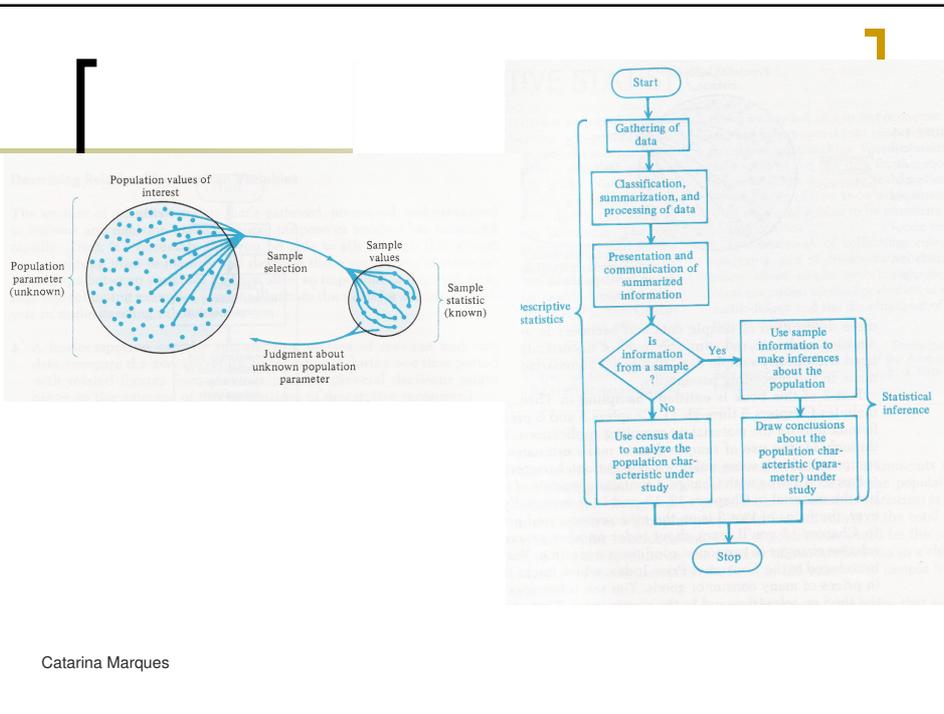
Catarina Marques

# Inferência Estatística

- **Inferência Estatística**
  - Permite generalizar à população certas conclusões, tendo em conta o resultados amostrais, com a indicação da precisão dessas inferências. O seu uso permite estimar as características (**parâmetros**) desconhecidas da população e testar se as hipóteses formuladas sobre as características desconhecidas são plausíveis.
- **Parâmetros** – Resumem as características da população e assumem valores numéricos fixos (constantes).
  - Ex: média da população ( $\mu$ ), variância populacional ( $\sigma^2$ ), desvio-padrão populacional ( $\sigma$ )
- **Estatísticas** – É toda a função que opera sobre a amostra. Dado que de uma população se podem retirar diversas amostras estamos perante valores que variam de amostra para amostra.
  - Ex: média amostral ( $\bar{x}$ ), a variância amostral ( $S^2$ ), desvio-padrão amostral ( $S$ )



Catarina Marques



Catarina Marques

# Distribuição de frequências

- **Frequência absoluta ( $F_i$ )** – corresponde ao número de vezes que um determinado valor se repete na amostra ou população
- **Frequência relativa ( $f_i$ )** – corresponde ao número de vezes que o valor da variável ocorre relativamente ao total da amostra ( $n$ ) ou total da população ( $N$ ).

$$f_i = \frac{F_i}{n} \quad n = \sum_i F_i$$

- **Frequências acumuladas absolutas (cum  $F_i$ ) ou relativas (cum  $f_i$ )** – representam o número de dados com valor inferior ou igual à observação em causa.

OUTPUT do SPSS:  
Exemplo de uma variável ordinal

Frequência prática desporto (mãe)

	Frequency	Percent	Cumulative Percent
De vez em quando	25	25,0	25,0
Algumas vezes	50	50,0	75,0
Sempre	25	25,0	100,0
Total	100	100,0	

**$F_i$**  (Frequency)

**$f_i$**  (expresso em %)

**Cum  $f_i$**  (expresso em %)

Catarina Marques

# OUTPUT do SPSS:

Exemplo de uma variável discreta, medida em escala de razão

Tabagismo materno (nº de cigarros por dia)

	Frequency	Percent	Cumulative Percent
0	70	70,0	70,0
10	10	10,0	80,0
15	8	8,0	88,0
20	8	8,0	96,0
25	2	2,0	98,0
30	2	2,0	100,0
Total	100	100,0	

OUTPUT do SPSS:  
Exemplo de uma variável discreta –  
**necessidade de construção de classes**

Idade da mãe (anos)

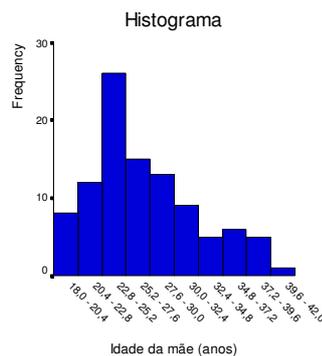
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 18	1	1,0	1,0	1,0
19	3	3,0	3,0	4,0
20	4	4,0	4,0	8,0
21	7	7,0	7,0	15,0
22	5	5,0	5,0	20,0
23	8	8,0	8,0	28,0
24	10	10,0	10,0	38,0
25	8	8,0	8,0	46,0
26	9	9,0	9,0	55,0
27	6	6,0	6,0	61,0
28	6	6,0	6,0	67,0
29	4	4,0	4,0	71,0
30	3	3,0	3,0	74,0
31	4	4,0	4,0	78,0
32	5	5,0	5,0	83,0
33	3	3,0	3,0	86,0
34	2	2,0	2,0	88,0
35	3	3,0	3,0	91,0
36	1	1,0	1,0	92,0
37	2	2,0	2,0	94,0
38	3	3,0	3,0	97,0
39	1	1,0	1,0	98,0
41	1	1,0	1,0	99,0
42	1	1,0	1,0	100,0
Total	100	100,0	100,0	

## Representações gráficas

**Histograma** – é um gráfico de barras adjacentes em que a área de cada barra representa a frequência da classe que representa

**Polígono de frequências** – linha poligonal fechada que se obtém unindo os pontos médios de cada classe, se estas tiverem a mesma amplitude. de modo a que a área total do histograma seja igual à área total

Na construção do histograma e polígono de frequências, se as classes tiverem amplitudes diferentes tomam-se, para representação, **os valores das frequências por unidade de classe**, ou seja,  $F_i / a_i$ , no caso das frequências absolutas  $f_i / a_i$ , no caso das frequências relativas



Catarina Marques

## Medidas descritivas

- **Medidas de localização** – localizam os valores da variável no eixo dos  $n^{\circ}$ s reais.
  - Medidas de tendência central – identificam o “centro” de uma distribuição.
    - Média, Moda, Mediana
  - Medidas de tendência não central – apontam para outras posições da distribuição.
    - Quantis (Quartis, Decis e Percentis)
- **Medidas de dispersão** – permitem destacar o grau de homogeneidade ou heterogeneidade existente entre os dados.
  - **Medidas Absolutas**: Intervalo de Variação, Intervalo Interquartilico, Desvio Médio, Variância, Desvio-Padrão
  - **Medidas Relativas**: Coeficiente de Variação
- Medidas de assimetria
- Medidas de curtose (ou achatamento)
- Medidas de concentração

Catarina Marques

## Medidas de localização

### Medidas de tendência central

**Média** – Corresponde ao valor que a variável teria se todas as observações fossem iguais.

- Só pode ser calculada em variáveis assumidas como quantitativas
- Pode ser um valor não assumido pela variável
- Tem de se encontrar entre o valor mais pequeno e o mais elevado da amostra

Dados simples

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Dados organizados em tabelas de frequência

$$\bar{x} = \frac{\sum_{i=1}^k x_i F_i}{n} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

Catarina Marques

## Medidas de localização

### Medidas de tendência central

**Mediana** – valor abaixo e acima do qual se registaram 50% das observações.

- Pode ser calculada para variáveis ordinais ou para variáveis quantitativas.
- Como consiste no valor da variável que divide uma dada distribuição ao meio, o seu cálculo obriga a que se ordene previamente os valores da variável de forma crescente (ou

decrecente).  $n$  par  $t = \frac{n}{2}$   $Me = \frac{x_t + x_{t+1}}{2}$   $n$  ímpar  $t = \frac{n+1}{2}$   $Me = x_t$

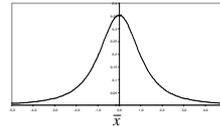
- Para a determinação da mediana de um conjunto de dados não agregados em classes há que ter em atenção se o nº de observações é par ou ímpar.

Catarina Marques

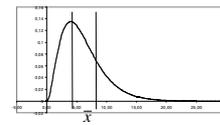
## Estudo da assimetria de uma distribuição através das medidas de tendência central

- Distribuição simétrica

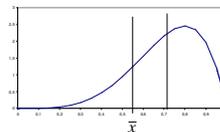
$$\bar{x} = Mo = Me$$



- Distribuição assimétrica positiva ou enviesada à esquerda  $\bar{x} \geq Me \geq Mo$



- Distribuição assimétrica negativa ou enviesada à direita  $\bar{x} \leq Me \leq Mo$



Catarina Marques

## Medidas de localização

- Medidas de tendência não central

**Quantis** – Por quantis de ordem  $k$  entende-se os  $k-1$  valores que dividem um conjunto de observações ordenadas em  $k$  partes.

Se  $k=4$  divide-se a distribuição em 4 partes iguais → QUARTIS  
Se  $k=10$  divide-se a distribuição em 10 partes iguais → DECIS  
Se  $k=100$  divide-se a distribuição em 100 partes iguais → PERCENTIS

- Podem calcular-se desde que a variável seja pelo menos ordinal
- O cálculo dos quantis efectua-se de modo semelhante ao da mediana.
- A posição do Quantil  $p$  é obtida por  $\frac{n \times p}{k}$ .

Catarina Marques

# Medidas de dispersão

## Variância e desvio-padrão

- Só podem ser calculados para variáveis numéricas
- A variância é uma medida de dispersão que avalia os afastamentos em relação à média das observações
  - É expressa no quadrado das unidades da variável
- O desvio-padrão é a raiz quadrada positiva da variância
  - É expressa nas mesmas unidades que a variável
- Nos programas informáticos é utilizado o desvio-padrão corrigido

$$s = \sqrt{s^2}$$

Dados organizados em tabelas de frequência

$$\sigma^2 = s^2 = \text{var}(x) = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 F_i}{n} = \sum_{i=1}^k (x_i - \bar{x})^2 f_i \quad \sigma^2 = s^2 = \text{var}(x) = \frac{\sum_{i=1}^k x_i^2 F_i}{n} - \bar{x}^2 = \sum_{i=1}^k x_i^2 f_i - \bar{x}^2$$

## Coefficiente de variação – medida relativa.

- Exprime o peso percentual do desvio-padrão face à média.
- É um indicador da grandeza da dispersão numa determinada distribuição.

Catarina Marques

$$CV = \frac{s}{\bar{x}} \times 100$$

# Medidas Descritivas

## Exemplo

Statistics		
Idade da mãe (anos)		
N	Valid	100
	Missing	0
Mean		27,06
Median		26,00
Mode		24
Std. Deviation		5,458
Variance		29,792
Range		24
Minimum		18
Maximum		42
Percentiles	25	23,00
	50	26,00
	60	27,00
	75	31,00
	80	32,00

Catarina Marques