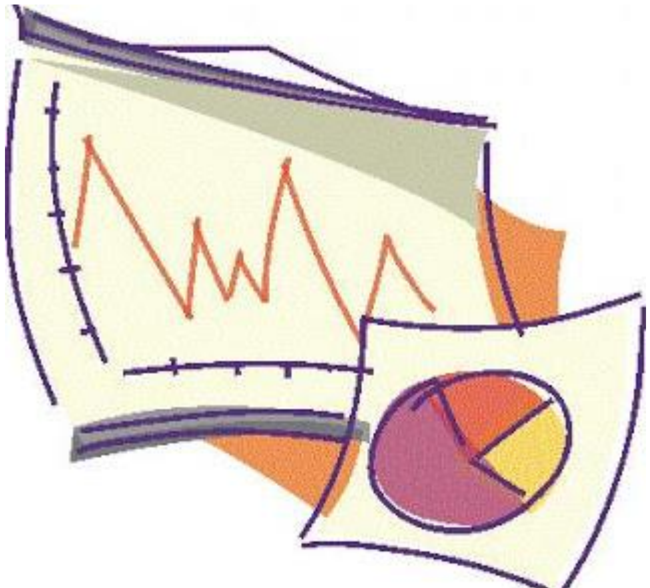
Estatística Descritiva



João Antunes

CONCEITOS

A ESTATÍSTICA

É um ramo da Matemática que nos ajuda a recolher, a organizar e a interpretar dados através de métodos e técnicas especificas, de modo a destacar o que é mais importante.

Estatística **DESCRITIVA**

Descreve as características importantes de um conjunto conhecido de dados populacionais

Estatística **INDUTIVA**

Utiliza os dados amostrais para fazer inferências (ou generalizações) sobre uma população

CONCEITOS

Nos mais variados aspetos da nossa vida, está presente a estatística:

- na política
- na investigação cientifica
- no desenvolvimento tecnológico
- na agricultura
- na industria farmacêutica
- na ciências exatas
- nas decisões das instituições governamentais
- no marketing
- na Internet
- no ensino

CONCEITOS

POPULAÇÃO

é uma coleção de elementos individuais, por exemplo, animais, pessoas, objetos, ou resultados experimentais que tenham uma ou mais características em comum e que se pretendam analisar.

AMOSTRA

Subconjunto da população escolhido corretamente, de modo a poder ser considerado como representativo da população.

DADOS ESTATÍSTICOS

são observações de atributos/variáveis (como medidas, respostas de inquéritos, registos de idade, sexo, naturalidade, etc.) de elementos que compõem uma população ou amostra

INSTITUTO SUPERIOR DE TECNOLOGIAS AVANÇADAS

CONCEITOS

CENSOS 2011
XV recenseamento geral da população
V recenseamento geral da habitação

RECENSEAMENTO

Processo em que para além da recolha de informação está também envolvida a análise de todos os elementos da população e o estudo das suas características importantes

SONDAGEM

Estudo estatístico de uma população, feito através de uma amostra, destinado a estudar uma ou mais características tal como elas se apresentam nessa população.

CONCEITOS

PARÂMETRO

uma medida numérica que descreve alguma característica de uma população.

POPULAÇÃO ⇔ **PARÂMETRO**

ESTATÍSTICA

uma medida numérica que descreve alguma característica de uma amostra

AMOSTRA ⇔ **ESTATÍSTICA**

CONCEITOS

PORQUE RAZÃO NÃO SE ESTUDA SEMPRE A POPULAÇÃO?

- Devido à população ser infinita ou poder ser considerada como tal.
  Ex: Temperaturas nos pontos da superfície da Terra, num dado instante.

- Devido à recolha da informação obrigar à destruição total dos elementos em estudo.
  Ex: Tempo de vida de determinado tipo de lâmpada.

- Devido ao tempo e ao custo dispendioso de recolha de toda a informação.
  Ex: Consulta de opinião pública sobre um candidato

CONCEITOS

AMOSTRA REPRESENTATIVA

- Todos os indivíduos da população devem ter igual probabilidade de ser selecionados

- A População deve estar bem definida logo à partida, ou seja desde o início do estudo

- Dimensão da amostragem deve ser adequada, ou seja nela devem figurar toda a variedade de subgrupos existentes na população

AMOSTRA ENVIESADA

- Não representativa de toda a população

- Quando é obtida por resposta voluntária

- Quando se recorre a processo de amostragem por conveniência

CONCEITOS

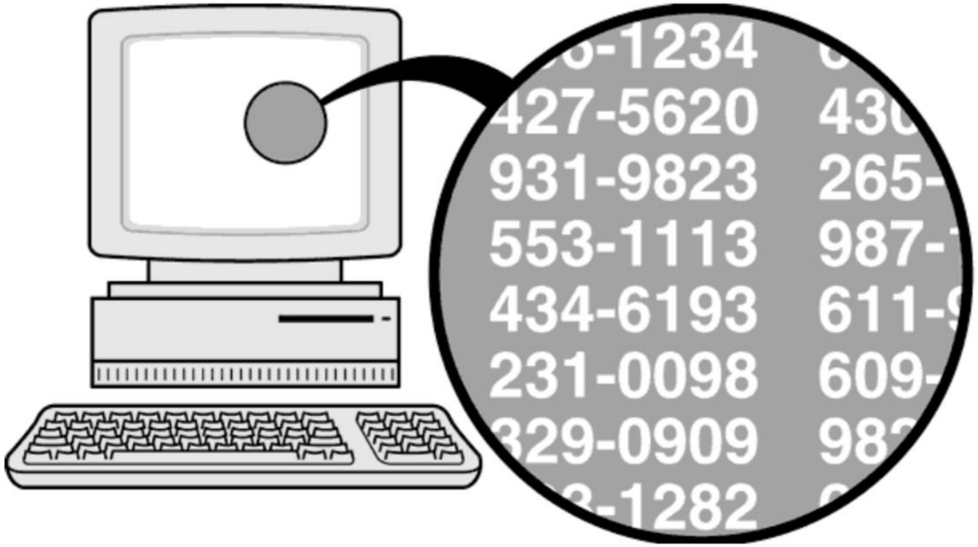
MÉTODOS DE AMOSTRAGEM

- Aleatória
- Sistemática
- por conveniência
- Estratificada
- por *clusters*

CONCEITOS

Amostragem Aleatória

Seleção tal que cada um tem igual possibilidade de ser escolhido.



CONCEITOS

TABELAS ALEATÓRIAS

é uma listagem dos dígitos 0, 1, 2, 3, 4, 5, 6, 7, 8 ou 9 tal que qualquer um dos dígitos considerados tem igual possibilidade de figurar em qualquer posição da lista e a posição em que figura cada digito é independente das posições dos outros dígitos.

Linha								
101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335
106	68417	35013	15529	72765	85089	57067	50211	47487
107	82739	57890	20807	47511	81676	55300	94383	14893
108	60940	72024	17868	24943	61790	90656	87964	18883
109	36009	19365	15412	39638	85453	46816	83485	41979

CONCEITOS

PR. 1

Considere a população constituída pelos todos os alunos do 1º ano do ISTEC (93 alunos) em que a característica de interesse a estudar é a ALTURA média desses alunos.

1º) A partir da tabela aleatória atrás apresentada, selecione uma amostra composta por 10 alunos

2º) Indique outro processo de obter uma amostra aleatória

CONCEITOS

Amostragem Sistemática

Selecionar a partir de um ponto inicial e depois selecionar a cada K elemento na população

CONCEITOS

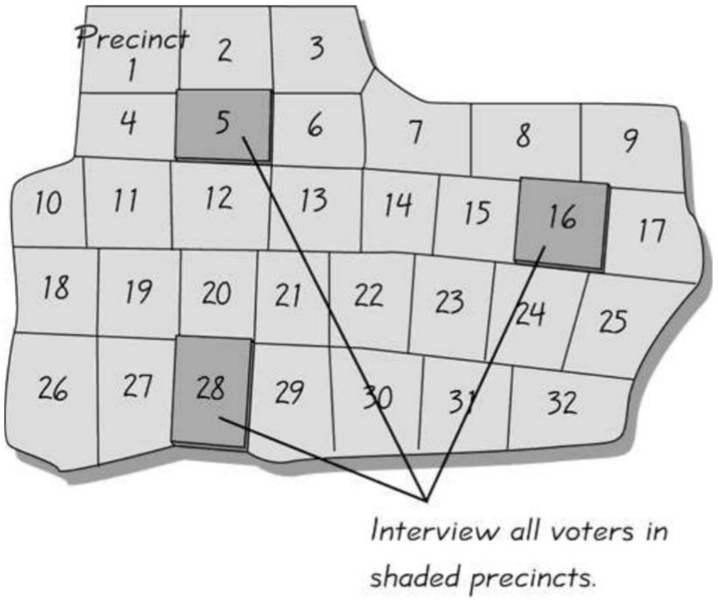
Amostragem Estratificada

subdividir a população em, pelo menos, dois distintos que partilham alguma característica e, em seguida, recolher uma amostra de cada um dos subgrupos

CONCEITOS

Amostragem por *clusters*

dividir a população em secções (ou clusters); selecionar aleatoriamente alguns desses clusters; escolher todos os membros dos clusters selecionados



CONCEITOS

Amostragem por Conveniência

Utilizar os resultados que são fáceis de obter

CONCEITOS

Atributos ou Variáveis

Qualitativos

Quando as características/atributos não são mensuráveis.
Exemplos: naturalidade, sexo

Quantitativos

As características são mensuráveis . O atributo varia de elemento para elemento da população   Exemplo: idade

Dados dicotómicos: apenas assumem dois valores

Tipo de dados

Discretos

Quando o número de valores possíveis é finito ou contável.

Exemplos: nº de reclamações, acidentes laborais

Contínuos

Quando o número de valores possíveis é infinito e corresponde a alguma escala contínua que contempla uma amplitude de valores sem interrupções ou saltos

Exemplos: Altura dos alunos; Temperatura

CONCEITOS

EX. 1

Num estudo para analisar a taxa de germinação de um certo tipo de cereal foram semeadas cinco sementes em cada um de 50 vasos iguais com o mesmo tipo de solo.

O número de sementes germinadas em cada vaso está registado a seguir:

1 0 1 2 1 3 2 0 0 1 4 0 2 1 0 2 4 1 2 0 3 5 3 0 2

1 3 3 0 4 0 2 5 3 0 2 5 1 1 0 4 4 1 2 1 0 5 1 2 3

Neste caso os **dados são de natureza discreta, com um número pequeno de valores distintos**.

CONCEITOS

EX. 2

Um dos principais indicadores da poluição atmosférica nas grandes cidades é a concentração de ozono na atmosfera. Num dado Verão registou-se 75 valores dessa concentração (em μg/m3), numa dada cidade:

3.5	6.2	3.0	3.1	5.1	6.0	1.4	5.7	1.7	4.4	6.2	4.4	6.8	9.4	1.1
6.6	3.1	4.7	2.4	6.8	7.5	5.4	5.8	5.6	6.8	6.6	5.8	5.6	4.7	6.0
5.5	2.5	3.4	5.3	5.7	5.8	2.0	6.2	5.6	4.0	7.6	4.7	7.6	7.4	3.7
2.8	3.4	3.5	3.8	5.5	4.4	2.5	11.7	4.1	4.5	5.8	4.7	3.7	6.6	6.7
4.2	5.9	3.0	3.3	4.1	3.9	5.4	1.6	6.0	9.4	6.6	6.1	6.5	1.4	1.4

Neste caso os **dados são de natureza contínua**

CONCEITOS

Escalas para os dados

- ✓Nominal
- ✓Ordinal
- ✓Intervalar
- ✓Absoluta ou de razões

CONCEITOS

- ✓ **Escala NOMINAL** - caracterizada por dados tais como nomes, etiquetas ou categorias. Os dados não têm qualquer relação de ordem (do mais pequeno para o maior)

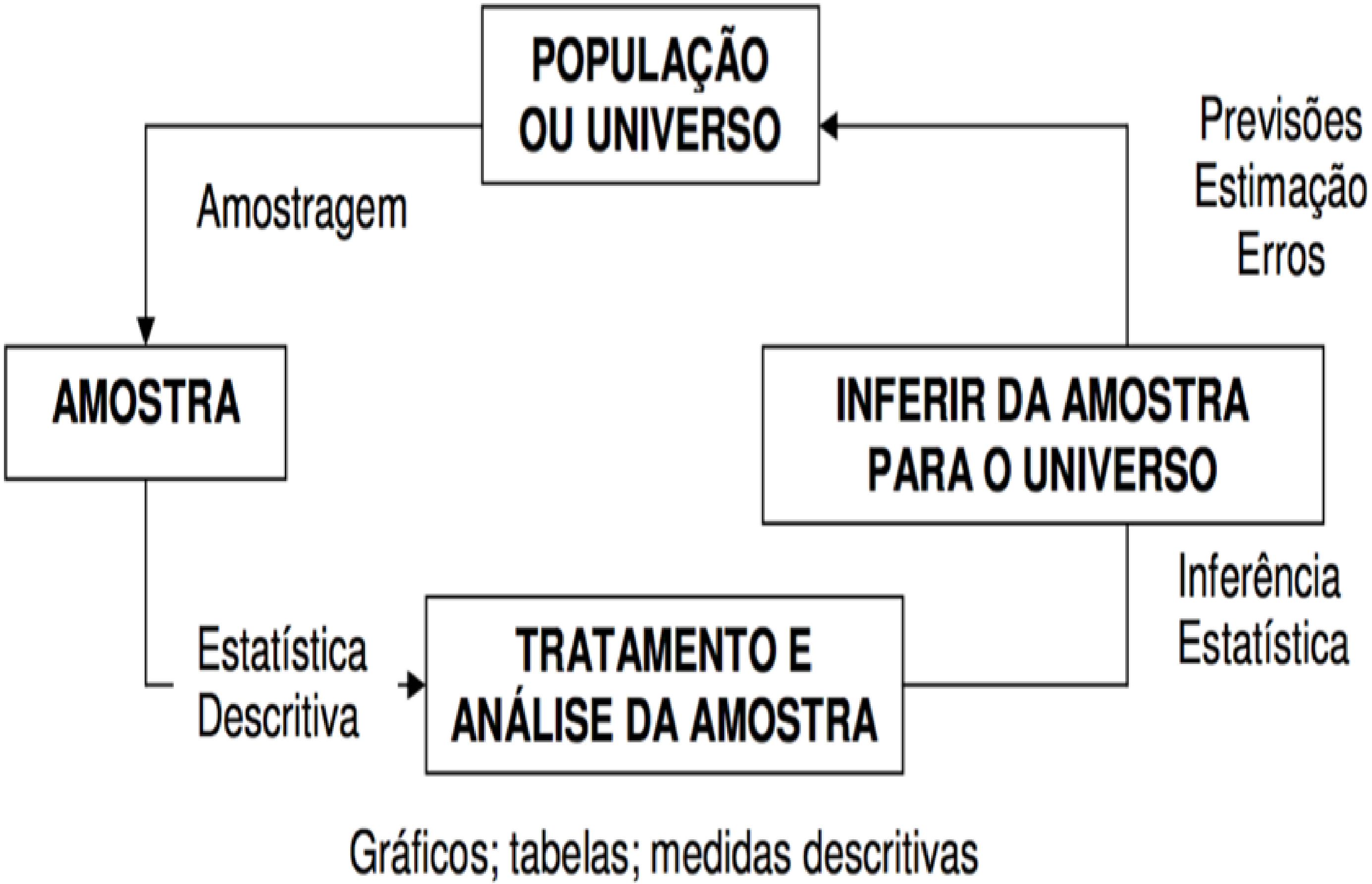
- ✓ **Escala ORDINAL -** envolve dados que podem ser ordenados, mas as diferenças entre eles ou não podem ser calculadas ou não fazem sentido

- ✓ **Escala INTERVALAR** - como a escala ordinal, mas onde é possível calcular diferenças. No entanto, não existe um zero natural (que pudesse corresponder a ausência da característica) diferenças possíveis mas sem zero natural

- ✓ **Escala ABSOLUTA -** o zero da escala corresponde à anulação da característica em estudo. São possíveis comparações quer através de diferenças quer através de quocientes.

CONCEITOS

O CICLO DA ESTATÍSTICA

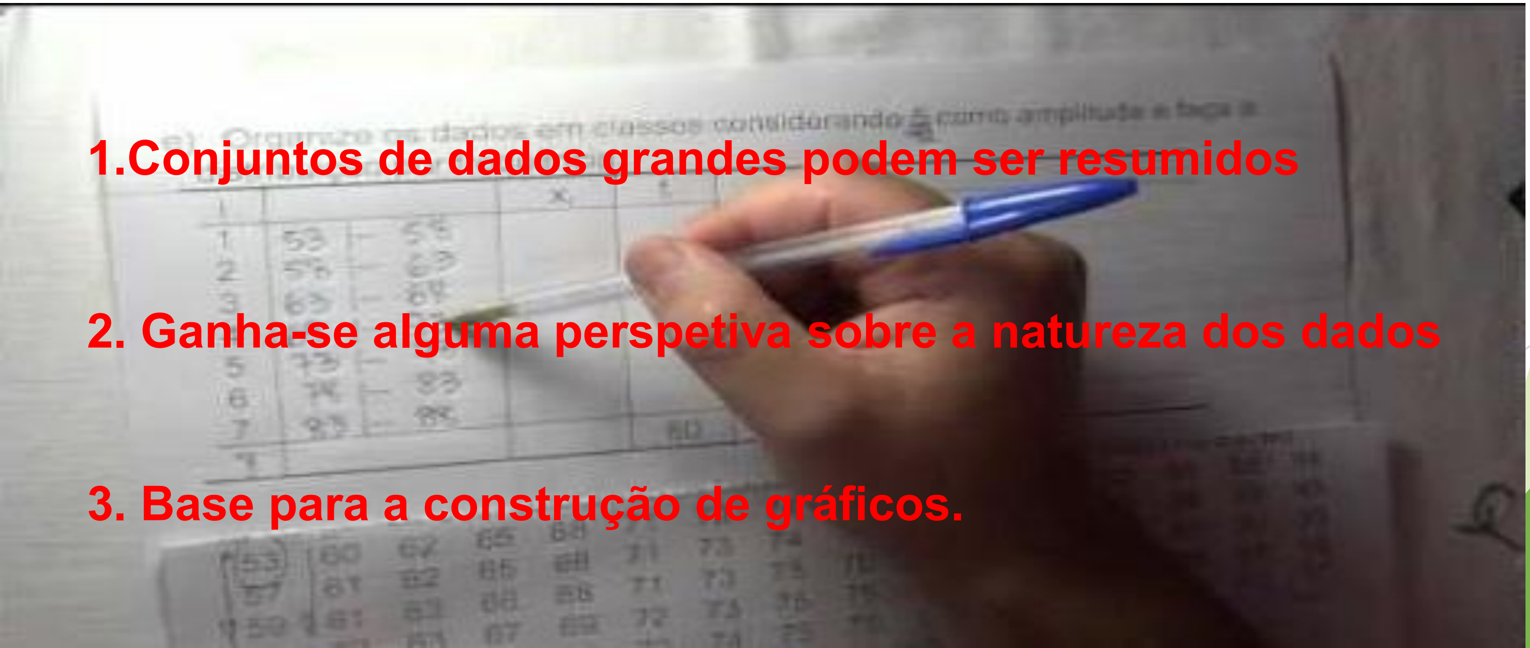


DISTRIBUIÇÃO DE FREQUÊNCIAS

O que é uma distribuição de frequências?

Lista ou rol dos valores dos dados (individuais ou por grupos de intervalos), juntamente com as correspondentes frequências ou contagens

Razões para construir distribuições de frequência



1.Conjuntos de dados grandes podem ser resumidos

2. Ganha-se alguma perspetiva sobre a natureza dos dados

3. Base para a construção de gráficos.

DISTRIBUIÇÃO DE FREQUÊNCIAS

Tabela de frequências

Uma das formas comuns de organizar os dados é através das tabelas de frequências, onde figuram entre outros valores, aqueles que dizem respeito às frequências absolutas simples frequências relativas simples, frequências absolutas acumuladas e frequências relativas acumuladas.

EX. 3

Relativamente a 50 motoristas que trabalham numa empresa de transportes, observou-se a quantidade de acidentes sofridos em serviço, tendo-se registado os seguintes dados:

3	1	0	3	1	1	2	1	0	1
0	0	2	1	0	2	2	0	1	2
1	0	1	2	2	0	1	0	2	3
0	3	0	1	0	2	0	0	1	2
0	2	1	0	0	1	0	1	0	0

Construa uma tabela de distribuição de frequências

DISTRIBUIÇÃO DE FREQUÊNCIAS

Frequência ABSOLUTA simples

Corresponde ao número de vezes que um determinado valor x_i é observado na população ou amostra estudada, e designa-se por n_i)

Frequência RELATIVA simples

Sendo ni a frequência absoluta do valor da variável x_i e N a dimensão da população ou amostra estudada, designa-se por frequência relativa simples de x_i , o quociente:

$$f_i = \frac{n_i}{N}$$

É comum a utilização das frequências relativas sob a forma de percentagem, que se obtêm calculando $fi \times 100$.

DISTRIBUIÇÃO DE FREQUÊNCIAS

Frequência ABSOLUTA acumulada

Designa-se por frequência absoluta acumulada de índice i a soma das frequências absolutas correspondentes aos valores da variável desde o primeiro valor até ao de ordem i, e representa-se por Ni.

Para k observações tem-se

$$N_1 = n_1$$

$$N_2 = n_1 + n_2 = N_1 + n_2$$

$$N_3 = n_1 + n_2 + n_3 = N_2 + n_3$$

$$N_k = n_1 + n_2 + \ldots + n_k = N_k - 1 + n_k$$

$$N = \sum_{i=1}^{k} n_i$$

DISTRIBUIÇÃO DE FREQUÊNCIAS

Frequência RELATIVA acumulada

Corresponde à soma das frequências relativas correspondentes aos valores da variável desde o primeiro valor até ao de ordem i, e representa-se por F_i.

$$F_1 = f_1$$
$$F_2 = f_1 + f_2 = F_1 + f_2$$
$$F_3 = f_1 + f_2 + f_3 = F_2 + f_3$$
$$F_k = f_1 + f_2 + \ldots + f_k = F_k - 1 + f_k$$

$$F = \sum_{i=1}^{k} f_i = 1$$

Em termos genéricos

x_i	n_i	$f_i = \frac{n_i}{n}$	N_i	F_i
x_1	n_1	$f_1 = \frac{n_1}{n}$	n_1	f_1
x_2	n_2	$f_2 = \frac{n_2}{n}$	$n_1 + n_2$	$f_1 + f_2$
x_3	n_3	$f_3 = \frac{n_3}{n}$	$n_1 + n_2 + n_3$	$f_1 + f_2 + f_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	$f_k = \frac{n_k}{n}$	$n_1 + n_2 + n_3 + \ldots + n_k$	$f_1 + f_2 + f_3 + \ldots + f_k$
		Total	$N = \sum_{i=1}^{k} n_i$	$\sum_{i=1}^{k} f_i = 1$

DISTRIBUIÇÃO DE FREQUÊNCIAS

EX. 1 *(Continuação)*

Atributo alvo da observação realizadasobre o conjunto dos motoristas

Quantidade de Acidentes	Quantidade de Motoristas	Proporção de Motoristas	Quantidade Acumulada de Motoristas	Proporção Acumulada de Motoristas
0	20	0,40	20	0,40
1	15	0,30	35	0,70
2	11	0,22	46	0,92
3	4	0,08	50	1,00
TOTAL	50	1,00		

Frequências Relativas Acumuladas

Modalidades numéricas (valores) assumidas pelo atributo VariávelDISCRETA

Qtd. de repetições de cada modalidade do atributo: frequências absolutas simples

Proporção de repetições de cada modalidade do atributo: frequências relativas simples

Frequências Absolutas Acumuladas

DISTRIBUIÇÃO DE FREQUÊNCIAS

Dados agrupados em classes

Como as variáveis contínuas podem assumir um número infinito de valores, torna-se necessário definir classes de valores

Exemplos: escalões do IRS; rendimento anual

Para se definir as classes tem que se ter em consideração:

- **Número de classes**
- **Amplitude**
- **Limites**
- **Ponto médio ou centro**

401-500
501-600
601-700

DISTRIBUIÇÃO DE FREQUÊNCIAS

QUESTÃO PARA DISCUSSÃO

“Se houver muitos intervalos, o resumo não constituirá grande melhoria com relação aos dados brutos. Se houver poucos, um grande volume de informação se perderá”

Pagano, 2004

DISTRIBUIÇÃO DE FREQUÊNCIAS

Regras na definição das classes

1. O número de classes (k) deverá estar compreendido entre quatro e catorze
2. Nenhuma classe deverá ter uma frequência nula
3. As classes deverão ter, sempre que possível, amplitudes iguais
4. Os pontos médios das classes deverão ser números e cálculo fácil
5. Classes abertas deverão ser evitadas
6. Os limites das classes são definidos de modo a que cada valor pertença a um único intervalo

DISTRIBUIÇÃO DE FREQUÊNCIAS

Como construir uma tabela de frequências

PASSO 1: Decidir o número de classes (k) seguindo uma das seguintes soluções

a) k = 5 para n < 25 e k igual ao valor arredondado da raiz quadrada de n quando n >= 25

b) Regra de Sturges: k =nº de classes ≈ 1+ 3,322 log n

c) Milone: k= - 1 + 2 ln n

PASSO 2: Calcular (aproximando por excesso) a amplitude da classe (h) ≈ [(maior valor) – (menor valor)] / número de classes

DISTRIBUIÇÃO DE FREQUÊNCIAS

Como construir uma tabela de frequências (continuação)

PASSO 3: Escolher o limite inferior da primeira classe**.**

PASSO 4: Usar o limite inferior da primeira classe e a amplitude de classe para listar, numa coluna vertical, todos os limites inferiores.

PASSO 5: Em seguida, listar os correspondentes limites superiores.

PASSO 6: Percorrer os dados, assinalando com um traço vertical a classe onde se encontra cada dado

DISTRIBUIÇÃO DE FREQUÊNCIAS

INSTITUTO SUPERIOR DE TECNOLOGIAS AVANÇADAS

PR. 2

Considere a variável "**idade**" (em nº de anos) dos alunos da turma de Estatística Descritiva

1.Como se classifica esta variável?

2.Construa uma tabela de frequências simples e sem perda de informação.

3.Determine as classes de distribuição, analisando as várias hipóteses à luz dos critérios estudados

4.Construa uma distribuição de frequências com intervalos de classes iguais

5.Construa uma distribuição de frequências com intervalos de classes diferentes

6.Admita, agora, que pretende estudar a "**altura**" dessa população. Responda novamente a todas as questões colocadas

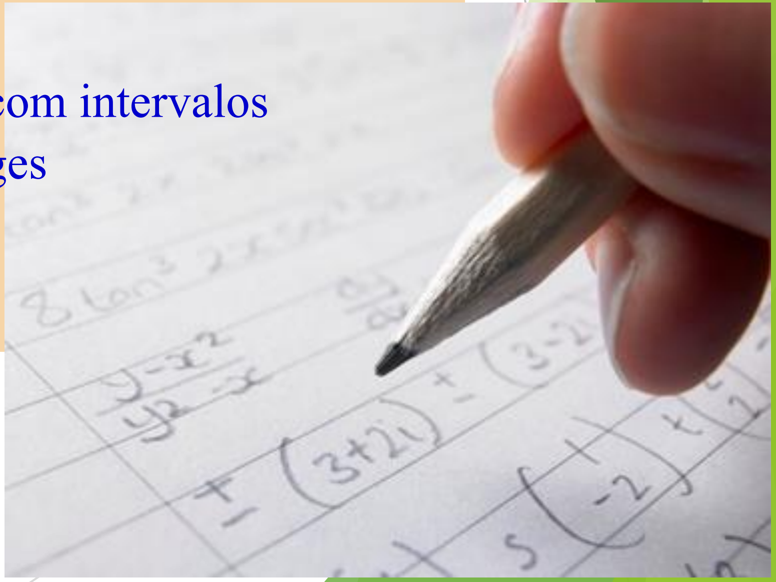
DISTRIBUIÇÃO DE FREQUÊNCIAS

PR. 3

Considere a amostra constituída pelas notas obtidas num ponto de Matemática, de uma determinada turma:

12.1	8.9	16.2	8.2	9.8	15.1
14.5	13.4	14.7	7.5	8.8	12.4
16.1	15.2	13.5	13.8	14.6	15.5
7.8	12.5	13.2	11.0	10.5	

Construa uma distribuição de frequências com intervalos de classes iguais, seguindo a regra de Sturges

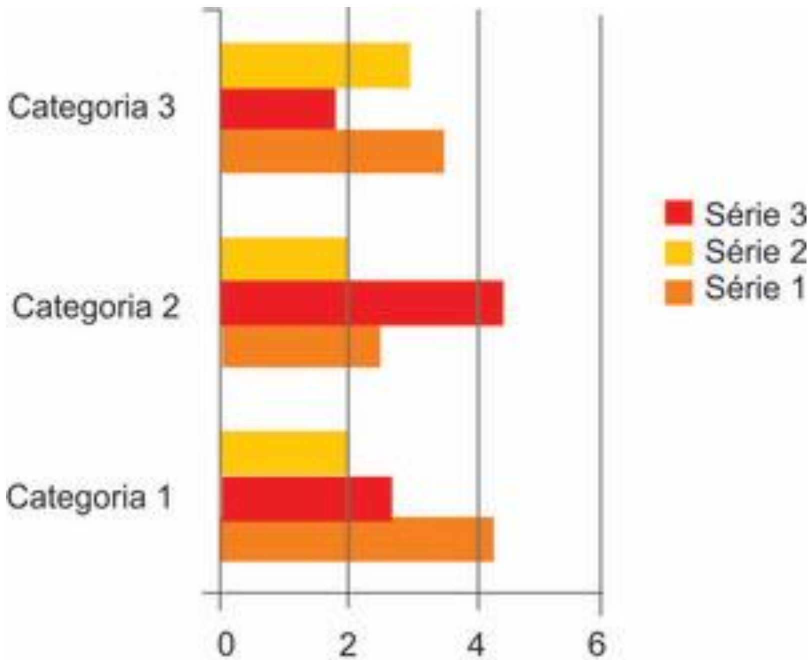


APRESENTAÇÃO DOS DADOS

✓**Tabelas**

✓**Gráficos**

Categoria	Frequência Absoluta	Frequência relativa
Castanhos	10	0,50
Pretos	7	0,35
Azuis	2	0,10
Verdes	1	0,05
Total	20	1.00



(Gráfico de barra)

APRESENTAÇÃO DOS DADOS

Tabelas

Toda tabela deve ser simples, clara, objetiva e autoexplicativa

Tabela *rs* – Título respondendo as perguntas: o quê, onde e quando?

Coluna indicadora	Cabeçalho
Conteúdo da linha	Célula Coluna

Fonte: Origem dos dados.

Nota: Informação esclarecedora.

Corpo da tabela

Figura 01 – Representação tabular dos dados.

APRESENTAÇÃO DOS DADOS

Gráficos

Os gráficos devem primar pela simplicidade, clareza e veracidade.

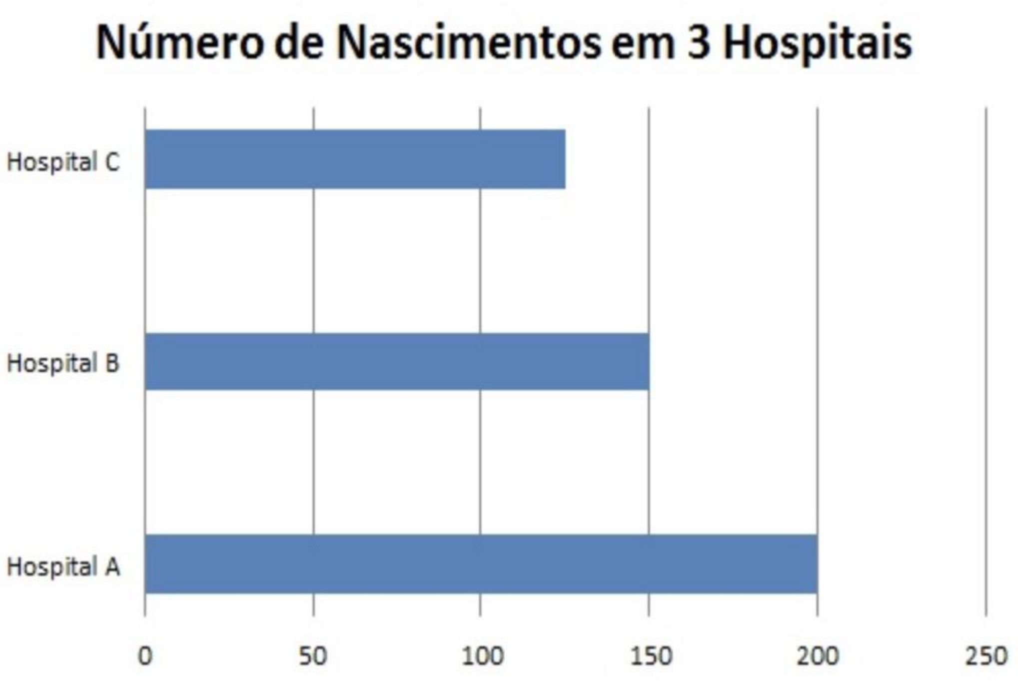
Antes de construir um GRÁFICO, não se esqueça de responder às seguintes questões:

- *Um gráfico realmente é a melhor opção?*
- *Qual é o público-alvo?*
- *Qual é o objetivo do gráfico?*
- *Que tipo de gráfico deve ser usado?*
- *Como o gráfico deve ser apresentado?*
- *Que tamanho o gráfico deve ter?*
- *Deverá ser usado apenas um gráfico?*
- *A qual meio técnico se deve recorrer?*

APRESENTAÇÃO DOS DADOS

Gráficos para variáveis qualitativas

GRÁFICO DE BARRAS



- É um gráfico formado por retângulos horizontais de larguras iguais, onde cada um deles representa a intensidade de uma modalidade ou atributo.

- É recomendável que cada coluna conserve uma distancia entre si de aproximadamente 2/3 da largura da base de cada barra, evidenciando deste modo, a não continuidade na sequência dos dados.

- O objetivo deste gráfico é de comparar grandezas e é recomendável para variáveis cujas categorias tenham designações extensas.

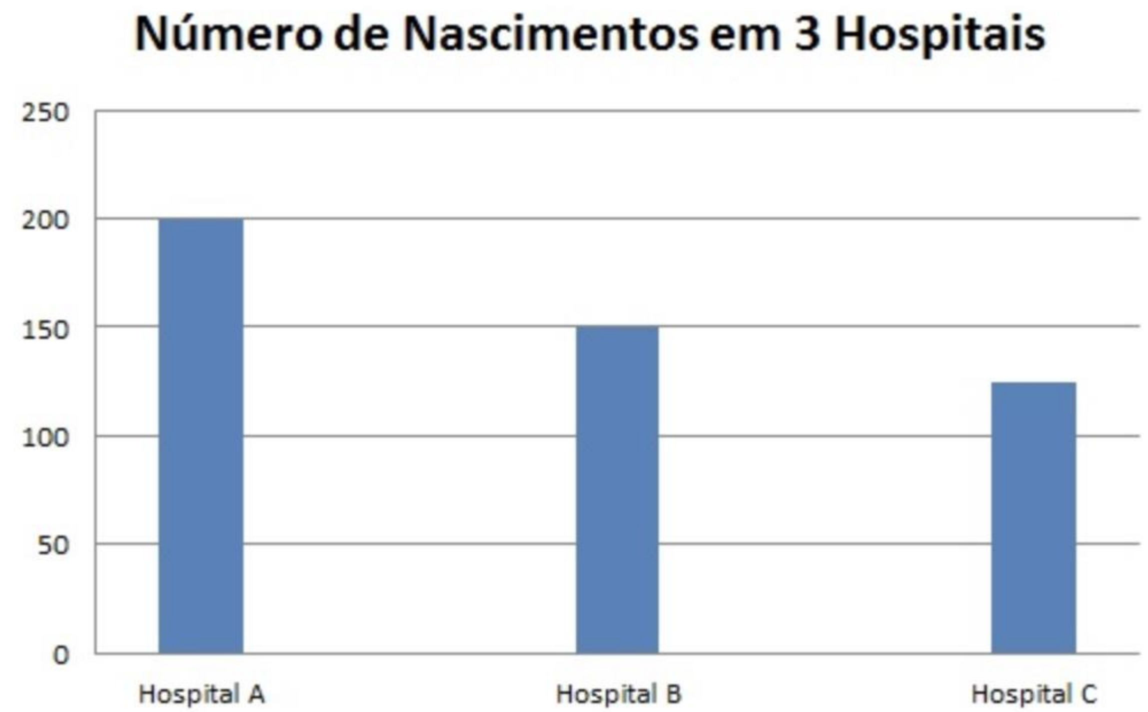
APRESENTAÇÃO DOS DADOS

Gráficos para variáveis qualitativas

GRÁFICO DE COLUNAS

- É o gráfico mais utilizado para representar variáveis qualitativas.

- Também para este tipo de gráfico deve ser preservada a distância entre cada retângulo de, aproximadamente, 2/3 da largura da base de cada coluna.



- O número de colunas ou barras do gráfico não deve ser superior a 12 (doze).

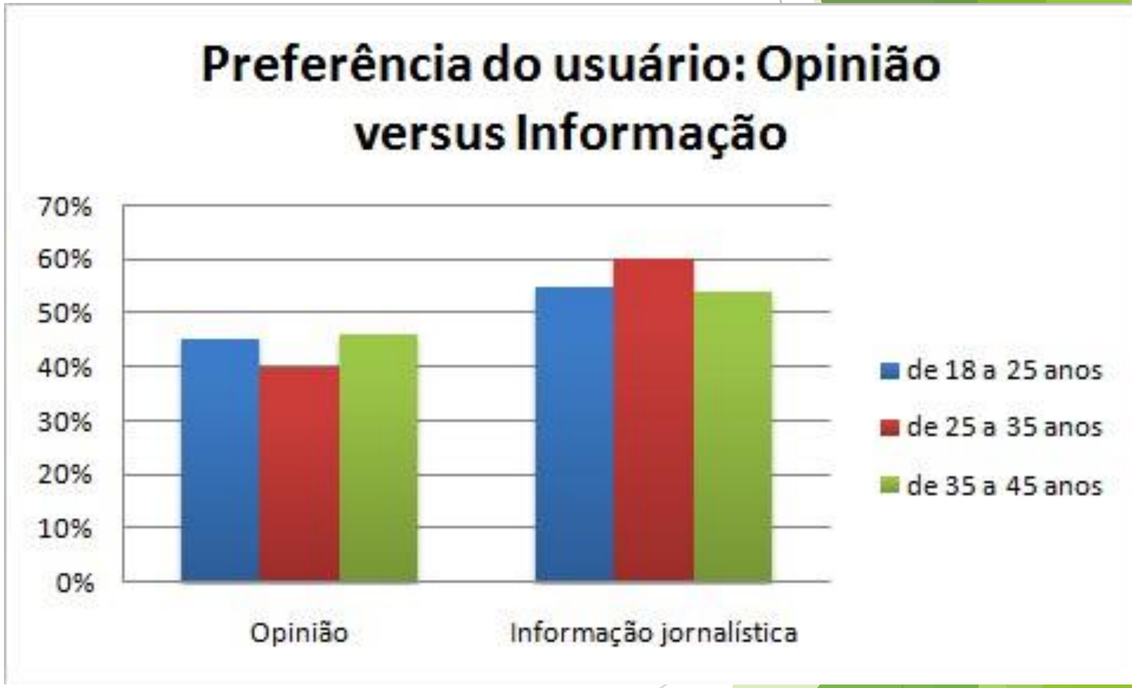
APRESENTAÇÃO DOS DADOS

Gráficos para variáveis qualitativas

GRÁFICO DE COLUNAS/ LINHAS JUSTAPOSTOS

- Ao se descrever simultaneamente duas ou mais categorias de uma variável, é conveniente fazer uso dos gráficos de barras ou colunas justapostas, chamados de gráficos **comparativos**.

- De acordo com as normas contidas em Gráficos, este tipo de gráfico só deve ser utilizado quando apresentar até três elementos para uma série de no máximo quatro valores.



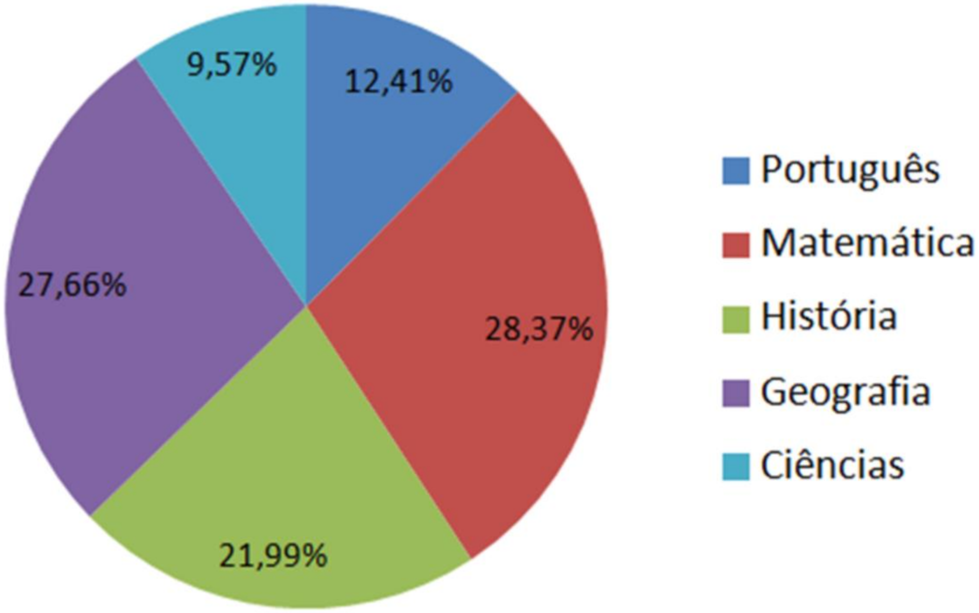
APRESENTAÇÃO DOS DADOS

Gráficos para variáveis qualitativas

GRÁFICO DE SETORES/CIRCULAR

- Tipo de gráfico onde a variável em estudo é projetada num circulo, de raio arbitrário, dividido em setores com áreas proporcionais às frequências das suas categorias.

- São indicados quando se deseja comparar cada valor da série com o total.



- Recomenda-se seu uso para o caso em que o número de categorias não é grande e não obedece a ordem especifica.

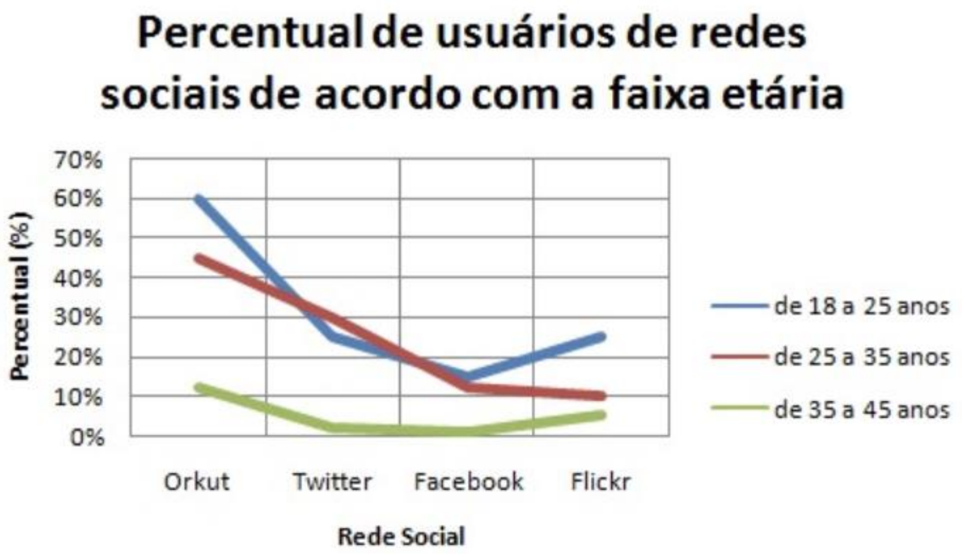
APRESENTAÇÃO DOS DADOS

Gráficos para variáveis qualitativas

GRÁFICO DE LINHAS

- É o mais indicado para representações de séries temporais sendo por tal razão, conhecidos também como gráficos de séries cronológicas.

- A sua construção é feita colocando-se no eixo vertical (y) a mensuração da variável em estudo e na abcissa (x), as unidades da variável numa ordem crescente.



- Este tipo de gráfico permite representar séries longas, o que auxilia detetar suas flutuações tanto quanto analisar tendências.

- Também podem ser representadas varias series em um mesmo gráfico.

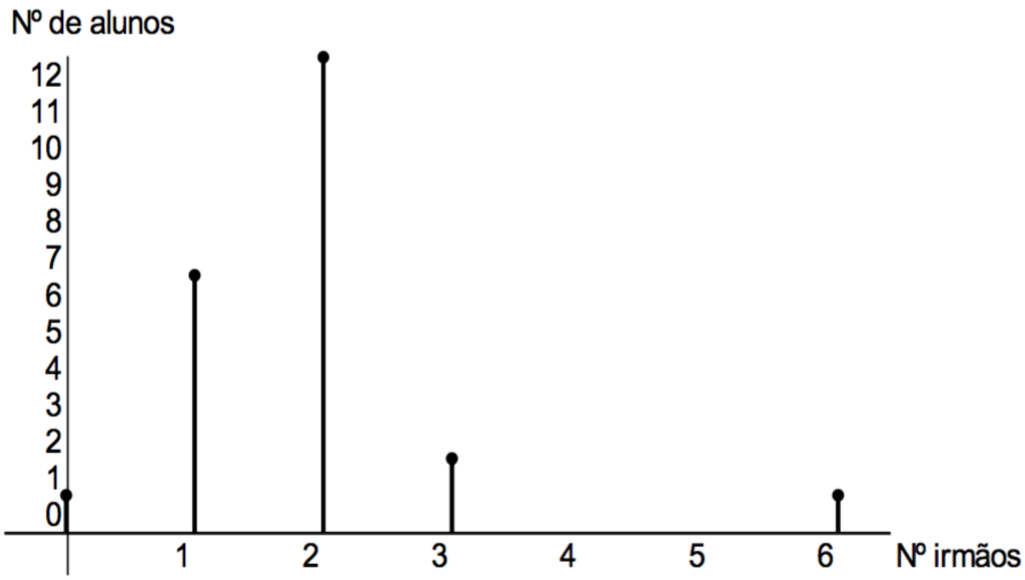
APRESENTAÇÃO DOS DADOS

Gráficos para variáveis quantitativas DISCRETAS

DIAGRAMA DIFERENCIAL

- É formado por segmentos de retas perpendiculares ao eixo horizontal cujo comprimento corresponde à frequência absoluta ou relativa de cada elemento da distribuição.

- As suas coordenadas não podem ser unidas porque a leitura do gráfico deve tornar claro que não há continuidade entre os valores individuais assumidos pela variável em estudo.



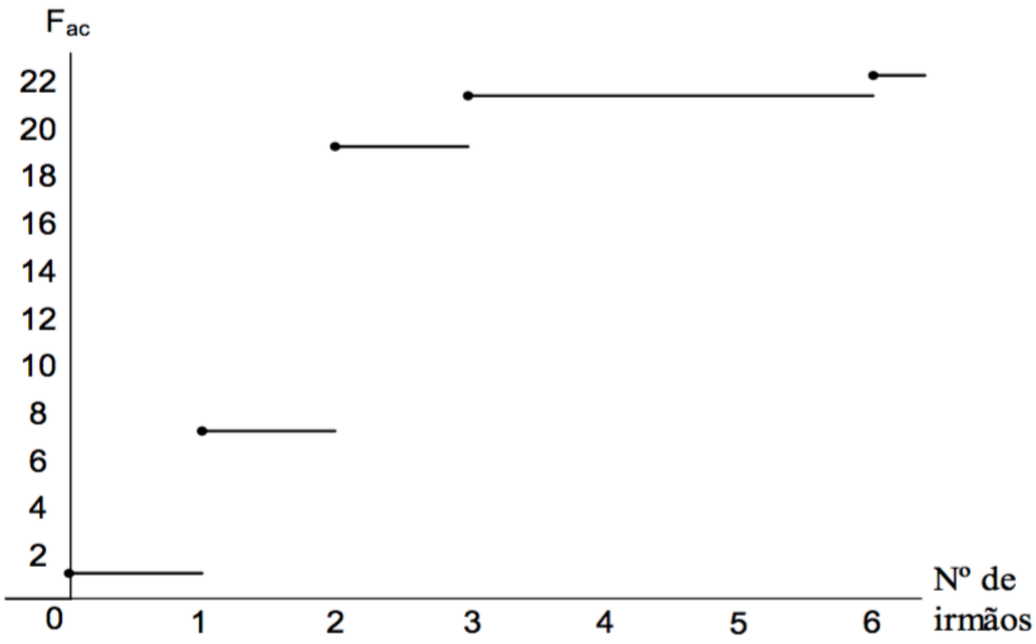
APRESENTAÇÃO DOS DADOS

Gráficos para variáveis quantitativas DISCRETAS

DIAGRAMA INTEGRAL

- No eixo das abcissas são alocados os valores assumidos pela variável "número de irmãos" e no eixo das ordenadas as suas frequências acumuladas.

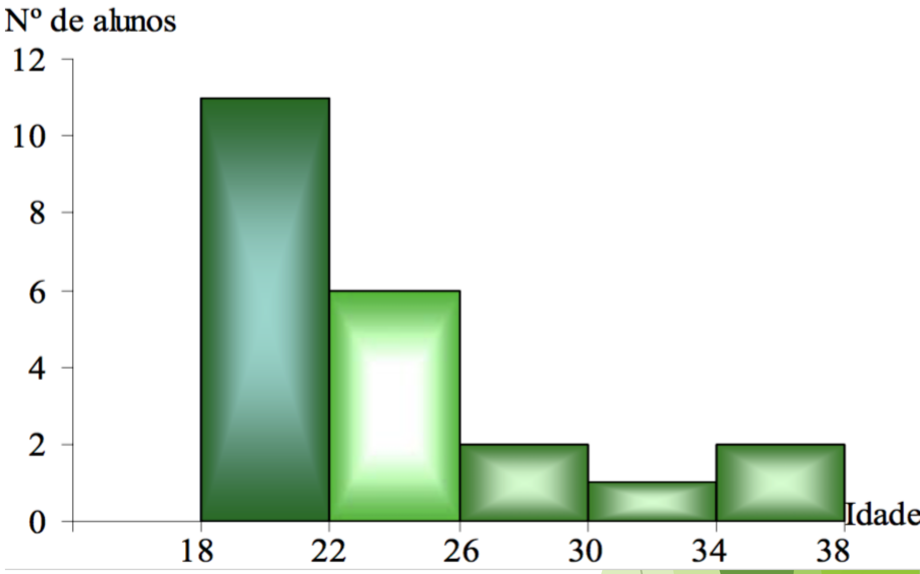
- Observa-se que a leitura do gráfico exige alguns cuidados básicos: caso o valor da variável esteja ou não incluído, a sua frequência acumulada difere.



APRESENTAÇÃO DOS DADOS

Gráficos para variáveis quantitativas CONTÍNUAS

HISTOGRAMA



- É um gráfico de colunas justapostas que representa uma distribuição de frequência para dados contínuos ou uma variável discreta quando esta apresentar muitos valores distintos.

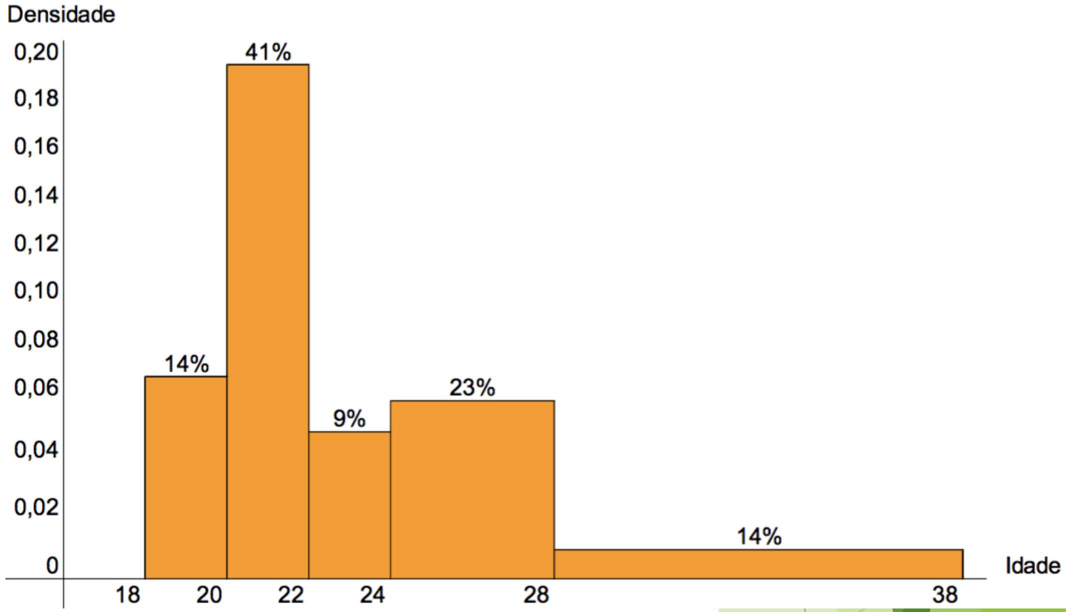
- No eixo horizontal são dispostos os limites das classes segundo as quais os dados foram agrupados enquanto que o eixo vertical corresponde às frequências absolutas ou relativas das mesmas.

- Quando os dados são distribuídos em classes de mesma amplitude, todas as colunas apresentam bases iguais com alturas variando em função das suas frequências absolutas ou relativas. Neste caso, tem-se que a área de cada retângulo depende apenas da sua altura

APRESENTAÇÃO DOS DADOS

Gráficos para variáveis quantitativas CONTÍNUAS

HISTOGRAMA (continuação)



- no caso de dados agrupados em classes de dimensões diferentes, a área de cada coluna já não é proporcional à sua altura.

- é necessário comparar as áreas para interpretar as informações que são expostas

APRESENTAÇÃO DOS DADOS

PR. 4

4. Considere a amostra obtida a partir das notas do exame de estatística descritiva dos alunos do 1º ano do ISTEC

12,1	14,5	16,1	7,8
8,9	13,4	15,2	12,5
16,2	14,7	13,5	13,2
8,2	7,5	13,8	11
9,8	8,8	14,6	10,5
15,1	12,4	15,5	

a) Construa em MS Excel a distribuição de frequências com intervalos iguais através da regra de Sturges (via tabela dinâmica)

b) Desenhe o histograma desta distribuição

c) Utilize a função Histograma do MS Excel para construir a distribuição e o respetivo gráfico

APRESENTAÇÃO DOS DADOS

PR. 5

3. A um conjunto de 25 alunos de uma escola perguntou-se quantos irmãos tinham, tendo-se obtido os seguintes valores

1	3	2	0	0
2	2	2	2	3
1	1	3	3	4
0	0	1	1	1
3	1	1	2	6

a) Construa uma tabela dinâmica em que constem as frequências absoluta e relativa desta distribuição

b) Represente a distribuição através de um diagrama de barras

c) Desenhe o polígono de frequências sobreposto ao diagrama de barras

APRESENTAÇÃO DOS DADOS

PR. 6

1. Admita uma empresa onde trabalham 100 funcionários. Resolva no MS Excel as seguintes questões:

a) Construa uma lista composta por 100 valores aleatórios de modo a representarem a idade desses funcionários (idade mínima: 18; idade máxima 65)

b) Defina, igualmente e de modo aleatório, o género de cada funcionário

c) Com base nos valores gerados nas alíneas a) e b), construa uma distribuição de frequências das idades dos funcionários com intervalos de classes iguais através da regra de Sturges,

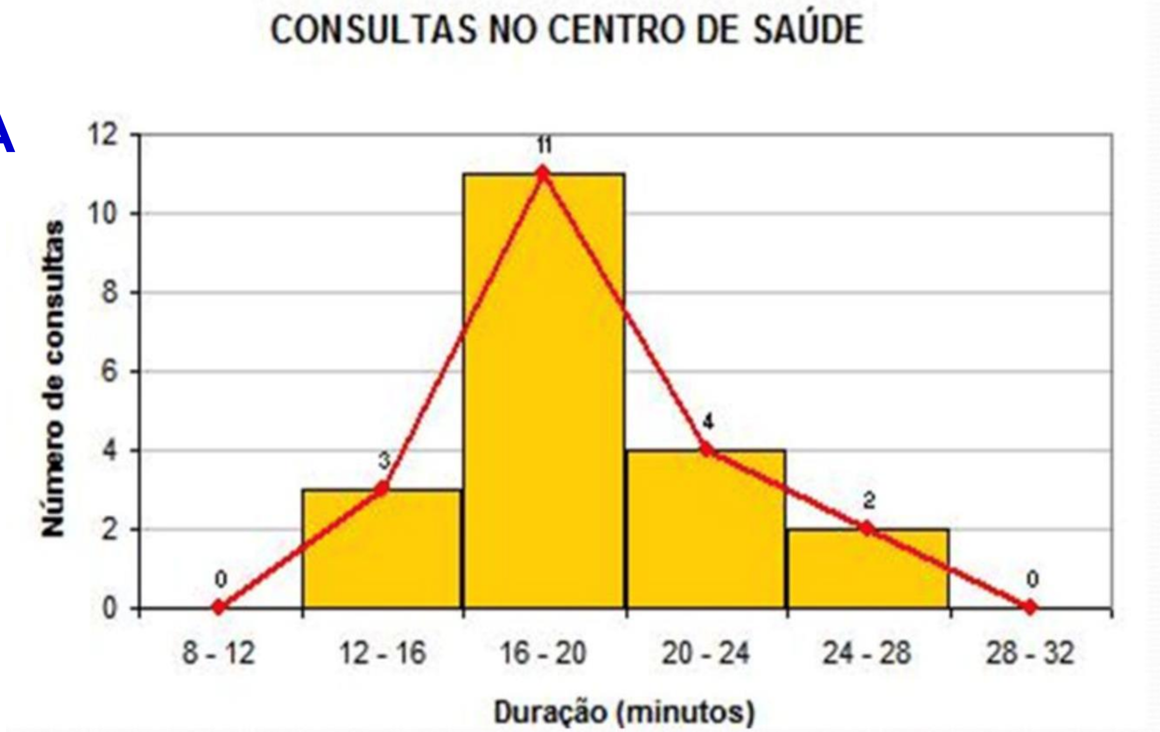
d) Represente a distribuição graficamente

e) Construa um diagrama circular que represente a distribuição dos funcionários por género

2. Represente graficamente a distribuição da PRÁTICA 3 (diapositivo 35).Justifique a sua escolha.

APRESENTAÇÃO DOS DADOS

Gráficos para variáveis <u>quantitativas CONTÍNUAS</u>



POLÍGNO DE FREQUÊNCIA

- é feito unindo-se os pontos de coordenadas de abscissas correspondentes aos pontos médios de cada classe e as ordenadas, às frequências absolutas ou relativas dessas mesmas classes.

- O polígono de frequência é um gráfico que deve ser fechado no eixo das abscissas.

- Por serem gráficos de linhas, permitem a comparação entre dois ou mais conjuntos de dados por meio da superposição dos mesmos.

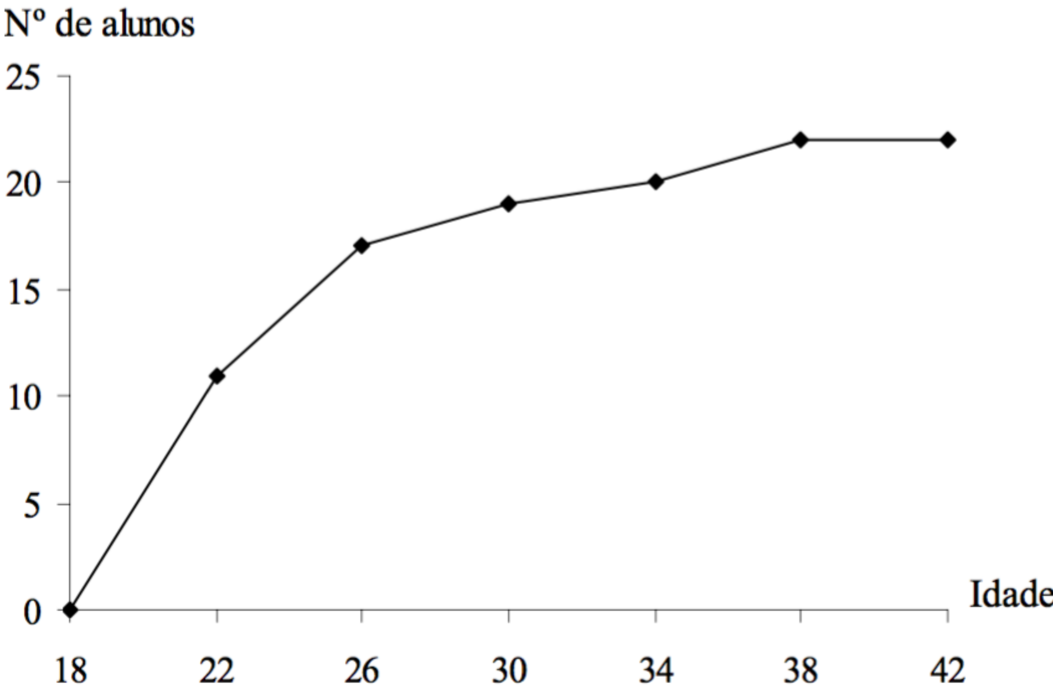
APRESENTAÇÃO DOS DADOS

Gráficos para variáveis quantitativas CONTÍNUAS

GRÁFICO OGIVA

- A ogiva é um gráfico de linha que une os pontos cujas abcissas são os limites superiores das classes, e, as ordenadas as suas respetivas frequências acumuladas.

- O ponto inicial do gráfico é o limite inferior do primeiro intervalo, com frequência acumulada zero, pois não existe qualquer valor inferior



APRESENTAÇÃO DOS DADOS

Gráficos para variáveis quantitativas CONTÍNUAS

CAULE-E-FOLHAS

35, 78, 50, 63, 86, 73, 57, 82, 59, 75, 66, 79, 83, 71, 94, 59

- O diagrama Caule-e-Folhas, criado por John Tukey, é um procedimento utilizado para armazenar os dados sem perda de informação

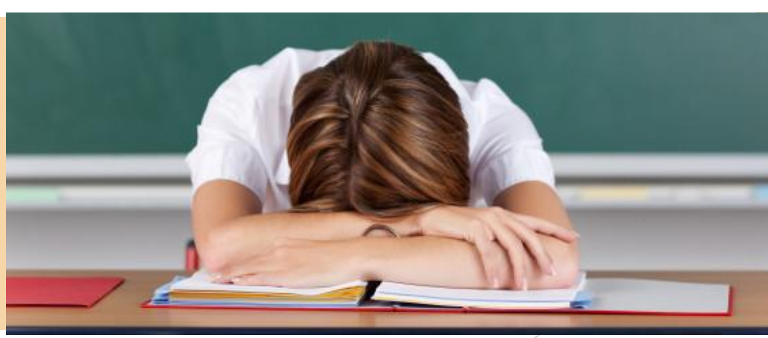
3	5
5	0 7 9 9
6	3 6
7	1 3 5 8 9
8	2 3 6
9	4

APRESENTAÇÃO DOS DADOS

PR. 7

A seguir apresentam-se os tempos de sono, medidos durante 30 noites seguidas, de dois jovens. Compare-os através de um caule-e-folhas

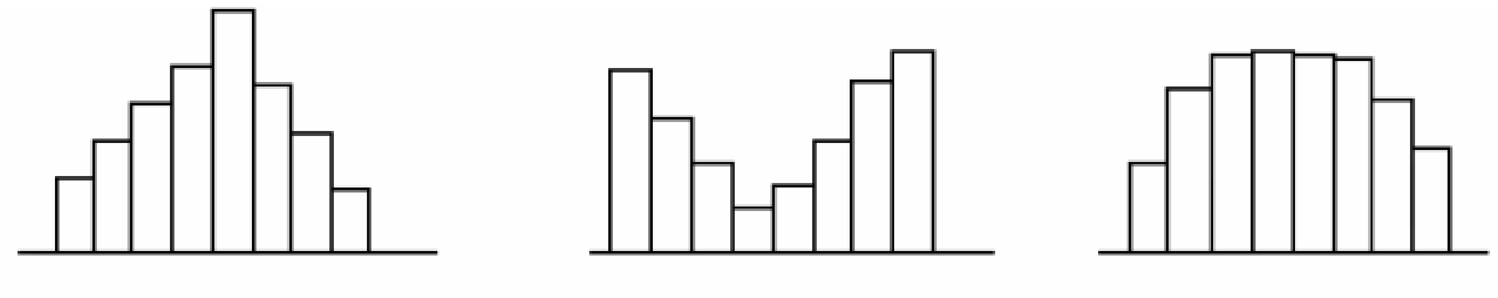
Pedro			David		
8.7	9.3	8.7	7.1	9.5	7.1
9.4	5.3	7.4	8.3	7.1	7.4
6.6	7.3	6.3	7.1	7.5	7.4
6.0	6.7	5.9	7.9	7.9	7.8
6.9	5.8	10.0	7.5	6.4	6.2
9.9	4.7	6.5	6.2	6.2	8.6
6.3	5.6	8.6	8.2	7.5	8.4
8.9	5.9	7.7	8.7	7.7	6.6
10.1	9.4	9.0	8.5	7.6	8.1
9.6	7.6	7.9	7.6	8.8	7.1



APRESENTAÇÃO DOS DADOS

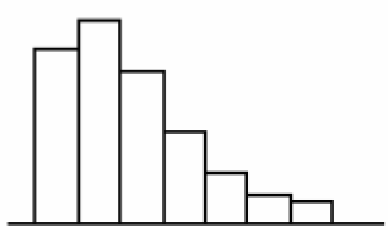
DISTRIBUIÇÕES SIMÉTRICAS

A distribuição das frequências faz-se de forma aproximadamente simétrica, relativamente a uma classe média:

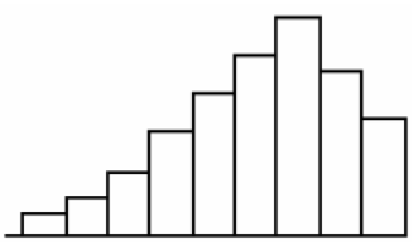


DISTRIBUIÇÕES ENVIESADAS

A distribuição das frequências faz-se de forma acentuadamente assimétrica, apresentando valores substancialmente mais pequenos num dos lados, relativamente ao outro



Enviesada para a direita

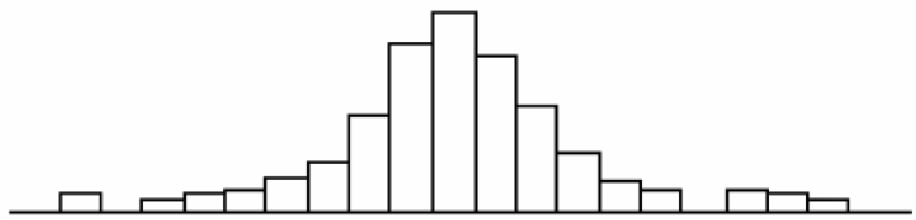


Enviesada para a esquerda

APRESENTAÇÃO DOS DADOS

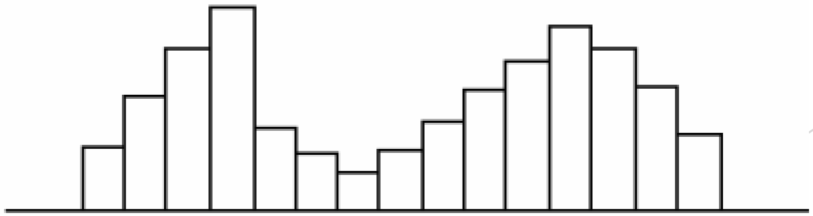
Distribuições com caudas longas

A distribuição das frequências faz-se de tal forma que existe um grande número de classes nos extremos, cujas frequências são pequenas, relativamente às classes centrais



Distribuições com vários "picos" ou modas

A distribuição das frequências apresenta 2 ou mais "picos" a que chamamos modas, sugerindo que os dados são constituídos por vários grupos distintos



APRESENTAÇÃO DOS DADOS

FUNÇÃO DE DISTRIBUIÇÃO EMPIRICA

É uma função definida para todo o número x real e que para cada x dá a proporção de elementos da amostra menores ou iguais a x.

Como se constrói?

1ºPASSO

Ordenar os n elementos da amostra, por ordem crescente

2º PASSO

Considerar um sistema de eixos coordenados e marcar no eixo do xx os valores da Amostra

3º PASSO

Começar a desenhar a função da esq. para a dir., atribuindo o valor 0 à esq do mínimo, o valor 1/n entre o mínimo e o 2º mínimo, o valor 2/n entre o 2º e o 3º mínimo, assim sucessivamente até se esgotar todos os valores da amostra

APRESENTAÇÃO DOS DADOS

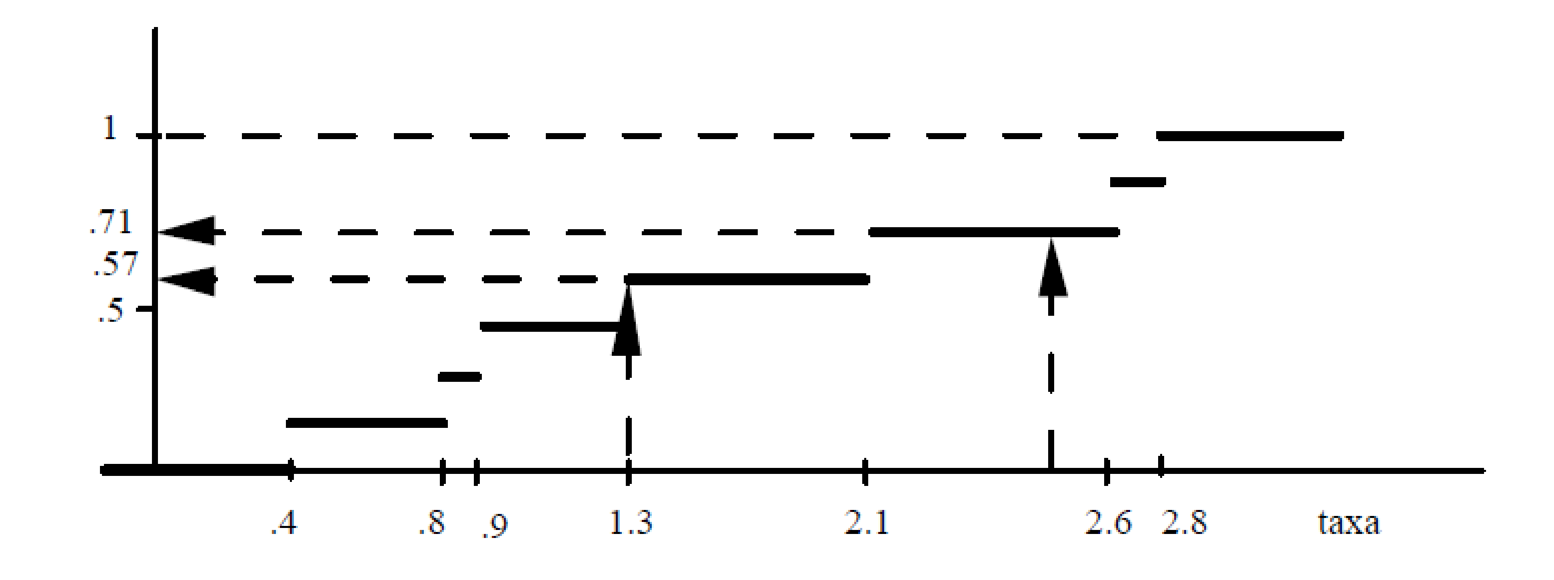
EX. 4

Construa o gráfico de uma função distribuição empírica para os seguintes valores, que representam a taxa de crescimento populacional, nas seguintes regiões

Região	Taxa de crescimento
África	2,8
Ásia	2,1
América do Norte	0,8
América Latina	2,6
Países de Leste	0,9
Europa	0,4
Oceânia	1,3

APRESENTAÇÃO DOS DADOS

EX. 4



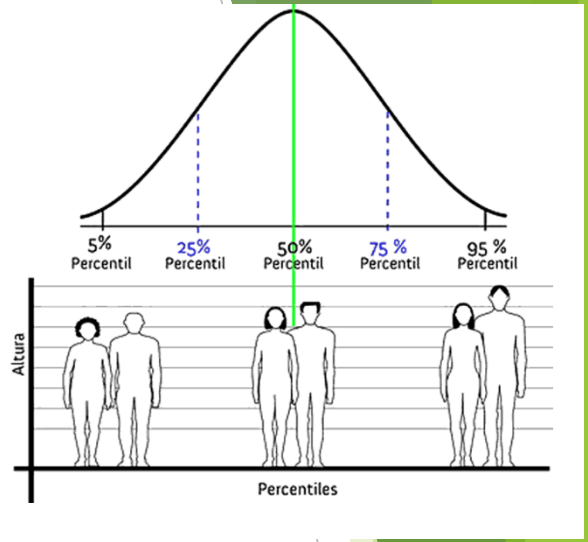
Qual a percentagem de taxas inferiores ou iguais a 1.3? **57%**

Qual a percentagem de taxas inferiores ou iguais a 2.5? **71%**

APRESENTAÇÃO DOS DADOS

PERCENTIS OU QUANTIS

Medidas que dão a localização dos valores de uma distribuição dividida em partes



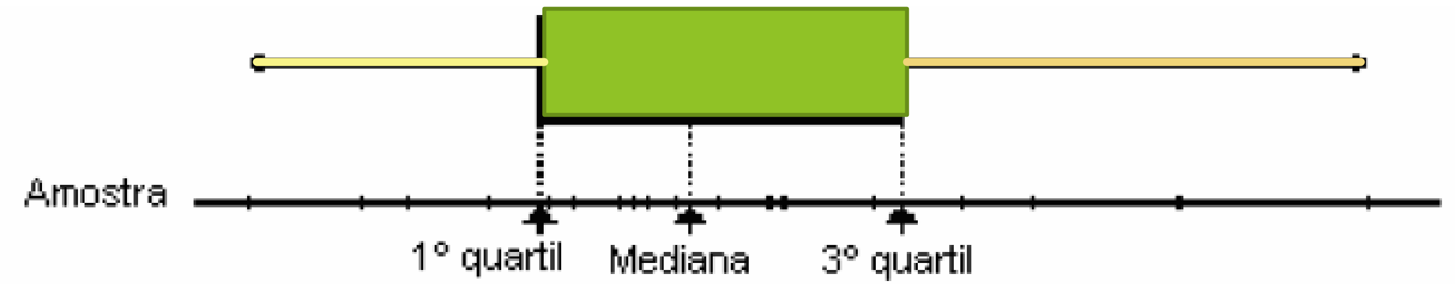
Existem alguns quantis que, pela sua importância, merecem uma referência especial:

Mediana- É o percentil correspondente à percentagem de 50%, o que significa que divide a amostra em duas partes com o mesmo número de elementos. Costuma-se representar por **m**.

Quartis - O **1º quartil** (ou quartil inferior) é o percentil, correspondente à percentagem de 25%, o que significa que 25% dos elementos da amostra são menores ou iguais a ele e os restantes são maiores ou iguais. O **3º quartil** (ou quartil superior) é o percentil correspondente à percentagem de 75%.

APRESENTAÇÃO DOS DADOS

"BOX-PLOT" OU "BOX-AND-WHISKER PLOT" (CAIXA-COM-BIGODES)



O conjunto dos valores da amostra compreendidos entre o 1º e o 3º *quartis*, Q.25 e Q.75 é representado por um retângulo (caixa) com a *mediana* indicada por uma barra.

Consideram-se seguidamente duas linhas que unem os meios dos lados dos retângulos com o menor e maior elementos da amostra que estão dentro das **barreiras**, definidas a seguir

APRESENTAÇÃO DOS DADOS

“BOX-PLOT” OU “BOX-AND-WHISKER PLOT” (CAIXA-COM-BIGODES)

O que são BARREIRAS?

Define-se **barreira inferior**, como sendo o valor

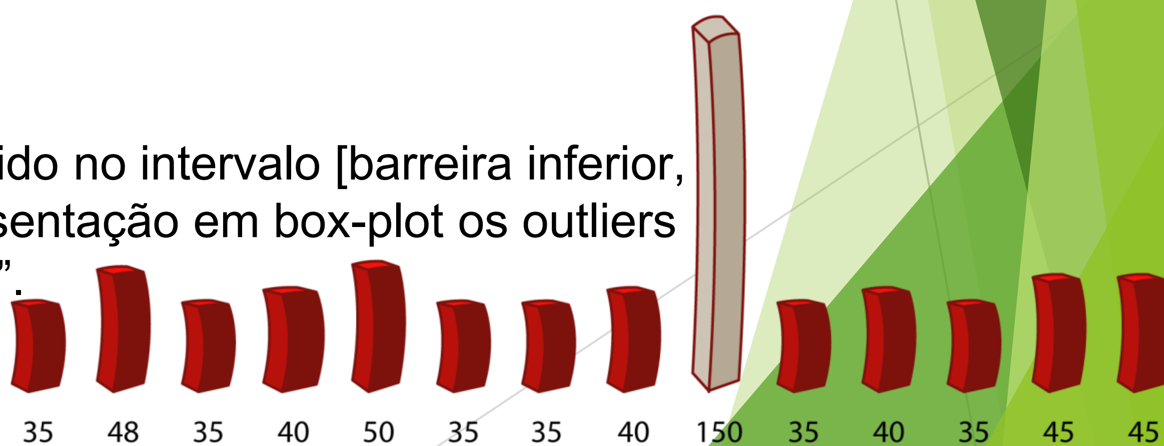
$$Q_{.25} - 1.5 \times (Q_{.75} - Q_{.25})$$

Define-se **barreira superior**, como sendo o valor

$$Q_{.75} + 1.5 \times (Q_{.75} - Q_{.25})$$

OUTLIERS

Valor que não está compreendido no intervalo [barreira inferior, barreira superior]. Numa representação em box-plot os outliers assinalam-se com o símbolo “*”.



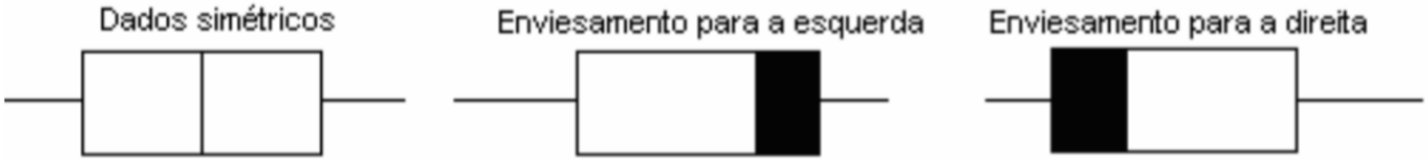
APRESENTAÇÃO DOS DADOS

“BOX-PLOT” OU “BOX-AND-WHISKER PLOT” (CAIXA-COM-BIGODES)

Como se pode reconhecer a simetria ou enviesamento dos dados?

- distância entre a linha indicadora da mediana e os lados do retângulo

- comprimento das linhas que saem dos lados dos retângulos



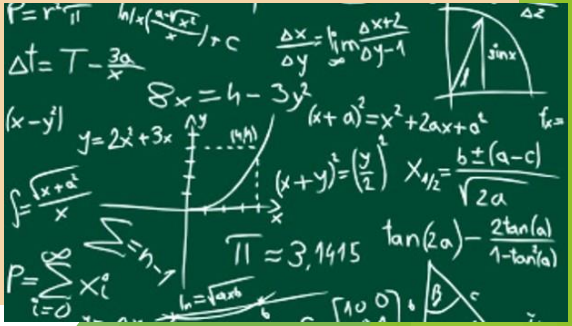
ISTEC INSTITUTO SUPERIOR DE TECNOLOGIAS AVANÇADAS

APRESENTAÇÃO DOS DADOS

PR. 8

Num inquérito à comunidade científica sobre a utilização de meios informáticos, realizado pela Fundação para o Desenvolvimento dos meios Nacionais de Cálculo Científico, obtiveram-se os seguintes resultados, quanto ao tipo de problemas tratados:

- Ajustamento de dados 337
- Eq. Diferenc. Ordinárias 54
- Análise de Fourier 195
- Gráfica Computacional 53
- Anál. Estatíst. de Dados 144
- Integração Numérica 38
- Desenv. de Software 116
- Inteligência Artificial 30
- Diferenças Finitas 96
- Interpolação 27
- Diferenciação Numérica 83
- Método Monte Carlo 19
- Elementos de Fronteira 75
- Métodos Numéricos 19
- Elementos Finitos 74
- Simulação 14
- Eq. Algébricas Lineares 70
- Valores e Vect. Próprios 11
- Eq. Algéb. não Lineares 59
- Outros 141



Represente em box-plot esta distribuição.

MEDIDAS DE ESTATÍSTICA DESCRITIVA

Dois alunos do 7º ano obtiveram as seguintes notas no 3º período:

Pedro ⟶ 4 3 3 3 3 3 4 3 4 3

João ⟶ 5 2 2 3 4 3 5 3 3 3

O Pedro e o João tiveram a mesma média de 3.3, mas o João não transitou de ano, pois teve duas negativas

A média não é suficiente para caracterizar e diferenciar os dois conjuntos de dados

3	3 3 3 3 3 3 3
4	4 4 4

2	2 2
3	3 3 3 3 3
4	4
5	5 5

MEDIDAS DE ESTATÍSTICA DESCRITIVA

MEDIDAS DE LOCALIZAÇÃO

- Medidas de tendência central (MÉDIA, MEDIANA, MODA
- Medidas de tendência não central (QUARTIS, DECIS, PERCENTIS)

MEDIDAS DE DISPERSÃO

- Amplitude de intervalos
- Desvio absoluto médio
- Variância e desvio-padrão
- Coeficiente de variação

MEDIDAS DE CONCENTRAÇÃO

- Curva de Lorenz
- Índice de Gini

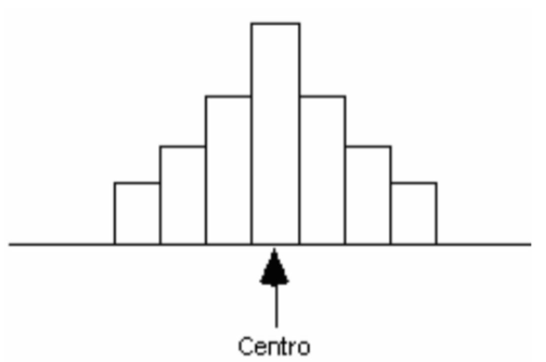
MEDIDAS DE ASSIMETRIA E CURTOSE

- Coeficientes de assimetria
- Achatamento e curtose

MEDIDAS DE LOCALIZAÇÃO

MEDIDAS DE TENDÊNCIA CENTRAL

São as que localizam o *centro de uma amostra ou população*

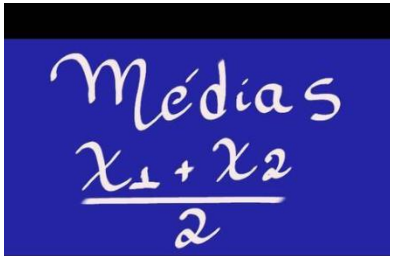


MÉDIA ARITMÉTICA

É a medida de localização mais vulgarmente utilizada. Representa-se por $\bar{x}$ quando se trate de uma estatística da amostra e por μ quando representa um parâmetro da população.

Calcula-se utilizando o seguinte processo:

- Somam-se todos os elementos da amostra;
- Divide-se o resultado da soma, pelo número de elementos da amostra



MEDIDAS DE LOCALIZAÇÃO

MÉDIA ARITMÉTICA (continuação)

VANTAGENS

- Fácil de calcular e de interpretar
- Utilizar toda a informação disponível e ter precisão matemática

INCONVENIENTES

- Ser influenciada por valores extremos
- Poder não corresponder a um valor em concreto da variável

MEDIDAS DE LOCALIZAÇÃO

MÉDIA ARITMÉTICA (continuação)

Xi - valores individuais observados
N - dimensão da população
n - dimensão da amostra

$$\mu = \frac{\sum_{i=1}^{N} Xi}{N}$$

Média da população (dados desagregados)

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

Média da amostra (dados desagregados)

PR. 9

O preço dos 6 modelos de telemóveis Apple à venda numa determinada loja constam do quadro seguinte: Calcule o preço médio dos vários modelos Apple.

Ref	Modelo	Preço (€)
1	**Apple iPhone 5s 16GB**	352,10
2	**Apple iPhone 6 16GB**	562,90
3	**Apple iPhone SE 16GB**	489,20
4	**Apple iPhone 6s 16GB**	640,90
5	**Apple iPhone 6s 64GB**	737,00
6	**Apple iPhone 6s Plus 16GB**	715,90

MEDIDAS DE LOCALIZAÇÃO

MÉDIA ARITMÉTICA (continuação)

PR. 10

Admita agora que as vendas por modelo são diferentes, seguindo as quantidades abaixo indicadas. Calcule agora o preço médio de cada telemóvel vendido.

Ref	Modelo	Quant.	Preço (€)
1	**Apple iPhone 5s 16GB**	3	352,10
2	**Apple iPhone 6 16GB**	5	562,90
3	**Apple iPhone SE 16GB**	2	489,20
4	**Apple iPhone 6s 16GB**	6	640,90
5	**Apple iPhone 6s 64GB**	3	737,00
6	**Apple iPhone 6s Plus 16GB**	2	715,90

$$\mu = \frac{\sum X_i \, . \, n_i}{N} = \sum X_i \, . \, f_i$$

Média da população (dados não agregados)

$$\bar{x} = \frac{\sum X_i \, . \, n_i}{n} = \sum X_i \, . \, f_i$$

Média da amostra (dados não agregados)

MEDIDAS DE LOCALIZAÇÃO

MÉDIA ARITMÉTICA (continuação)

PR. 11

Admita agora que as seguintes vendas de telemóveis por intervalo de preço. Calcule agora o preço médio de venda dos telemóveis.

Preço	Quant.
De 300 euros até 400 euros	12
De 400 euros até 500 euros	23
De 500 euros até 600 euros	8
De 600 euros até 700 euros	14
De 700 euros até 800 euros	5

$$\mu = \frac{\sum C_i \cdot n_i}{N} = \sum C_i \cdot f_i$$

Média da população (dados agregados em classes)

C_i – centro da classe

MEDIDAS DE LOCALIZAÇÃO

MÉDIA ARITMÉTICA (continuação)

Propriedades da média aritmética

-O somatório dos desvios dos valores da variável em relação à média é zero

$$\sum f_i \, (X_i - \mu) = 0$$

-O somatório do quadrado dos desvios dos valores da variável em relação é o mínimo

$$\sum f_i \, (X_i - \mu)^2 \text{ é } mínimo$$

MEDIDAS DE LOCALIZAÇÃO

MÉDIA GEOMÉTRICA

Utiliza-se em distribuições cujas variações são proporcionais a um valor original.

A média geométrica (M_g) de N números positivos é a raiz de Índice N do produto desses números

$$M_g = \sqrt[N]{X_1 . X_2 . X_3 \ldots X_N} = \sqrt[N]{\prod_{i=1}^{N} X_i}$$

Média geométrica (dados desagregados)

com Xi >0

Año	Nº nacimientos	Índice (Base 1994=100)
1940	20.287	100
1950	17.645	87,0
1960	19.164	94,5
1970	19.606	96,6
1980	18.204	89,7
1990	13.869	68,4

Se todos os valores de X forem iguais, a média geométrica é igual à média aritmética

MEDIDAS DE LOCALIZAÇÃO

MÉDIA GEOMÉTRICA (continuação)

VANTAGENS

- Utilizar toda a informação disponível e ter precisão matemática desde que a variável não tenha valores nulos ou negativos

- É o único tipo de média que serve para trabalhar com taxas de variação (ex. taxas de inflação)

- Não é condicionada por valores extremos como a média aritmética

INCONVENIENTES

- Complexidade de cálculo e de interpretação

- Poder não corresponder a um valor em concreto da variável

- Limitada a variáveis que só tenham valores positivos

AUXILIARES

LOGARITMO

Chama-se logaritmo de **x** na base **a**, a um número **b** tal que se elevarmos **a** ao expoente **b** obtemos **x**

Isto é $\log_a x = b \Leftrightarrow a^b = x$

EXEMPLOS

$$\log_2 8 = 3 \text{ pois } 2^3 = 8$$

b será portanto o logaritmo de x na base a o que significa que b é o expoente a que deve ser elevado a para obter x.

$\mathbf{\log_{10} 1000=3}$ pois $\mathbf{10^3 =1000}$

$\mathbf{\log_3 81 = 4}$ já que $\mathbf{3^4 = 81}$

AUXILIARES

Propriedades e regras dos LOGARITMOS

- $\log_a 1 = 0$
- $\log_a a = 1$
- $\log_a a^x = x$
- $a^{\log_a x} = x$
- $\log_a (u \cdot v) = \log_a u + \log_a v$
- $\log_a \left(\frac{u}{v} \right) = \log_a u - \log_a v$
- $\log_a (u^n) = n \cdot \log_a u$
- $\log_a \sqrt[n]{u} = \frac{1}{n} \log_a u$

ANTILOGARITMO

É o número que corresponde a um logaritmo dado. Consiste no problema inverso do cálculo do logaritmo de um número,

$$\log_a x = y \Leftrightarrow Anti\log_a y = x \Leftrightarrow a^y = x$$

ou seja, consiste em elevar a base ao número obtido no logaritmo

EXEMPLO

log 78=1.8921 logo antilog 1.892=78 <=>$10^{1.892} = 78$

MEDIDAS DE LOCALIZAÇÃO

MÉDIA GEOMÉTRICA (continuação)

Quando o n é maior que dois, o cálculo necessário para apurar a média geométrica fica simplificado através dos logaritmos

$$\text{Log Mg} = \text{Log}\left(\prod_{i=1}^{N} Xi\right)^{\frac{1}{N}} \Leftrightarrow \frac{1}{N}\log\left(\prod_{i=1}^{N} Xi\right) \Leftrightarrow \frac{1}{N}\sum_{i=1}^{N} \log Xi$$

$$\text{Mg} = \text{ant}\log\left[\frac{1}{N}\sum_{i=1}^{N} \log Xi\right]$$

Média geométrica (dados desagregados)

MEDIDAS DE LOCALIZAÇÃO

MÉDIA GEOMÉTRICA (continuação)

Para dados agregados a fórmula para o cálculo da média geométrica é a seguinte:

$$Mg = \text{ant}\log\left[\sum_{i=1}^{N} fi \log Xi\right]$$

Média geométrica (dados agregados)

MEDIDAS DE LOCALIZAÇÃO

MÉDIA GEOMÉTRICA (continuação)

PR. 12

Considere as seguintes taxas de crescimento populacional:

Ano	Mundo	Europa
1985	1,089	1,020
1990	1,090	1,022
1995	1,078	1,008
2000	1,070	1,001
2005	1,063	0,996

Calcule a taxa média de crescimento da população nos últimos 20 anos.

MEDIDAS DE LOCALIZAÇÃO

MÉDIA GEOMÉTRICA (continuação)

PR. 13

Suponha que contrata um empréstimo para habitação a 25 anos com spread variável.

Nos primeiros 15 anos, o spread é de 3%. Depois, reduz-se nos cinco anos seguintes para 2,5% e por fim, é fixado em 2%.

Qual a taxa média de juro a pagar pelo empréstimo admitindo que uma taxa Euribor igual a 1%

MEDIDAS DE LOCALIZAÇÃO

MÉDIA HARMÓNICA

Em situações onde a proporcionalidade inversa esteja presente é aconselhável o cálculo da Média Harmónica.

Por exemplo quando se estudam fenómenos como a velocidade média ou o custo médio de bens comprados com uma quantia fixa, deverá ser utilizada a média harmónica

$$\boldsymbol{Mh} = \frac{1}{\frac{1}{N}\sum_{\boldsymbol{i}=1}^{\boldsymbol{N}} \frac{\boldsymbol{Fi}}{\boldsymbol{xi}}} = \frac{N}{\sum_{i=1}^{N} Fi \frac{1}{X_{\boldsymbol{i}}}} = \frac{1}{\sum_{i=1}^{N} fi \frac{1}{Xi}}$$

Fi – Frequências absolutas simples
fi – Frequências relativas simples

A média harmónica é o inverso da média aritmética dos inversos dos valores observados

MEDIDAS DE LOCALIZAÇÃO

MÉDIA HARMÓNICA (continuação)

EX. 5

Um avião voa três distâncias iguais a uma velocidade de 300, 400 e 300 milhas por hora. Calcule a velocidade média utilizada para cobrir a distância total.

Neste caso aconselha-se a utilização da média harmónica uma vez que se trata de percorrer uma distância fixa a três velocidades diferentes.

$$Mh = \frac{3}{\frac{1}{300}+\frac{1}{400}+\frac{1}{300}} = 327{,}8 \text{ milhas por hora}$$

Se, pelo contrário, o avião voasse durante períodos fixos a velocidades diferentes, já seria correto utilizar a média aritmética.

MEDIDAS DE LOCALIZAÇÃO

MÉDIA HARMÓNICA (continuação)

PR. 14

Admita que tem um orçamento mensal de 400 euros para alojamento em hotéis de 4 estrelas, que durante os meses de julho, agosto e setembro registaram os preços diários de 40, 100 e 140 euros, respetivamente.

Qual o custo médio pago pelo alojamento diário nesse período?

MEDIDAS DE LOCALIZAÇÃO

MEDIANA É o valor central de um conjunto de dados

Principais características

1)Quando o número de elementos da série estatística for ímpar, haverá coincidência da mediana com um dos elementos da série

2)Quando o nº de elementos da série estatística for par, a mediana será sempre a média aritmética dos dois elementos centrais da série

3)Numa série, a mediana, a média, e a moda não têm necessariamente o mesmo valor

4)A mediana depende da posição e não dos valores dos elementos na série ordenada ou seja, a mediana é determinada pelo nº de observações e não pelo seu valor. Deste modo, os valores extremos, quer sejam grandes ou pequenos, não afetam o valor da mediana

5)É uma medida muito utilizada sobretudo quando se trata de distribuições fortemente assimétricas, pelo facto de não ser afetada por valores extremos

6)Para fins de inferência estatística, a mediana não satisfaz as propriedades de um bom estimador

MEDIDAS DE LOCALIZAÇÃO

MEDIANA

EX. 6

Considere o seguinte conjunto de notas de um aluno de Engenharia Informática

10 10 10 11 11 11 11 12

A média e a mediana deste conjunto de dados são, respetivamente,

$\bar{x} = 10.75$ e $m = 11$

Admitamos que uma das notas de 10 foi substituída por uma de 18. Então neste caso a mediana continuaria a ser 11, enquanto que a média subiria para 11.75 . **Como medida de localização, a mediana é mais resistente do que a media, pois não é tão sensível aos dados.**

Então qual destas medidas é preferível? Média ou mediana?

- Quando a distribuição é simétrica, a média e a mediana coincidem.
- A mediana não é tão sensível, como a média, às observações que são muito maiores ou muito menores do que as restantes (*outliers*). Por outro lado, a média reflete o valor de todas as observações.

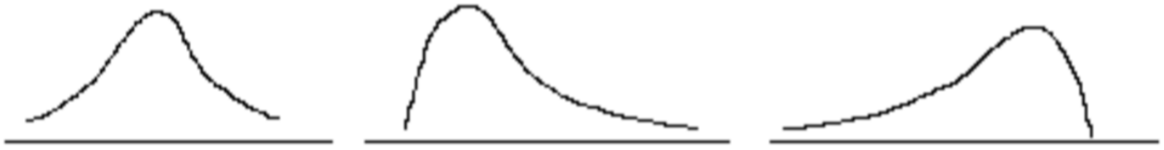
Assim, não se pode dizer, em termos absolutos, qual destas medidas é preferível, dependendo do contexto em que estão a ser utilizadas.

MEDIDAS DE LOCALIZAÇÃO

MEDIANA

Como a média é influenciada quer por valores muito grandes, quer por valores muito pequenos, conclui-se que:

-se a distribuição dos dados for enviesada para a direita (alguns valores grandes como *outliers*), a média tende a ser maior que a mediana;
-se for aproximadamente simétrica, a média aproxima-se da mediana
-e se for enviesada para a esquerda (alguns valores pequenos como *outliers*), a média tende a ser inferior à mediana.



média ≈ mediana   média > mediana   média < mediana

MEDIDAS DE LOCALIZAÇÃO

MEDIANA (continuação)

Como se obtém a mediana

Obtém-se ordenando os dados de forma crescente ou decrescente e identificando o valor central, no caso do número de dados ser ímpar, ou calculando a média entre os dois valores centrais, no caso do número de dados seja par.

Dados não agrupados com N ímpar (variáveis discretas)

$$\text{Mediana} = \frac{N+1}{2} \Leftrightarrow \text{Mediana} = \frac{\sum F_i + 1}{2}$$

Dados não agrupados com N par (variáveis discretas)

$$\text{Mediana} = \left[\frac{n}{2} + \left(\frac{n}{2} + 1\right)\right] / 2 \Leftrightarrow \text{Mediana} = \frac{\frac{N}{2} + \frac{N+2}{2}}{2}$$

MEDIDAS DE LOCALIZAÇÃO

MEDIANA (continuação)

PR. 15

Determinar a mediana para os pontos obtidos em 25 lançamentos de um dado.

Pontos X_i	Freq. Absoluta F_i	Freq. Absol. Acumulada F_A
1	4	4
2	3	7
3	4	11
4	6	17
5	5	22
6	3	25

MEDIDAS DE LOCALIZAÇÃO

MEDIANA (continuação)

Dados agrupados em classes (variáveis contínuas)

1) Determinar as frequências abolutas simples e acumuladas
2) Calcular a posição da mediana
3) Marcar a classe mediana
4) Calcular a mediana recorrendo à seguinte fórmula

Li – limite inferior da classe mediana

FAA – frequência acumulada da classe imediatamente anterior à classe mediana

Fi – frequência absoluta da classe mediana

h – amplitude do intervalo da classe mediana

$$Me = \mathrm{Li} + \frac{\left[\left(\frac{\mathrm{N}}{2} - \mathrm{FAA}\right) * \mathrm{h}\right]}{Fi}$$

PR. 16

Determine a mediana com base nos dados constantes na seguinte tabela de distribuição de frequências,

Classes	Fi	Fa
[50 a 54 [	4	4
[54 a 58 [	9	13
[58 a 62 [	11	24
[62 a 66 [	8	32
[66 a 70 [	5	37
[70 a 74 [	3	40
TOTAL	40	

MEDIDAS DE LOCALIZAÇÃO

MODA

É o valor que surge com mais frequência, se os dados são discretos, ou intervalo de classe com maior frequência, se os dados são contínuos

À semelhança da mediana, a **moda** também é uma medida de posição e pode ser determinada para dados quantitativos e qualitativos.

Considere os seguintes dados [7, 8, 9, 10, 10, 10, 11, 12]

Neste exemplo a moda é 10 já que é a variável mais frequente (com maior número de observações).

Por vezes, pode acontecer haver duas ou mais variáveis nestas condições. Nestes casos, estamos na presença de uma série **multimodal**

Ex: [7, 7, 8, 9, 10, 10, 11, 11, 12]

Pode ainda acontecer que numa determinada série não exista um valor modal, isto é, séries em que nenhuma variável apareça com maior frequência que as outras. Nestes casos a série é **amodal**

Ex: [7, 8, 9, 10, 11, 12]

MEDIDAS DE LOCALIZAÇÃO

MODA (continuação)

Moda de variáveis quantitativas agrupadas em classes (contínuas**)**

Para variáveis quantitativas agrupadas em classes (contínuas) torna-se necessário

1) Identificar a classe modal
2) Determinar o valor da moda dentro da classe

Para identificar a classe modal basta identificar qual a classe que apresenta maior frequência

O método mais simples para o cálculo da moda consiste em determinar o **ponto médio da classe modal (moda bruta)**

$$Mo = (Li + Ls) / 2$$

Li – limite inferior da classe modal

Ls – Limite superior da classe modal

MEDIDAS DE LOCALIZAÇÃO

MODA (continuação)

Moda de variáveis quantitativas agrupadas em classes (contínuas)

No método de **KING**, o cálculo da moda leva em conta a influência das classes adjacentes à classe modal, "deslocando" a moda em direção a aquelas

$$Mo = Li(Mo) + \left(\frac{F(Mo+1)}{F(Mo-1)+F(Mo+1)}\right) . h\ (Mo)$$

Li – limite inferior da classe modal
F (Mo + 1) - frequência absoluta da classe seguinte à modal
F (Mo - 1) - frequência absoluta da classe anterior à modal
h (Mo) - amplitude da classe modal

MEDIDAS DE LOCALIZAÇÃO

MODA (continuação)

Moda de variáveis quantitativas agrupadas em classes (contínuas)

Existe, no entanto, outro método mais elaborado e que nos permite calcular a moda com um maior grau de precisão. Este método consiste no recurso à fórmula de **CZUBER**

Emanuel Czuber
1851-1925

$$Mo = \mathrm{Li}\,(\mathrm{Mo}) + \frac{\mathrm{F}\,(\mathrm{Mo}) - \mathrm{F}(\mathrm{Mo}-1)}{2 \;\mathrm{x}\, \mathrm{F}\,(\mathrm{Mo}) - (\mathrm{F}\,(\mathrm{Mo}-1) + \mathrm{F}(\mathrm{Mo}+1))} \;\mathrm{x}\; \mathrm{h}\,(\mathrm{Mo})$$

Li – limite inferior da classe modal

F (Mo + 1) - frequência absoluta da classe seguinte à modal

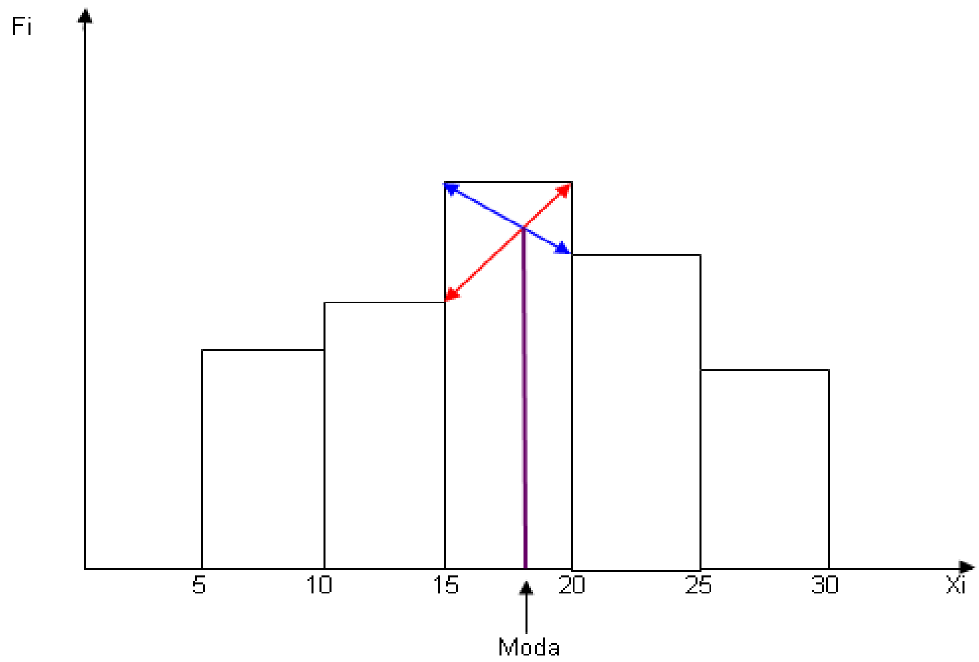
F (Mo - 1) - frequência absoluta da classe anterior à modal

h (Mo) - amplitude da classe modal

MEDIDAS DE LOCALIZAÇÃO

MODA (continuação)

- A moda pode ser determinada graficamente;



- A moda pode ser determinada em qualquer situação, mesmo quando a distribuição está definida em classes abertas;

- O valor da moda não sofre influência de valores extremos.

MEDIDAS DE LOCALIZAÇÃO

MODA (continuação)

PR. 17

Num exame de Estatística Descritiva composto por 50 questões e realizado por 220 alunos, registaram-se as seguintes respostas erradas:

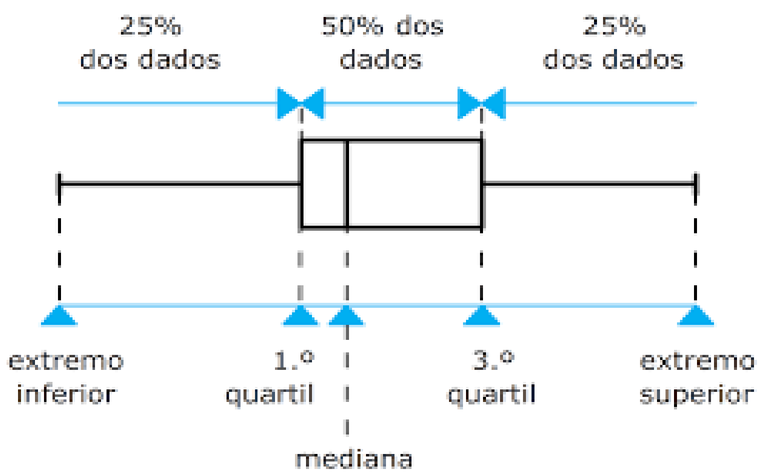
Questões erradas	Nº alunos
[0 a 10 [	30
[10 a 20 [	50
[20 a 30 [	70
[30 a 40 [	60
[40 a 50 [	10

Determine a média, a mediana, a moda bruta, a moda de King e a moda Czuber, histograma e polígono de frequências

MEDIDAS DE LOCALIZAÇÃO

MEDIDAS DE TENDÊNCIA NÃO CENTRAL

QUARTIS



Os quartis são os valores da variável observada que dividem a distribuição de frequências em 4 partes iguais.

Para determinar os quartis, tal como acontece na determinação da mediana, é necessário ordenar os dados.

Fórmulas para cálculo das posições de Q_1 e Q_3
(DADOS NÃO AGRUPADOS)

n par		n ímpar		
Q_1	Q_3	Q_1	Q_3	
$k = \frac{n+2}{4}$	$k = \frac{3n+2}{4}$	$k = \frac{n+3}{4}$	$k = \frac{3n+1}{4}$	Método Inclusivo
		$k = \frac{n+1}{4}$	$k = \frac{3n+3}{4}$	Método Exclusivo

Fórmulas para cálculo do valor de Qp dada a sua posição k

k inteiro	$Q_p = x_k$
k não inteiro $i < k < i+1$	$Q_p = \frac{x_i + x_{i+1}}{2}$

MEDIDAS DE LOCALIZAÇÃO

MEDIDAS DE TENDÊNCIA NÃO CENTRAL

QUARTIS (continuação) **para DADOS AGRUPADOS**

$$Qi = Li\,(Qi) + \frac{ni - FAA}{Fi} \times h\,(Qi)$$

Li – limite inferior da classe correspondente ao quartil i

ni - nº de observações acumuladas até ao quartil em causa (não é mais do que o valor obtido para a posição do quartil .

FAA – frequência acumulada da classe imediatamente anterior à classe do quartil

Fi – frequência absoluta da classe onde se situa o quartil

PR. 17

Os dados contidos na tabela respeitam às classificações obtidas pelos alunos da cadeira de Estatística Descritiva (4 turmas). Determine os quartis desta distribuição.

Classes	Fi	fi	Fa	fa
[0 a 4[	27	24,55%	27	24,5%
[4 a 8[	16	14,55%	43	39,1%
[8 a 12[	34	30,91%	77	70,0%
[12 a 16[	17	15,45%	94	85,5%
[16 a 20[	16	14,55%	110	100,0%
	110	**100,00%**		

ISTEC INSTITUTO SUPERIOR DE TECNOLOGIAS AVANÇADAS

MEDIDAS DE DISPERSÃO

MEDIDAS DE ESTATÍSTICA DESCRITIVA

Dois alunos do 7º ano obtiveram as seguintes notas no 3º período:

Pedro ⟶ 4 3 3 3 3 3 4 3 4 3

João ⟶ 5 2 2 3 4 3 5 3 3 3

O Pedro e o João tiveram a mesma média de 3.3, mas o João não transitou de ano, pois teve duas negativas

A média não é suficiente para caracterizar e diferenciar os dois conjuntos de dados

A dispersão faz com que a medida considerada para a média possa não ser representativa destas duas distribuições por pouca exatidão. A exatidão é uma medida do desvio entre o valor obtido e o verdadeiro valor.

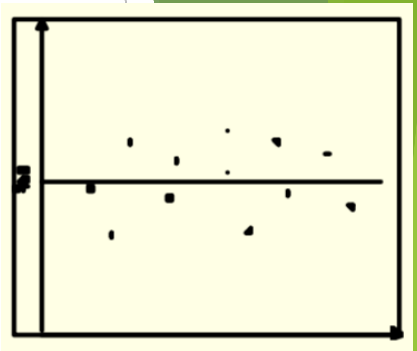
As medidas de localização não são suficientes para sintetizar um conjunto de observações isto porque, para além da localização torna-se, também, necessário conhecer a forma como estão distribuídos os valores ao longo do intervalo de variação.

Então, as **Medidas de Dispersão** têm por finalidade verificar a representatividade das medidas de localização.

MEDIDAS DE DISPERSÃO

Como medir a dispersão?

Não podemos limitar-nos a somar os desvios em relação à média? Porquê?



Porque soma dos desvios é zero

SOLUÇÃO

Considerar medida que não leve em conta o sinal dos desvios (o que importa é a magnitude do desvio)

- *Determinar o valor absoluto dos desvios* — Desvio absoluto médio
- *Determinar o quadrado dos desvios* — Variância

Medidas de Dispersão

- **ABSOLUTAS** (dependem das unidades de medida)
- **RELATIVAS** (não dependem das unidades de medida)

MEDIDAS DE DISPERSÃO

As medidas de dispersão **ABSOLUTAS** podem ser de dois tipos:

1) as que incluem as medidas de distância, apresentando-se os seus valores na mesma unidade de medida dos dados originais, não recorrendo, por regra, ao cálculo prévio de uma medida de localização (**Intervalo de Variação** e o **Intervalo Inter-quartis**)

2) as que utilizam uma medida de localização como termo de comparação (**Desvio Absoluto Médio,** a **Variância** e o **Desvio-padrão).**

MEDIDAS DE DISPERSÃO

INTERVALO de VARIAÇÃO

O **Intervalo de Variação (R) ou Amplitude do Intervalo de Variação** é a medida de dispersão mais intuitiva e mais simples de calcular. Para calcular o Intervalo de Variação basta fazer a diferença entre os valores máximo e mínimo da variável.

$$R = X_{Max} - X_{Min}$$

PR. 18

O exemplo 4 apresenta uma tabela onde constam as taxas de crescimento populacional em sete regiões do mundo.

Determine o intervalo de variação desta distribuição

Região	Taxa de crescimento
África	2,8
Ásia	2,1
América do Norte	0,8
América Latina	2,6
Países de Leste	0,9
Europa	0,4
Oceânia	1,3

MEDIDAS DE DISPERSÃO

INTERVALO INTERQUARTIS

O **Intervalo Interquartis ou Amplitude Interquartis (IQ)** é definido como a diferença entre o terceiro e o primeiro quartis e corresponde a um intervalo que engloba 50% das observações centrais.

PR. 19

Retomando novamente o exemplo 4, determine o intervalo interquartis (IQ) desta distribuição

Região	Taxa de crescimento
África	2,8
Ásia	2,1
América do Norte	0,8
América Latina	2,6
Países de Leste	0,9
Europa	0,4
Oceânia	1,3

MEDIDAS DE DISPERSÃO

DESVIO MÉDIO ABSOLUTO

O **Desvio Médio Absoluto (DM)** é a média aritmética entre os desvios absolutos de todos os seus elementos relativamente à média.

$$DM = \frac{\sum | x_i - \overline{x} |}{N}$$

O **processo** de cálculo do **desvio médio absoluto** é o seguinte:

1. Calcula-se o valor da média
2. Determinar a diferença entre cada valor observado e a média. Atendendo que o que interessa é a magnitude da diferença, o sinal desta operação será ignorado (deve considerar-se o valor absoluto).
3. Somar as diferenças e dividir pelo nº total de observações.

PR. 20

As idades das jogadoras de basquetebol de um determinado colégio são: 15; 16; 14; 17; 18. Calcule o desvio médio absoluto.

MEDIDAS DE DISPERSÃO

DESVIO MÉDIO ABSOLUTO (continuação)

Principais características

•É uma medida mais sensível que as anteriores por considerar a diferença de todas as observações relativamente a um valor central

• Essa diferenças são calculadas em relação a uma medida de tendência central que poderá ser a média ou a mediana

• A utilização de módulos faz com que o seu cálculo não seja tão imediato como as medidas anteriores

• É uma medida menos influenciada por valores extremos que o desvio padrão

•Em certas ocasiões, a utilização do desvio absoluto não é aconselhada por ignorar os sinais dos desvios.

MEDIDAS DE DISPERSÃO

VARIÂNCIA

Define-se a **variância** e representa-se por $\mathbf{s}^2$ ou $\boldsymbol{\sigma}^2$, como sendo a medida que se obtém somando os quadrados dos desvios das observações relativamente à média, e dividindo pelo número de observações.

Dados não agrupados ⟶

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

Os passos a seguir no cálculo da Variância são em tudo semelhantes aos indicados para o cálculo do desvio médio absoluto, exceto no que respeita às diferenças entre os valores da variável e a sua média que deverão ser elevados ao quadrado em vez de tomadas no seu valor absoluto.

PR. 21

Determine a variância da distribuição relativa às idades das jogadoras de basquetebol cujo enunciado consta da PR. 20.

MEDIDAS DE DISPERSÃO

VARIÂNCIA (continuação)

Dados agrupados

$$\sigma^2 = \frac{\sum_{i=1}^{N} Fi(X_i - \bar{X})^2}{N} \Leftrightarrow \sigma^2 = \sum_{i=1}^{N} fi\,(Xi - \bar{X})^2$$

De forma simplificada, a variância é a diferença entre a média aritmética do quadrado dos valores da variável e o quadrado da média

$$\sigma^2 = \sum f_i \,.\, X_i^2 - \bar{X}^2$$

$$\sigma^2 = \sum \frac{Fi}{n} \,.\, X_i^2 - \bar{X}^2$$

DESVANTAGEM da VARIÂNCIA

A variância envolve a soma de quadrados e por isso a unidade em que se exprime não é a mesma que a dos dados

MEDIDAS DE DISPERSÃO

DESVIO PADRÃO

Calculado a partir da raiz quadrada da variância, o desvio padrão é uma medida que **só pode assumir valores não negativos** e quanto **maior for**, **maior** será **a dispersão** dos dados.

Dados não agrupados

$$S \text{ ou } \sigma = +\sqrt{\sigma^2} \Leftrightarrow S \text{ ou } \sigma = +\sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{N}}$$

Dados agrupados

$$S \text{ ou } \sigma = +\sqrt{\frac{\sum_{i=1}^{n} F_i (X_i - \overline{X})^2}{N}}$$

$$S \text{ ou } \sigma = +\sqrt{\sum_{i=1}^{n} f_i (X_i - \overline{X})^2}$$

O **desvio padrão**, da mesma forma que a média, é *muito sensível à presença de outliers*, sendo portanto uma medida de dispersão pouco resistente.

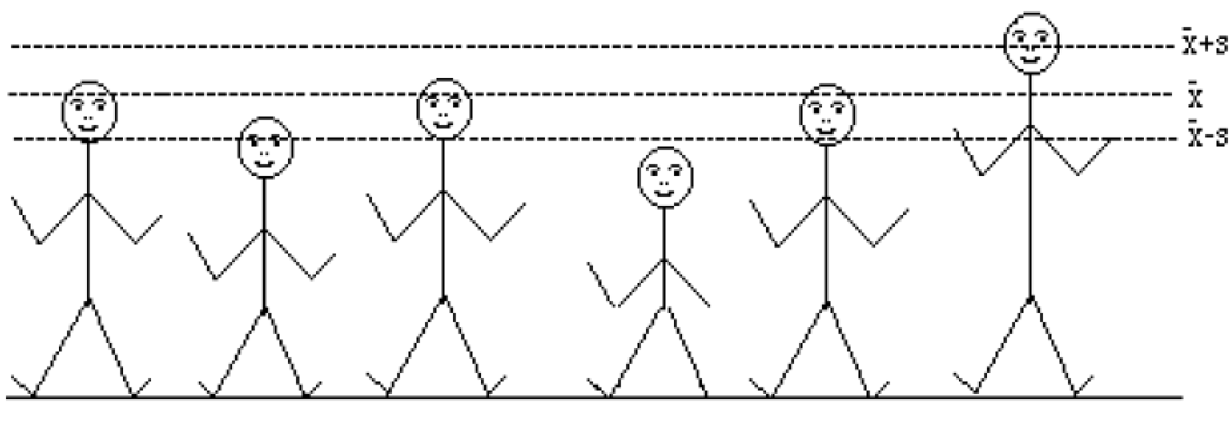
MEDIDAS DE DISPERSÃO

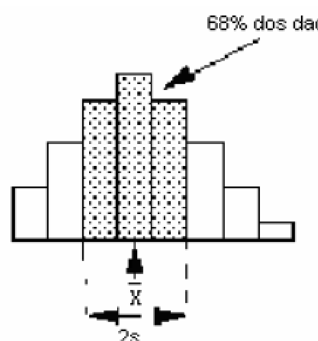
DESVIO PADRÃO (continuação)

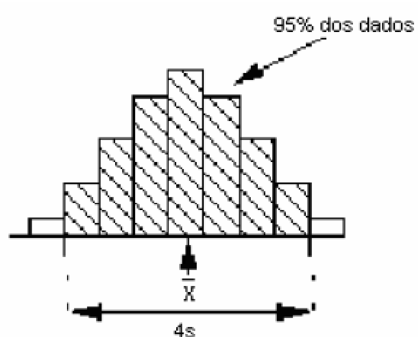
Propriedade para dados com distribuição aproximadamente normal

Uma propriedade que se verifica se os dados se distribuem de forma aproximadamente ***normal*** (ou seja, quando o histograma apresenta uma forma característica com uma classe média predominante e as outras classes distribuindo-se à volta desta de forma aproximadamente simétrica e com frequências a decrescer à medida que se afastam da classe média), é a seguinte:

Aproximadamente 68% dos dados estão no intervalo $[\bar{x} - s , \bar{x} + s]$







Aproximadamente 95% dos dados estão no intervalo $[\bar{x} - 2s , \bar{x} + 2s]$

ISTEC INSTITUTO SUPERIOR DE TECNOLOGIAS AVANÇADAS

MEDIDAS DE DISPERSÃO

PR. 22

Uma agência de viagens vende programas turísticos para os Açores, cujo número de dias é variável. Calcule o desvio médio absoluto, a variância e o desvio padrão da seguinte distribuição de frequências correspondente às vendas do mês de abril

Número de dias	5	7	8	9	11
Qt vendida	2	4	5	4	2

MEDIDAS DE DISPERSÃO

COEFICIENTE DE VARIAÇÃO

É uma medida relativa de dispersão, útil para a compreensão do grau de concentração em torno das médias de distribuições de frequências distintas.

É dado pela relação, em termos percentuais, entre o desvio padrão e a média

$$Cv = \frac{\sigma}{\overline{X}} * 100$$

EX. 6

Numa empresa o salário médio dos homens é de 4.000 com desvio padrão de 1.500 e o das mulheres é em média 3.000 com desvio padrão de 1.200. O coeficiente de variação é:

$$Cv_H = \frac{1.500}{4.000} \text{ x } 100 = 37{,}5\% \qquad Cv_M = \frac{1.200}{3.000} \text{ x } 100 = 40{,}0\%$$

Podemos concluir que os salários das mulheres apresentam uma maior dispersão relativa que os salários dos homens. Em termos práticos é usual considerar-se que um CV superior a 50% indica alto grau de dispersão relativa e consequentemente uma pequena representatividade da média como medida estatística

MEDIDAS DE DISPERSÃO

INSTITUTO SUPERIOR DE TECNOLOGIAS AVANÇADAS

PR. 23

Admita as seguintes estimativas referentes a três amostras de 10 telemóveis de marcas diferentes, onde $\bar{x}_1$ é a média de autonomia, em horas, da Huavei, $\bar{x}_2$ da Sony e $\bar{x}_3$ da Samsung, e S_1, S_2 e S_3 os respetivos desvios-padrão.

Marca	Média de autonomia	Desvio-padrão
Huawei	10h53m	3h34m
Sony	9h29m	3h07m
Samsung	8h58m	2h57m

Qual das marcas apresenta uma maior dispersão relativa no que respeita à autonomia?

MEDIDAS DE CONCENTRAÇÃO

Para medir a forma como determinada característica de uma população se distribui pelos seus elementos, é muitas vezes utilizada uma medida de concentração em vez de uma medida de dispersão.

Exemplos

•*Como é que a riqueza de um país está distribuída pelos seus cidadãos. Será que está igualmente distribuída, ou será que uma fração grande dessa riqueza está concentrada em poucos indivíduos?*

•*No total dos pontos obtidos pelos clubes da primeira liga, como é que esses pontos são distribuídos? Será que há um pequeno grupo de clubes que tem a maioria dos pontos?*

•*No total de dias por ano que os trabalhadores faltam ao emprego, como é que esses dias se distribuem pelos vários trabalhadores? Será que todos os trabalhadores faltam um igual número de dias, ou será que há um pequeno grupo de trabalhadores que são responsáveis por uma grande fração do total de dias de absentismo?*

MEDIDAS DE CONCENTRAÇÃO

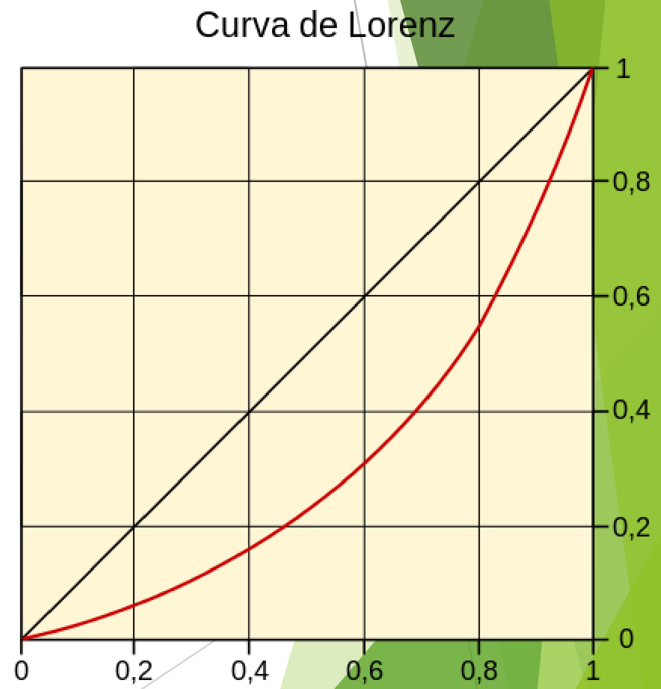
Curva de LORENZ

Dada uma característica (Yi) e as respetivas frequências (Fi) para cada intervalo ou classe i, obtém-se a Curva de Lorenz unindo, num referencial cartesiano, as frequências acumuladas para cada classe relativamente ao total

$$\text{Cum}\, f_i = \frac{\text{Cum}\, F_i}{\sum F_i}$$

com as frequências acumuladas do atributo em estudo, também relativamente ao total

$$\text{Cum}\, Y_i = \frac{\text{Cum}\, Y_i}{\sum Y_i}$$



Se a concentração for mínima o resultado será uma reta – reta de igual distribuição.

MEDIDAS DE CONCENTRAÇÃO

Curva de LORENZ (continuação)

EX. 7

Uma empresa tem 1500 clientes com os quais realizou no último ano um volume de negócios de 6.375 milhares de euros. Depois de ordenar os clientes por ordem crescente de volume de negócios, obteve a seguinte repartição:

• Com os 120 mais pequenos realiza negócios no valor de	30 mil euros
• Com os 240 seguintes realiza negócios no valor de	180 mil euros
• Com os 405 seguintes realiza negócios no valor de	607,5 mil euros
• Com os 270 seguintes realiza negócios no valor de	945 mil euros
• Com os 315 seguintes realiza negócios no valor de	2.362,5 mil euros
• Com os 150 finais realiza negócios no valor de	2.250 mil euros
No total 1500 clientes correspondem a	**6.375 mil euros**

Desenhe a curva de Lorenz.

MEDIDAS DE CONCENTRAÇÃO

Curva de LORENZ (continuação)

EX. 7

A repartição atrás indicada pode ser apresentada sob a forma de frequências relativas:

f_i			Y_i		
120 / 1500	8,0%	Dos clientes correspondem a	30 / 6375	0,5%	do volume de negócios
240 / 1500	16,0%		180 / 6375	2,8%	
405 / 1500	27,0%		607,5 / 6375	9,5%	
270 / 1500	18,0%		945 / 6375	14,8%	
315 / 1500	21,0%		2362,5 / 6375	37,1%	
150 / 1500	10,0%		2250 / 6375	35,3%	

Acumulando estes valores podem retirar-se as seguintes conclusões:

f_a			Y_a		
120 / 1500	8,0%	Dos clientes correspondem a	30 / 6375	0,5%	do volume de negócios
360 / 1500	24,0%		210 / 6375	3,3%	
765 / 1500	51,0%		817,5 / 6375	12,8%	
1035 / 1500	69,0%		1762,5 / 6375	27,6%	
1350 / 1500	90,0%		4125 / 6375	64,7%	
1500 / 1500	100,0%		6375 / 6375	100,0%	

MEDIDAS DE CONCENTRAÇÃO

Curva de LORENZ (continuação)

EX. 7

É a partir da representação gráfica das frequências acumuladas que se obtém a Curva de Lorenz.

- no eixo das abcissas, colocam-se os valores acumulados, em termos relativos, do número de ocorrências (cum fi)

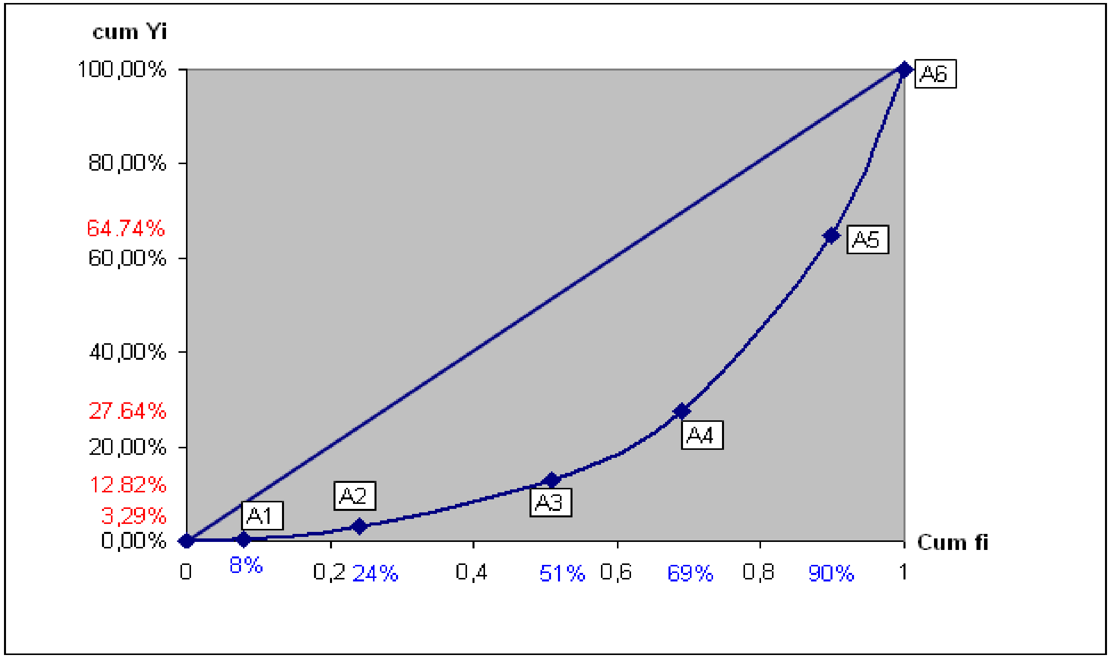
- no eixo das ordenadas colocam-se os valores correspondentes relativos ao atributo em estudo [neste exemplo % acumuladas de clientes (cum Yi)]

Pontos	Abcissas	Ordenadas
A_0	0	0
A_1	8%	0,47%
A_2	24%	3,29%
A_3	51%	12,82%
A_4	69%	27,65%
A_5	90%	64,71%
A_6	100%	100,00%

MEDIDAS DE CONCENTRAÇÃO

Curva de LORENZ (continuação)

EX. 7



Para este exemplo pode concluir-se que a empresa realiza metade do seu negócio com apenas 18% dos clientes e ainda que 51% dos clientes não ultrapassam 13% do volume de negócios. No seu conjunto estes valores indicam um grau relativamente elevado de concentração.

MEDIDAS DE CONCENTRAÇÃO

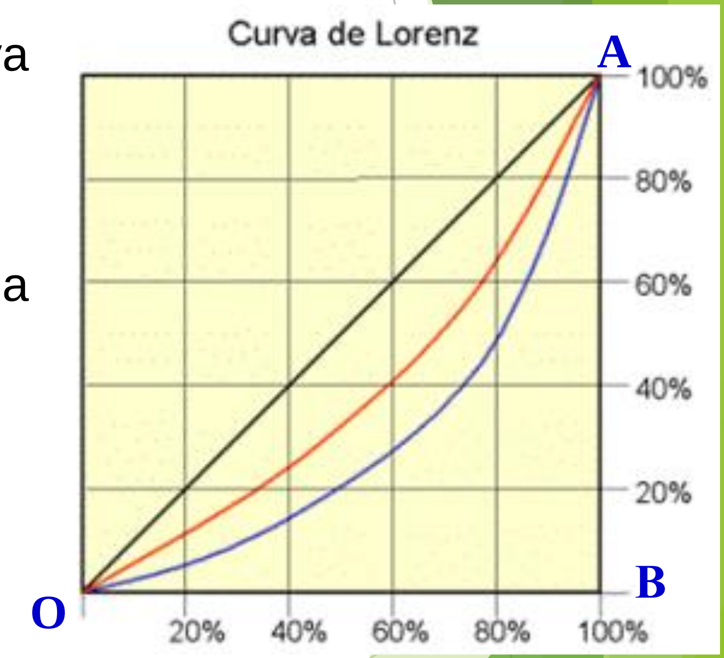
Curva de LORENZ (continuação)

Principais caraterísticas

- Quanto maior a concentração, mais a curva traçada se afastará da reta de igual distribuição.

- Quando a concentração é máxima, a área da concentração corresponderá ao triângulo OAB

- Seja $p_i = \dfrac{\text{Cum } F_i}{\sum F_i}$ e $q_i = \dfrac{\text{Cum } Y_i}{\sum Y_i}$



Quanto maior for a diferença entre p_i e q_i maior será a concentração

MEDIDAS DE CONCENTRAÇÃO

Índice de GINI

Com base nos valores de p_i e q_i nesta constatação Gini propôs o seguinte índice para medir o grau de concentração

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

***n** corresponde ao número de classes*

Este índice varia entre 0 e 1; é nulo quando em todas as classes houver igual distribuição do atributo pelos elementos em estudo e toma o valor máximo quando todo o atributo estiver concentrado nos indivíduos da última classe.

MEDIDAS DE CONCENTRAÇÃO

Índice de GINI (continuação)

PR. 24

A empresa GAMA pretende estudar como se distribuem as suas vendas pelos clientes. Pretende-se saber se as vendas se distribuem de igual modo por pequenas, médias e grandes empresas consumidoras ou se as grandes empresas constituem os principais clientes da empresa GAMA. Para tal, as empresas clientes foram distribuídas quanto à sua dimensão (o nº de empregados). Construa a curva de Lorenz da empresa GAMA e determine o índice de Gini

Nº de empregados	Nº de Empresas Fi	Volume de Vendas Yi
25-49	36	2.078
50-99	58	6.695
100-199	51	14.733
200-299	35	17.453
300-399	8	5.261
400-499	7	5.465
500-749	12	16.831
750-999	8	12.860
1000-1499	10	27.507
1500-2999	5	21.236
>=3000	3	23.302

REVISÃO

PR. 25

Uma escola avalia o seu curso através de um questionário com 50 perguntas sobre diversos aspetos de interesse. Cada pergunta tem uma resposta numa escala de 1 a 5, onde a maior nota significa melhor desempenho. Para cada aluno é então encontrada uma avaliação média. Na último semestre recorreu-se a uma amostra de 42 alunos, estando os resultados abaixo indicados:

4.2	2.7	4.6	2.5	3.3	4.7	4.0	2.4	3.9	1.2	4.1	4.0	3.1	2.4
3.8	3.8	1.8	4.5	2.7	2.2	3.7	2.2	4.4	2.8	2.3	1.9	3.6	3.9
2.3	3.4	3.3	1.8	3.5	4.1	2.2	3.0	4.1	3.4	3.2	2.2	3.0	2.8

(a) Proceda à organização dos dados construindo um quadro de frequências onde figurem as frequências absolutas, absolutas acumuladas e relativas acumuladas.

(b) Desenhe o respetivo histograma.

(c) Identifique as classes modal e mediana.

(d) Calcule a média e o desvio padrão usando os dados agrupados e também usando os dados não agrupados. Compare os resultados.

(e) Calcule os 1º e 3º quartis.

MEDIDAS DE ASSIMETRIA E CURTOSE

Como já foi dito anteriormente, as distribuições de frequências não diferem apenas quanto ao valor médio e à variabilidade, como também quanto à sua forma.

Do ponto de vista desse último aspeto, as características mais importantes são o ***grau de deformação*** **(assimetria**) e o ***grau de achatamento ou afilamento*** **(curtose)** da curva de frequências ou do histograma.

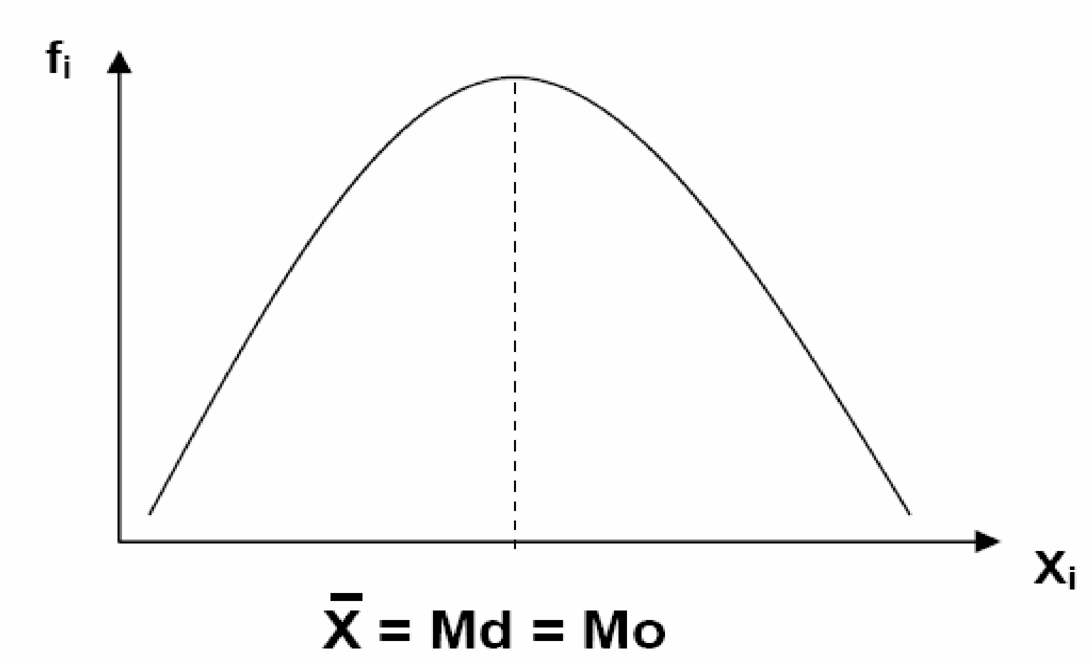
O que é assimetria?

Tal como indica a designação, significa desvio ou afastamento da simetria. Ou seja, assimetria é o grau de deformação de uma curva de frequências.

Para se determinar a sua existência, é fundamental analisar as relações entre: **média, mediana e moda**, ou seja:

MEDIDAS DE ASSIMETRIA E CURTOSE

Uma distribuição é **simétrica** quando apresenta o mesmo valor para a moda, a média e a mediana.



MEDIDAS DE ASSIMETRIA E CURTOSE

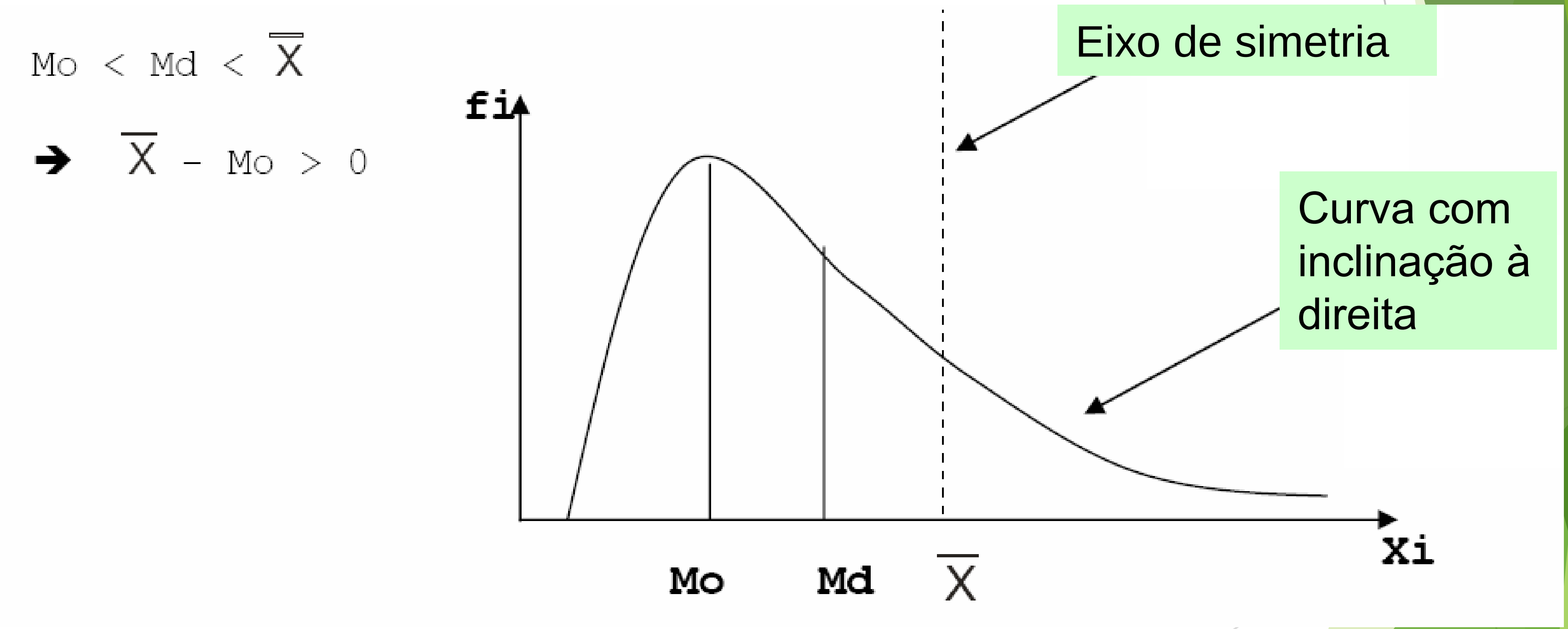
Quando esta igualdade não acontece, tem-se uma **distribuição assimétrica**.

Se considerar um eixo de referência, a que se dá o nome de eixo de simetria, traçado sobre o valor da média da distribuição, então sempre que a curva da distribuição se afastar do referido eixo, será considerada como tendo um certo grau de afastamento, que é considerado como uma assimetria da distribuição.

Ou seja assimetria é o grau de afastamento que uma distribuição apresenta do seu eixo de simetria. Este afastamento pode acontecer do lado esquerdo ou do lado direito da distribuição, chamado de assimetria negativa ou positiva respetivamente.

MEDIDAS DE ASSIMETRIA E CURTOSE

Quando a cauda da curva da distribuição declina para direita, temos uma distribuição com curva assimétrica **positiva**:

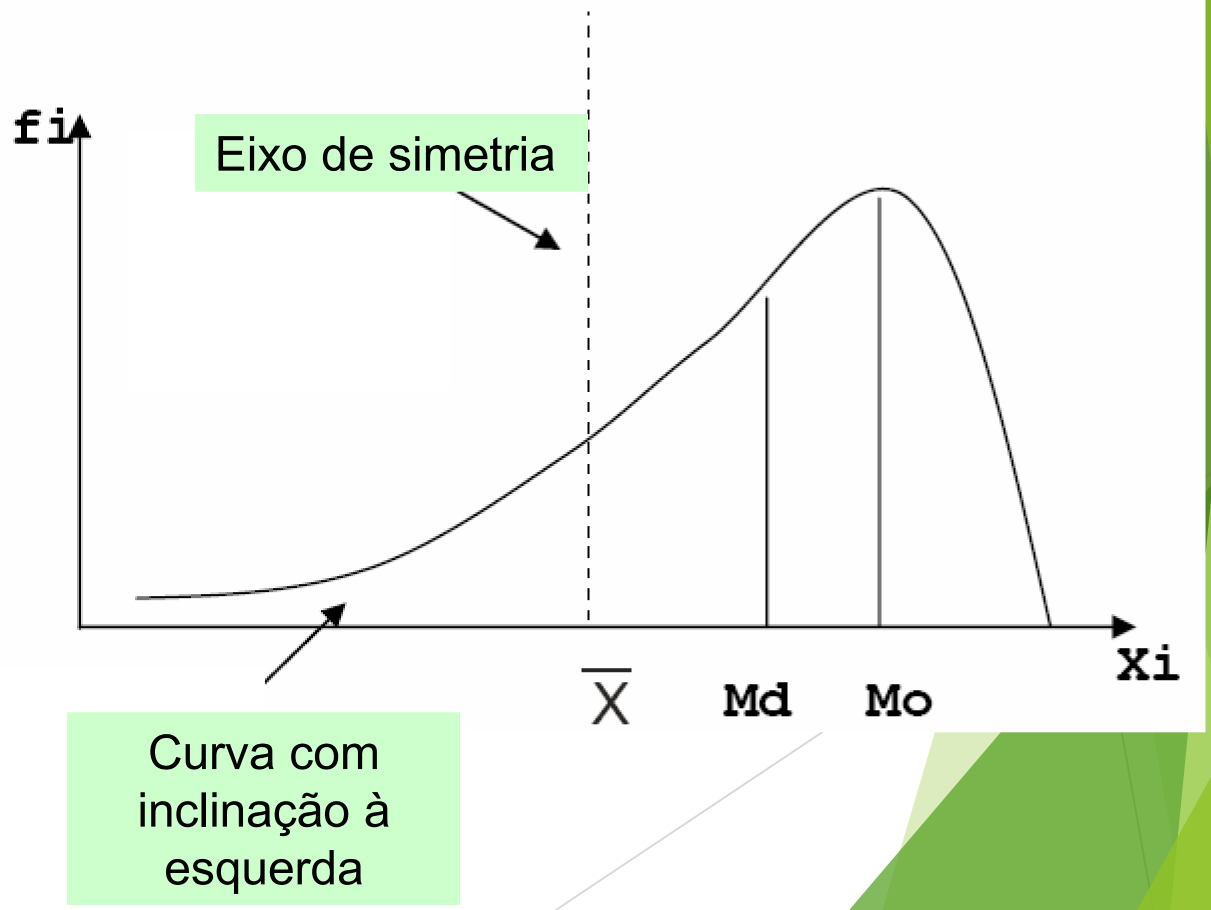


MEDIDAS DE ASSIMETRIA E CURTOSE

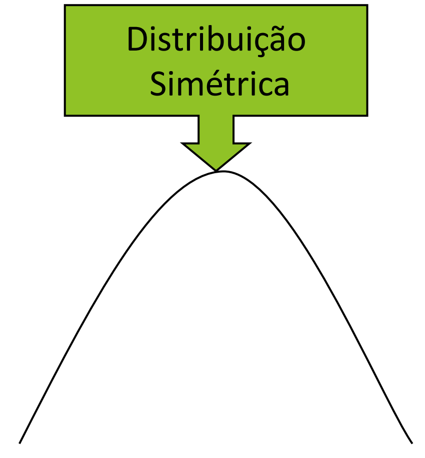
Analogamente quando a cauda da curva da distribuição declina para esquerda, temos uma distribuição com curva assimétrica **negativa**:

$$Mo > Md > \overline{X}$$

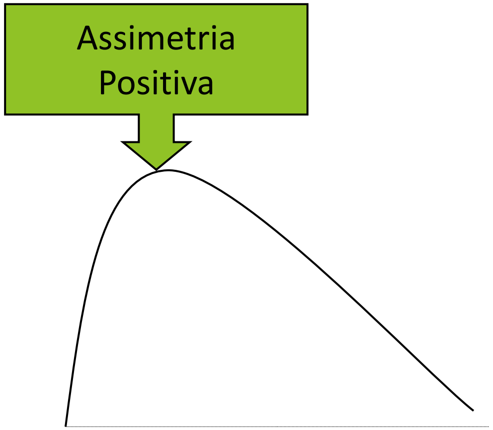
$$\rightarrow \quad \overline{X} - Mo < 0$$



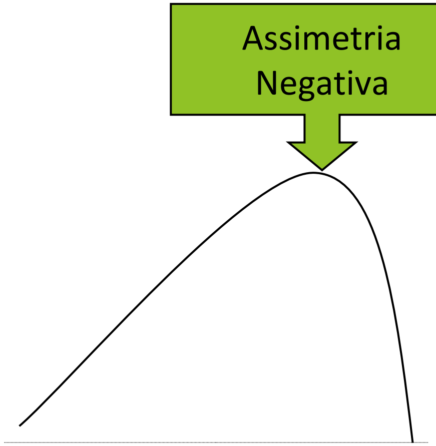
MEDIDAS DE ASSIMETRIA E CURTOSE



Media = Mediana = Moda



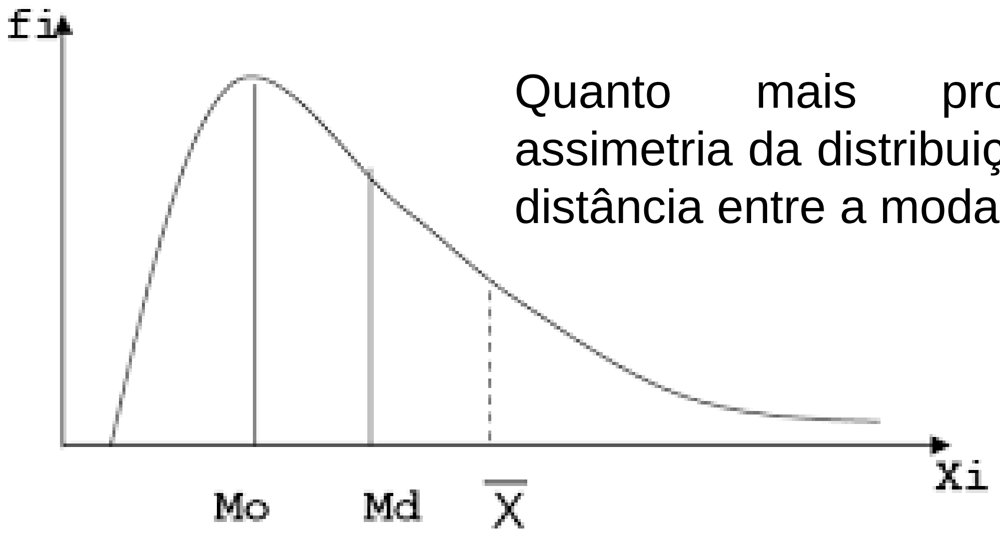
Moda <= Mediana <= Média



Média <= Mediana <= Moda

MEDIDAS DE ASSIMETRIA E CURTOSE

COEFICIENTE DE ASSIMETRIA



Quanto mais pronunciada for assimetria da distribuição maior será a distância entre a moda e a mediana

O coeficiente de assimetria da distribuição pode ser medido através de

$$G = \frac{3\,(média\ - mediana)}{desvio\ padrão}$$

MEDIDAS DE ASSIMETRIA E CURTOSE

COEFICIENTE DE ASSIMETRIA (continuação)

Segundo Pearson pode-se avaliar este grau de deformação pelos seguintes processos:

a) Primeiro Coeficiente de Pearson: baseia-se no peso que a diferença entre a moda e média têm relativamente ao desvio padrão

$$G_1 = \frac{\text{Média} - \text{Moda}}{\text{Desvio Padrão}}$$

Se G1 = 0 a distribuição é simétrica

Se G1 > 0 a distribuição é assimétrica positiva

Se G1 < 0 a distribuição é assimétrica negativa

MEDIDAS DE ASSIMETRIA E CURTOSE

COEFICIENTE DE ASSIMETRIA

Se não se dispuser da média e do desvio padrão, é possível calcular o grau de assimetria utilizando apenas os valores dos quartis, através do **Segundo Coeficiente de Pearson** (também chamado grau de Bowley)

$$G_2 = \frac{Q_3 + Q_1 - 2.mediana}{Q_3 - Q_1}$$

MEDIDAS DE ASSIMETRIA E CURTOSE

COEFICIENTE DE ASSIMETRIA (continuação)

Resumindo, o **Grau de Assimetria** pode ser medido da seguinte forma:

Aumenta a precisão

$$G = \frac{3(\text{Média - Mediana})}{\text{Desvio Padrão}}$$

$$G_1 = \frac{\text{Média} - \text{Moda}}{\text{Desvio Padrão}}$$

- Se G1 = 0 a distribuição é simétrica
- Se G1 > 0 a distribuição é assimétrica positiva
- Se G1 < 0 a distribuição é assimétrica negativa

$$G_2 = \frac{Q_3 + Q_1 - 2 * \text{Mediana}}{Q_3 - Q_1}$$

2º Coeficiente de Pearson

MEDIDAS DE ASSIMETRIA E CURTOSE

COEFICIENTE DE ASSIMETRIA (continuação)

PR. 26

Estudar o grau de simetria da seguinte distribuição de frequências relativa aos gastos dos alunos de engenharia multimédia com aquisição de livros técnicos durante a licenciatura.

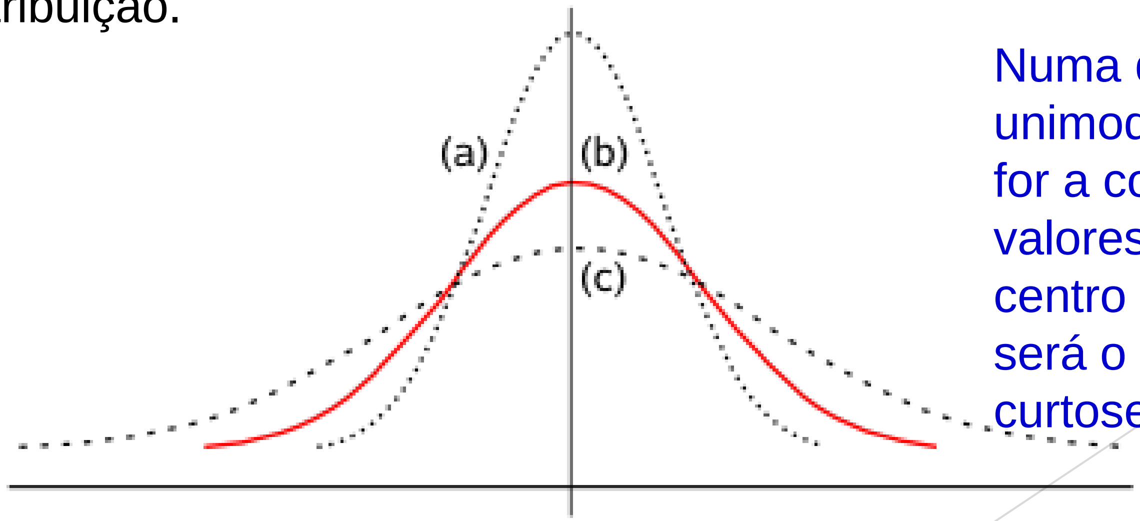
Gastos (€)	Frequência
0 – 20	21
20 – 40	35
40 – 60	30
60 – 80	23
80 – 100	28

MEDIDAS DE ASSIMETRIA E CURTOSE

ACHATAMENTO ou

Denomina-se curtose ao grau de “achatamento” de uma distribuição de frequências, geralmente unimodal, medido em relação ao de uma distribuição normal (de Gauss) que é tomada como padrão.

Muito embora seja comum explicar a curtose como o “grau de achatamento” de uma distribuição de frequências, o que as medidas de curtose tentam realmente indicar é o grau de concentração de valores da distribuição em torno do centro desta distribuição.



Numa distribuição unimodal, quanto maior for a concentração de valores em torno do centro da mesma, maior será o valor da sua curtose.

MEDIDAS DE ASSIMETRIA E CURTOSE

ACHATAMENTO ou CURTOSE (continuação)

COEFICIENTE PERCENTÍLICO DE CURTOSE

Este coeficiente é definido como o quociente entre a amplitude semi - interquartílica e a amplitude entre o 10º e o 90º percentis.

$$C_p = \frac{Q_3 - Q_1}{2\,(P_{90} - P_{10})}$$

O valor deste coeficiente para a curva normal é 0, 26367...

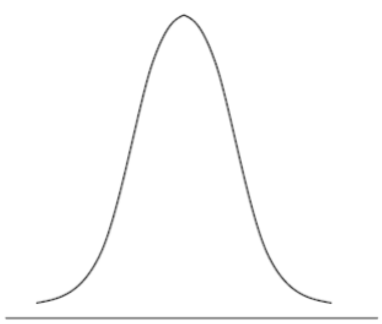
Assim sendo, ao calcular o coeficiente percentílico de curtose de uma distribuição qualquer tem-se:

- Quando Cp = 0,263 ➔ a distribuição é ***mesocúrtica***
- Quando Cp > 0,263 ➔ a distribuição é ***platicúrtica***
- Quando Cp < 0,263 ➔ a distribuição é ***leptocúrtica***

MEDIDAS DE ASSIMETRIA E CURTOSE

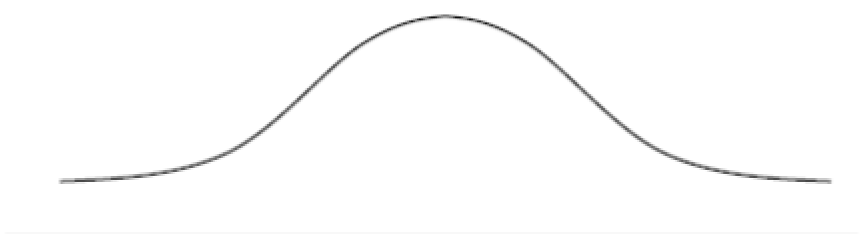
ACHATAMENTO ou CURTOSE (continuação)

Curva LEPTOCÚRTICA Alongada (K < 0.263).



Curva PLATICÚRTICA – achatada - (K > 0.263)

Curva MESOCÚRTICA- normal.
Nem achatada nem alongada- (K = 0.263).



MEDIDAS DE ASSIMETRIA E CURTOSE

ACHATAMENTO ou CURTOSE (continuação)

PR. 27

Analise a distribuição da prática 25 relativamente a simetria. Determine o grau da achatamento da distribuição.

4.2	2.7	4.6	2.5	3.3	4.7	4.0	2.4	3.9	1.2	4.1	4.0	3.1	2.4
3.8	3.8	1.8	4.5	2.7	2.2	3.7	2.2	4.4	2.8	2.3	1.9	3.6	3.9
2.3	3.4	3.3	1.8	3.5	4.1	2.2	3.0	4.1	3.4	3.2	2.2	3.0	2.8

REVISÃO

PR. 28

Admita que um grupo de 50 analistas financeiros efetuou uma previsão do ganho por ação, em euros, da **Sport Lisboa e Benfica SAD** no próximo ano, sendo os resultados apresentados em 7 classes de igual amplitude, na tabela seguinte:

Classes	Pontos Médios	Nr. de Analistas	Fa	fa
	5		4	0,08
			8	
[8 - 10[				
[10 - 12[		8	27	
	13		37	0,74
	17	5		1

a) Complete a tabela
b) Calcule a média, moda e mediana e interprete o seu valor

REVISÃO

PR. 29

Os dados da tabela que se segue dizem respeito ao preço (em euros) dos 50 eletrodomésticos vendidos numa loja durante o último fim-de-semana.

Preço de Venda (Ponto Médio da Classe)	Objectos Vendidos (Frequência Relativa)
150	?
250	0,2
?	?
?	?

Sabe-se que os preços médios e medianos dos artigos vendidos foram respetivamente 300 € e 290 € e ainda que neste estudo se utilizaram classes com a mesma amplitude.

a) Complete a tabela

b) Calcule o desvio absoluto médio e a amplitude do intervalo que engloba 4/5 do número total de observações (sob a condição de se situarem nas posições centrais da distribuição). Interprete os resultados obtidos.

c) Caracterize a distribuição em estudo relativamente à assimetria.

d) Construa a curva de Lorenz e comente os resultados obtidos

REVISÃO

PR. 30

No exame final de Estatística Descritiva, a nota média dos 50 alunos de Engenharia de Informática foi de 7,8 e o desvio padrão de 0,9.

Por seu turno, os alunos de Engenharia Multimédia obtiveram nota média final de 7,4 e desvio padrão de 0,8.

Que alunos apresentaram maior dispersão relativa?

NÚMEROS ÍNDICES

Definição

Um número índice é um quociente entre dois valores de uma mesma variável, referentes a diferentes pontos no espaço ou no tempo e expressos em percentagens

Porquê utilizar números índices?

Os Números Índices podem ser definidos como nºs relativos que permitem comparar os valores de uma variável em momentos diferentes ou em regiões diferentes. Para a realização desta comparação, em vez de se usarem os valores absolutos de uma grandeza utilizam-se os valores em relação a um período ou um valor tomado como base .

$$I_{t/_0} = \frac{V_t}{V_0} X\ 100$$

Número índice para o período t com base em 0

A grande vantagem destes índices é a de permitirem uma avaliação rápida da variação de um determinado período em relação ao período tomado como referência (período base).

NÚMEROS ÍNDICES

ÍNDICES SIMPLES

Quando se considera apenas uma variável ou grandeza

Sejam X_t e X_0 os valores de uma grandeza X, respetivamente no momento ou período t e no momento ou período de base “0”.

O índice $I_{t/0}$ relaciona o valor da grandeza no momento t com o valor no momento “0” e define-se como o quociente entre X_t e X_0

$$I_{t/0} = \frac{X_t}{X_0}$$

No caso de se desejar que o índice venha expresso em percentagem basta multiplicar por 100

$$I_{t/0} = \frac{X_t}{X_0} \times 100$$

NÚMEROS ÍNDICES

ÍNDICES SIMPLES (continuação)

Em que tipo de **grandezas** são utilizados?

Os números índices simples podem ser:

- de **preço** (quando se calcula a razão entre o preço observado de um artigo num período qualquer e o preço do mesmo artigo no período base)

- de **quantidade** (quando se calcula a razão entre a quantidade observada de um artigo num período qualquer e a quantidade no período base),

- e de **valor** (quando a razão é calculada pelo produto de preço e quantidade do artigo num período qualquer e o produto de preço e quantidade do mesmo artigo no período base)

Preço	Quantidade	Valor
$P_{t/0} = \frac{P_t}{P_0} x\, 100$	$Q_{t/0} = \frac{Q_t}{Q_0} x\, 100$	$V_{t/0} = \frac{V_t}{V_0} x\, 100$

Onde P_0 é o preço do artigo no período base, P_t é o preço do artigo em um período qualquer, Q_0 é quantidade do artigo no período base e Q_t é a quantidade do artigo em um período qualquer.

NÚMEROS ÍNDICES

ÍNDICES SIMPLES (continuação)

PROPRIEDADES

1	**Transitividade**	Um índice para o período I cuja base é o período 0 pode ser calculado através do produto de dois índices: o do período i com base em k e do período k com base em 0	
	$I_{i/0} = \frac{I_{i/k}}{100} \cdot \frac{I_{k/0}}{100} \cdot 100 \; com \; i > k > 0$	Ex: $I_{16/14} = I_{16/15} \cdot I_{15/14}$	
2	**Reversibilidade**	O índice para o período i com base 0 é igual ao inverso do mesmo índice para o ano 0 com base em i	
	$\frac{I_{i/0}}{100} \cdot \frac{100}{I_{0/i}}$	Ex: $\frac{I_{16/15}}{100} = \frac{100}{I_{15/16}}$	
3	**Encadeamento**	Um índice para o período i com base 0 pode ser descomposto no produto dos índices para todos os períodos desde i até 0, cada um deles tendo como base o período imediatamente anterior	
	$I_{i/0} = \frac{I_{i/i-1}}{100} \cdot \frac{I_{i-1/i-2}}{100} \ldots \frac{I_{1/0}}{100} \cdot 100$	Ex: $I_{16/13} = \frac{I_{16/15}}{100} \cdot \frac{I_{15/14}}{100} \cdot \frac{I_{14/13}}{100} \cdot 100$	

NÚMEROS ÍNDICES

ÍNDICES SIMPLES (continuação)

EX.8

Considerando os fluxos financeiros de Portugal com a União Europeia, construa os números índices com base em 1994 e determine as respetivas variações.

Anos	Pagamentos	Recebimentos	Saldo
1994	247,4	507,9	260,5
1995	196,2	715,1	518,9
1996	208,4	709,9	501,5
1997	214,6	750,1	535,5

NÚMEROS ÍNDICES

ÍNDICES SIMPLES (continuação)

EX.8

Anos	Pagamentos	Índice	Recebimentos	Índice	Saldo	Índice
1994	247,4	100	507,9	100	260,5	100
1995	196,2	79,3	715,1	140,8	518,9	199,2
1996	208,4	84,2	709,9	139,8	501,5	192,5
1997	214,6	86,7	750,1	147,7	535,5	205,6

Anos	Pagamentos	Índice	Var %	Recebimentos	Índice	Var %	Saldo	Índice	Var %
1994	247,4	100		507,9	100		260,5	100	
1995	196,2	79,3	-20,7	715,1	140,8	40,8	518,9	199,2	99,19
1996	208,4	84,2	-15,8	709,9	139,8	39,8	501,5	192,5	92,51
1997	214,6	86,7	-13,3	750,1	147,7	47,7	535,5	205,6	105,6

NÚMEROS ÍNDICES

ÍNDICES SIMPLES (continuação)

PR. 31

Uma siderurgia produz chapas de aço. No ano de 2013 a chapa custava 45 euros, e em 2014 atingiu o valor de 47,5 euros. Em 2013, a empresa produziu 1.500 chapas, e em 2014, 1.567 chapas.

Calcule os números índices de preço, quantidade e valor para a chapa de aço tomando o ano de 2013 como base.

NÚMEROS ÍNDICES

ÍNDICES AGREGADOS ou COMPOSTOS

Os números índices compostos expressam variações no preço, quantidade ou valor de um grupo de itens. Podem ser de dois tipos:

- agregados **simples** quando atribuem a mesma ponderação para todos os itens, desconsiderando a importância relativa de cada um.

- agregados **ponderados** atribuem ponderações diferentes para os itens, o que pode permitir dar maior enfâse às variações em determinado item, sendo a forma mais utilizada.

Os índices compostos mais utilizados são:

- Índice de **Laspeyres** (época básica): ponderação é feita em função dos preços ou do período base. Podem ser calculados índices de preço e quantidade.

- Índice de **Paasche** (época atual): ponderação é feita em função dos preços ou quantidades do período "atual". Podem ser calculados índices de preço e quantidade.

- Índice de **valores** : Não tem utilização generalizada como variável económica, mas é bastante utilizado na análise financeira. É calculado da mesma forma que um índice simples

NÚMEROS ÍNDICES

ÍNDICES AGREGADOS ou COMPOSTOS (continuação)

Índice de LASPEYRES

No índice de Laspeyres a ponderação é feita em função dos preços e quantidades do período base

Étienne Laspeyres
1834 - 1913

$$L_{Pi} = \frac{\sum P_i \, Q_0}{\sum P_0 \, Q_0} \, x \, 100 \longrightarrow$$

Índice de **preços** de Laspeyres para o período i

$$L_{Qi} = \frac{\sum P_0 \, Q_i}{\sum P_0 \, Q_0} \, x \, 100 \longrightarrow$$

Índice de **quantidades** de Laspeyres para o período i

São de cálculo fácil e rápido
Menor informação necessária

Resultados desatualizados se a variável sofrer grandes alterações ao longo do tempo

NÚMEROS ÍNDICES

ÍNDICES AGREGADOS ou COMPOSTOS (continuação)

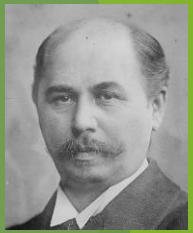
PR. 32

Para se produzir cabazes de natal, utilizam-se 3 artigos diferentes. Com os dados da tabela a seguir, e usando 2012 como base, obtenha os índices de Laspeyres de preço e quantidade.

Artigos	2012		2013		2014	
	Preço	Quantidade	Preço	Quantidade	Preço	Quantidade
1	2	4	2	5	3	6
2	3	3	4	2	6	3
3	5	2	6	5	8	6

NÚMEROS ÍNDICES

ÍNDICES AGREGADOS ou COMPOSTOS (continuação)



Hermann Paasche
1851 - 1925

Índice de PAASCHE

No índice de Paasche a ponderação é feita em função dos preços e quantidades do período atual

$$P_{Pi} = \frac{\sum P_i\, Q_i}{\sum P_0\, Q_i}\ x\ 100 \longrightarrow$$

Índice de **preços** de Paasche para o período i

$$P_{Qi} = \frac{\sum P_i\, Q_i}{\sum P_i\, Q_0}\ x\ 100 \longrightarrow$$

Índice de **quantidades** de Paasche para o período i

É mais dinâmico

Requer mais informação

NÚMEROS ÍNDICES

ÍNDICES AGREGADOS ou COMPOSTOS (continuação)

PR. 33

Utilizando os dados da Prática 33, e usando 2012 como base, obtenha os índices de Paasche de preços e quantidades.

Artigos	2012		2013		2014	
	Preço	Quantidade	Preço	Quantidade	Preço	Quantidade
1	2	4	2	5	3	6
2	3	3	4	2	6	3
3	5	2	6	5	8	6

NÚMEROS ÍNDICES

ÍNDICES AGREGADOS ou COMPOSTOS (continuação)

Índice de valores

É calculado multiplicando as quantidades pelos preços correspondentes, relacionando-se o resultado com o período base

$$I_{Vi} = \frac{\sum P_i Q_i}{\sum P_0 Q_0} \; x \; 100 \longrightarrow$$

Índice de **valores** para o período i

Pode ser apurado a partir do produto de um índice de Laspeyres por um de Paasche

$$I_{Vi} = \frac{L_{Pi}}{100} \cdot \frac{P_{Qi}}{100} \cdot 100 \quad \text{ou} \quad I_{Vi} = \frac{L_{Qi}}{100} \cdot \frac{P_{Pi}}{100} \cdot 100$$

É utilizado para analisar a *performance* das vendas

Não é comum como variável económica

NÚMEROS ÍNDICES

ÍNDICES AGREGADOS ou COMPOSTOS (continuação)

PR. 34

Utilizando os dados da Prática 33, obtenha os índices de valores dos artigos para 2013 e 2014

Artigos	2012		2013		2014	
	Preço	Quantidade	Preço	Quantidade	Preço	Quantidade
1	2	4	2	5	3	6
2	3	3	4	2	6	3
3	5	2	6	5	8	6

NÚMEROS ÍNDICES

MANIPULAÇÃO DE NÚMEROS ÍNDICES

Variações percentuais entre períodos

$$\Delta = \frac{I\ i{+}1 - Ii}{I_i}\ x\ 100$$

Alteração percentual entre dois períodos i e i+1

EX.9

Calcule as alterações percentuais para cada um dos anos

Anos	Índice de preços	Alteração %
5	125	-
6	135	(135-125)/125 x 100 = 8%
7	145	(145-135)/135 x 100 = 7,4%
8	155	(155-145)/145 x 100 = 6,8%

Apesar das diferenças nos valores do índices serem constantes, as alterações evidenciam uma quebra no crescimento dos preços

NÚMEROS ÍNDICES

MANIPULAÇÃO DE NÚMEROS ÍNDICES (continuação)

Mudança de base de um número índice

O procedimento é extremamente simples: basta dividir toda a série de números índices originais pelo número índice do período escolhido como nova base. Isso preservará as diferenças relativas entre eles.

EX.9

Mudar a base da série de números índices abaixo para 2011.

Ano	**2009**	2010	2011	2012	2013	2014
Índice	**100**	109,12	113,86	116,69	126,53	133,20

Ano	2009	2010	**2011**	2012	2013	2014
Índice	87,83	95,84	**100**	102,49	111,13	116,99

NÚMEROS ÍNDICES

DEFLAÇÃO DE SÉRIES CRONOLÓGICAS

"As variações de preço, causadas por inflação ou deflação, podem obscurecer as variações de quantidade". Isso significa que às vezes o que parece ser um crescimento de vendas, ou aumento na participação no mercado (por apresentar maior faturação) deve-se mais a flutuações de preços, ou desvalorizações cambiais, do que realmente a acréscimos nas quantidades vendidas. Este problema torna-se mais grave se estiverem a ser examinadas longas séries temporais

É preciso fazer a deflação da série temporal. Por outras palavras, remover o efeito da inflação nos valores da série temporal recorre-se a um índice de preços

Valor deflacionado = (valor original/índice) x 100

NÚMEROS ÍNDICES

DEFLAÇÃO DE SÉRIES CRONOLÓGICAS (continuação)

PR. 34

A tabela abaixo indica as vendas de uma empresa a preços correntes entre 1983 e 1987. Deflacione as vendas recorrendo ao índice de preços no consumidor (IPC)

Ano	Faturação	IPC
1983	120.345	100
1984	150.765	118
1985	180.555	135
1986	198.589	150
1987	256 650	162

AUXILIARES

Abreviatura	O que representa
F_i	Frequência absoluta simples
F_a	Frequência absoluta acumulada
f_i	Frequência relativa simples
f_a	Frequência relativa acumulada
P	Posição do quantil
Li	Limite inferior do intervalo ou classe
Ls	Limite superior do intervalo ou classe
h	Amplitude do intervalo (da classe)
r	Amplitude de variação (da distribuição)
C_i	Ponto central do intervalo ou classe
$\bar{X}$	Média
mo	Moda
m	Mediana

AUXILIARES

Pronúncia	Minúscula	Maiúscula
alfa	α	Α
beta	β	Β
gama	γ	Γ
delta	δ	Δ
épsilon	ε	Ε
dzeta	ζ	Ζ
eta	η	Η
teta	θ	Θ
iota	ι	Ι
capa	κ	Κ
lâmbda	λ	Λ
mi	μ	Μ

Pronúncia	Minúscula	Maiúscula
ni	ν	Ν
ksi	ξ	Ξ
omicron	ο	Ο
pi	π	Π
rho	ρ	Ρ
sigma	σ	Σ
tau	τ	Τ
upsilon	υ	Υ
phi	φ	Φ
khi	χ	Χ
psi	ψ	Ψ
ômega	ω	Ω