



INTRODUÇÃO À ESTATÍSTICA



População

Colecção de unidades individuais, com uma ou mais características comuns, que se pretendem analisar.

- Exemplos:
 - Alunos da Universidade de Coimbra
 - Potenciais eleitores para as eleições presidenciais

Parâmetros

Características numéricas que descrevem a população. Estas características são, em geral desconhecidas.

- Exemplos:
 - Alunos da Universidade de Coimbra
 - Altura média dos alunos
 - Classificações médias obtidas a Estatística...
 - Potenciais eleitores para as eleições presidenciais
 - Proporção de eleitores decididos a votar
 - Idade média dos eleitores...



Amostra

Subconjunto da população, que se observa com o objectivo de tirar conclusões para a população de onde foi retirada.

- Exemplos:
 - Alunos da Universidade de Coimbra
 - Amostra de 100 alunos da Universidades de Coimbra
 - Potenciais eleitores para as eleições presidenciais
 - Amostra de 1000 potenciais eleitores para as eleições presidenciais



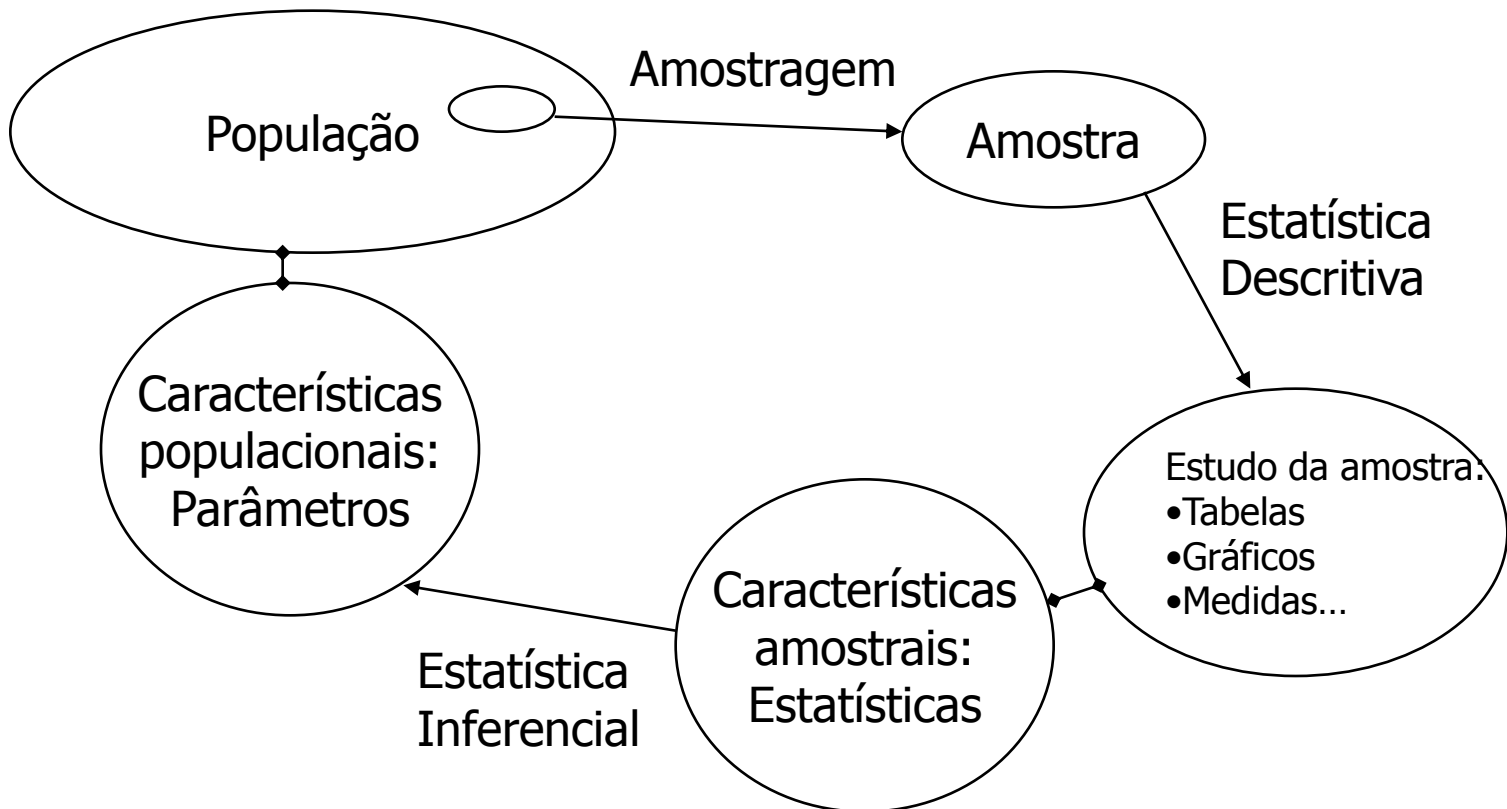
Estatística

Característica numérica que descreve a amostra. Calcula-se o valor da estatística a partir dos valores observados na amostra. Utiliza-se a estatística para estimar um parâmetro desconhecido.

- Exemplos:
 - Alunos da Universidade de Coimbra
 - Altura média dos 100 alunos da amostra
 - Potenciais eleitores para as eleições presidenciais
 - Proporção de eleitores que estão decididos a votar dos 1000 eleitores da amostra



Procedimento estatístico





Recenseamento ou censo

Estudo científico de um universo de pessoas, instituições ou objectos físicos com o propósito de adquirir conhecimentos, observando todos os seus elementos, e fazer juízos quantitativos acerca das características importantes desse universo.

- Exemplo:
 - XIV Recenseamento Geral da População Portuguesa (2001) (Instituto Nacional de Estatística – INE)

<http://www.ine.pt>



Sondagem

Estudo científico de uma parte de uma população com o objectivo de estudar atitudes, hábitos e preferências da população relativamente a acontecimentos, circunstâncias e assuntos de interesse comum.

- Exemplos:
 - Sondagens para obter informação acerca da atitude dos eleitores
 - Sondagens para testar as preferências dos consumidores



Fases de uma sondagem

- Escolha da amostra (amostragem)
- Obtenção de informação
- Análise dos dados
- Relatório final



Representatividade da amostra

A amostra deve ser tão representativa quanto possível da população que se está a estudar – deve-se evitar o enviesamento.

- Exemplos de amostras enviesadas ou tendenciosas:
 - Amostragem por conveniência
 - Utilizar uma amostra de sócios do FCP para prever o vencedor de próximo Benfica-Porto
 - Utilizar uma amostra de alunos de um curso para tirar conclusões acerca do aproveitamento dos alunos universitários
 - Amostragem por resposta voluntária
 - Certas “sondagens” realizadas pelas estações de televisão utilizando respostas voluntárias, por exemplo, por SMS



AMOSTRAGEM



Amostragem

- Com reposição
 - O elemento seleccionado é repostado na população, podendo, eventualmente, ser seleccionado mais vezes
- Sem reposição
 - Um elemento, uma vez seleccionado, é removido da população, não podendo, portanto, ser outra vez seleccionado



Exemplo

- Suponha que num almoço de 4 amigos, o número de moedas que cada um tem nos bolsos foi o seguinte:

$$x_1=6 \quad x_2=2 \quad x_3=4 \quad x_4=3$$

- Apresente todas as amostras possíveis de dimensão 2, em amostragem com e sem reposição



Solução

Com reposição

(x_1, x_1) 6,6	(x_2, x_2) 2,2	(x_3, x_3) 4,4	(x_4, x_4) 3,3
(x_1, x_2) 6,2	(x_2, x_3) 2,4	(x_3, x_4) 4,3	
(x_1, x_3) 6,4	(x_2, x_4) 2,3		
(x_1, x_4) 6,3			

Sem reposição

(x_1, x_2) 6,2	(x_2, x_3) 2,4	(x_3, x_4) 4,3
(x_1, x_3) 6,4	(x_2, x_4) 2,3	
(x_1, x_4) 6,3		



Solução

Com reposição

6	2,6	3,6	4,6	6,6
4	2,4	3,4	4,4	6,4
3	2,3	3,3	4,3	6,3
2	2,2	3,2	4,2	6,2
	2	3	4	6



Solução

Sem reposição

6	2,6	3,6	4,6	-
4	2,4	3,4	-	6,4
3	2,3	-	4,3	6,3
2	-	3,2	4,2	6,2
	2	3	4	6



Comentários

- Todas as combinações possíveis na amostragem sem reposição são também possíveis na amostragem com reposição
- Contudo, o inverso não é verdade já que não se observam repetições na amostragem sem reposição
- Uma vez que uma observação repetida não contribui com nova informação sobre a população amostrada, a amostragem sem reposição é, por isso, considerada mais eficiente.



Amostra aleatória simples

Dada uma população, uma amostra aleatória simples de dimensão n é um conjunto de n unidades da população, tal que qualquer outro conjunto de n unidades teria igual probabilidade de ser seleccionado.

- Minimiza o enviesamento
- Tem em conta o princípio da aleatoriedade
- Recolha sem reposição



ESTATÍSTICA DESCRITIVA



Classificação das Variáveis

- **Qualitativas**

- *informação que identifica alguma qualidade, categoria ou característica, não susceptível de medida, mas de classificação, assumindo várias modalidades.*

Ex.: **cor dos olhos, desporto preferido, ...**

- **Quantitativas**

- *informação resultante de características susceptíveis de serem medidas, apresentando-se com diferentes intensidades.*
 - **Discretas** – *pode tomar um n^o finito (ou infinito numerável) de valores distintos*

Ex.: **n^o de acidentes, resultado do lançamento de um dado, ...**

- **Contínuas** – *pode tomar todos os valores numéricos, compreendidos no seu intervalo de variação*

Ex.: **peso, altura, nível de colesterol no sangue, ...**



Organização dos Dados

- Tabelas de frequências (absolutas e relativas)
- Representações gráficas
 - Dados qualitativos
 - Diagrama circular
 - Diagrama de barras
 - Dados quantitativos
 - Discretos
 - Diagrama de barras
 - Contínuos
 - Histograma
 - Diagrama Caixa de Bigodes



Construção de um Histograma

- Número de observações, n
- Amplitude, R
- Número de classes, k
 - Regra de *Sturges*

$$k = 1 + 3.3 \log(n)$$

- Intervalo de classe, R/k
- Extremos de classe

n	k
25	5-6
50	6-7
100	7-8
500	9-10
1000	10-11



Exemplo

A tabela apresenta os tempos de espera numa fila de supermercado de sujeitos seleccionados aleatoriamente

4	18	8	25	5.5	7
7	26	8	16	2	1
12	3	2	9	16	4
21	7	13	27	8	8
27	4	34.5	19	7	5
18	9	12	16	2	6
12	10	7	21	3	1
0.5	11	10	13	4	5
20	1.5	5	7	12	2
8.5	12	5	10	18	0.5



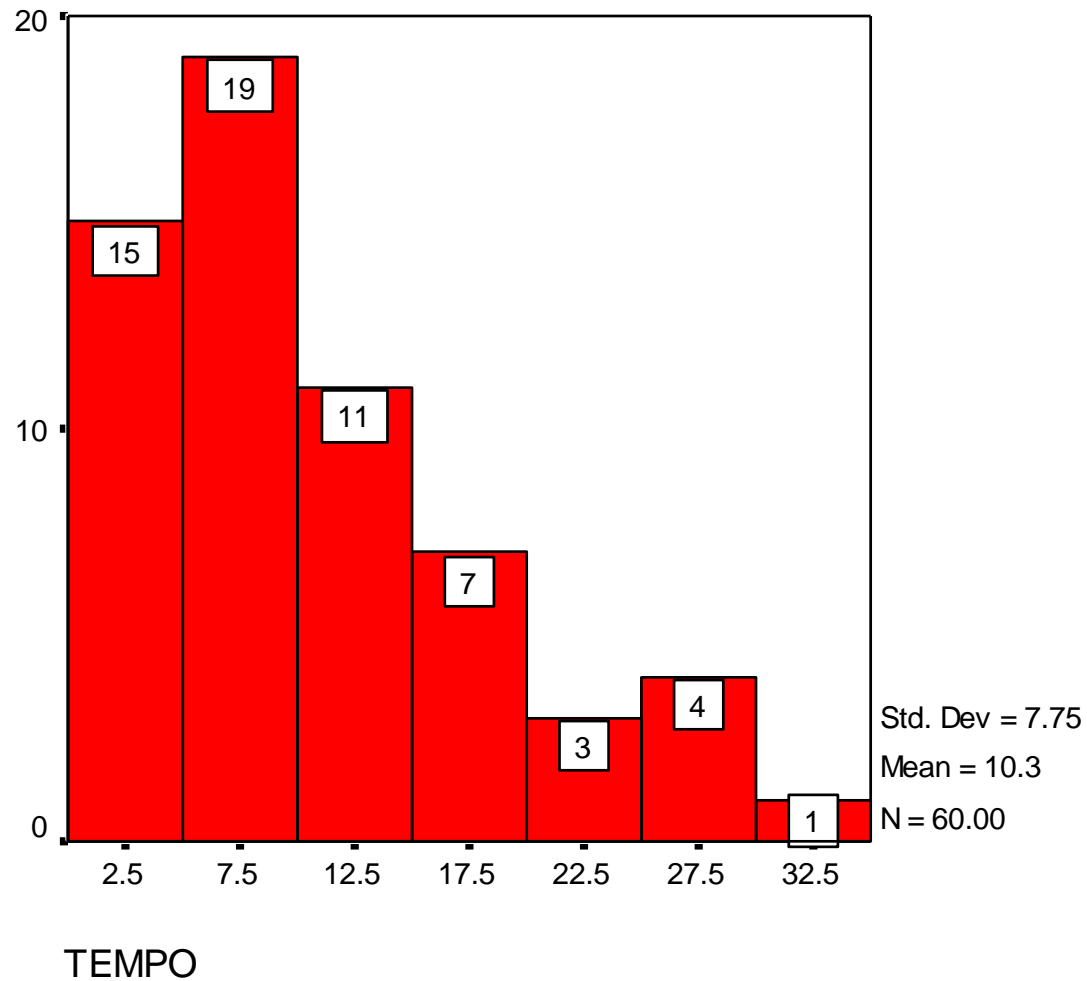
Tabela de Frequências

- Número de observações, $n=60$
- Amplitude, $R=34.5-0.5=34.0$
- Número de classes, $k=7$
- Intervalo de classe, $R/k=34/7 \approx 4.8 \approx 5.0$
- Extremos de classe, $\min=0.5$

0-5	15
5-10	19
10-15	11
15-20	7
20-25	3
25-30	4
30-35	1



Histograma





Estatísticas

- Medidas de localização
 - Média
 - Mediana
 - Quartis e percentis
 - Moda
- Medidas de dispersão ou variabilidade
 - Desvio padrão e variância
 - Amplitude
 - Amplitude inter-quartil



Estatísticas

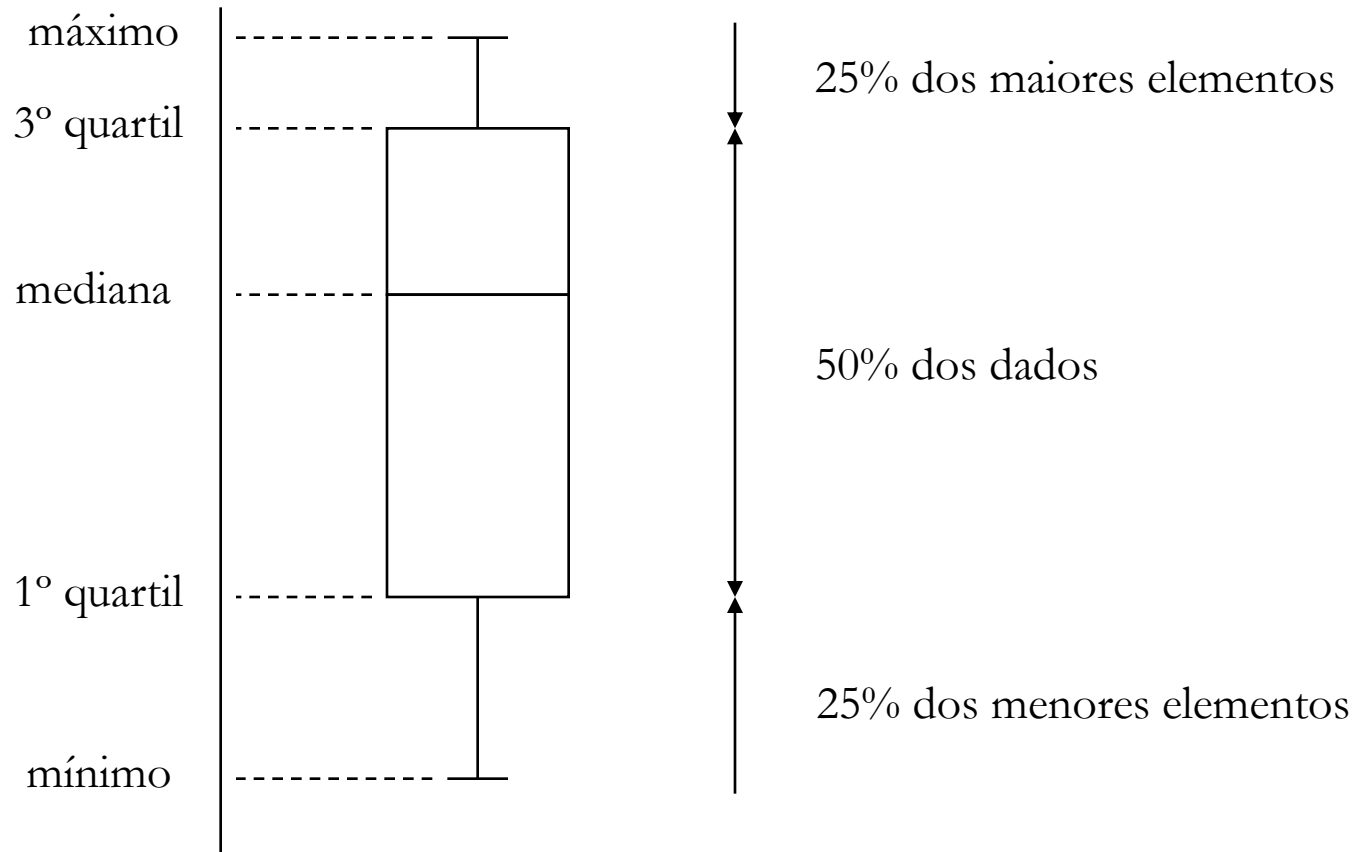
Statistics

TEMPO

N	Valid	60
	Missing	0
Mean		10.267
Median		8.000
Std. Deviation		7.7462
Variance		60.0040
Range		34.0
Minimum		.5
Maximum		34.5
Percentiles	10	2.000
	20	4.000
	25	4.250
	30	5.000
	40	7.000
	50	8.000
	60	10.000
	70	12.000
	75	15.250
	80	17.600
	90	21.000

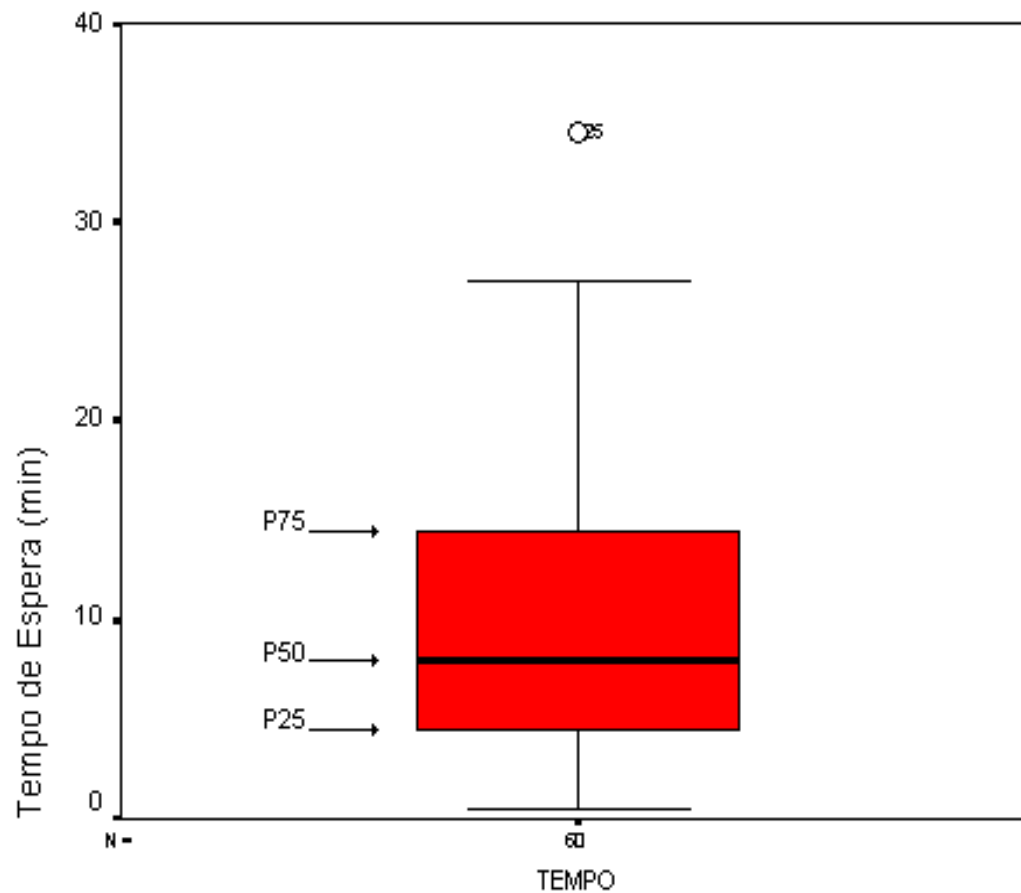


Caixa de Bigodes





Caixa de Bigodes





Medidas de Localização

- Média aritmética

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \qquad \bar{X} = \frac{\sum_{i=1}^k f_i X_i}{n}$$

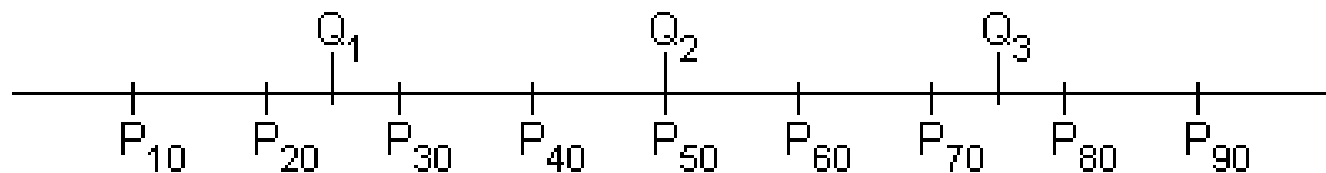
- Mediana

$$Md = X_{(n+1)/2}$$

- Quartis

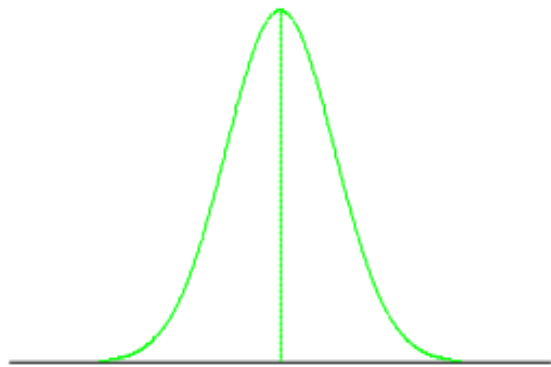
$$Q_1 = X_{(n+1)/4} \qquad Q_2 = Md \qquad Q_3 = X_{(n+1-\text{subscrito de } Q_1)}$$

- Percentis





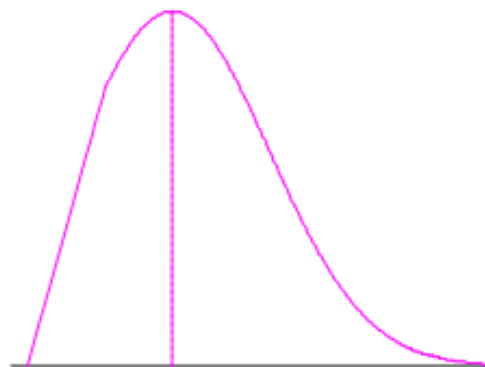
Comparação entre as medidas de localização



Distribuição Simétrica



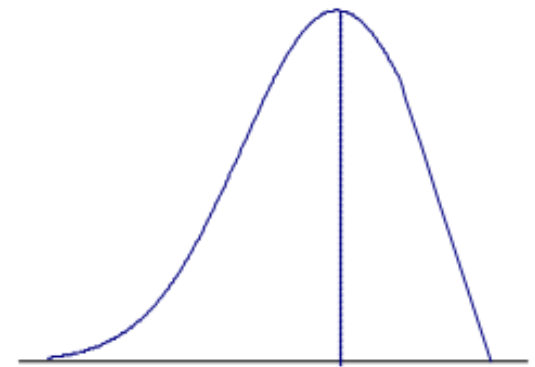
Moda = Mediana = Média



Distribuição Assimétrica
à direita



Moda < Mediana < Média



Distribuição Assimétrica
à esquerda



Moda > Mediana > Média



Medidas de Dispersão

- Amplitude

$$R = X_{(n)} - X_{(1)}$$

- Distância interquartílica

$$DIQ = Q_3 - Q_1$$

- Variância

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

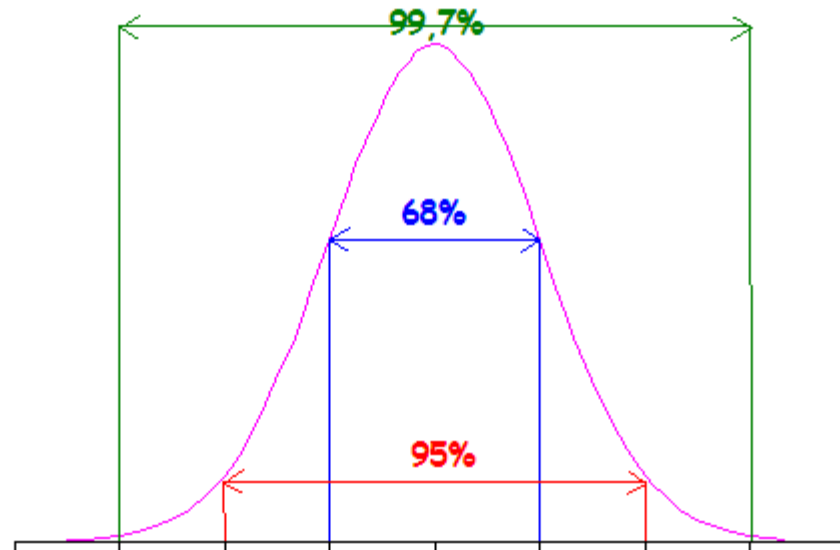
- Desvio padrão

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

- Coeficiente de variação

$$V = \frac{s}{\bar{X}}$$

Distribuição Normal



- ❖ Distribuição simétrica em forma de sino, centrada em μ .
- ❖ 68% das observações pertencem a $]\mu - \sigma ; \mu + \sigma [$;
- ❖ 95% das observações pertencem a $]\mu - 2\sigma ; \mu + 2\sigma [$;
- ❖ 99,7% das observações pertencem a $]\mu - 3\sigma ; \mu + 3\sigma [$



Alguns sites interessantes...

- ALEA - Acção Local de Estatística Aplicada
 - <http://alea-estp.ine.pt/>
- INE – Instituto Nacional de Estatística
 - <http://www.ine.pt/>