

Medidas de localização central

Profª Cesaltina Pires
cpires@uevora.pt

Plano da Apresentação

- ✍ **Medidas de localização central**
 - ✍ Média
 - ✍ Mediana
 - ✍ Moda
 - ✍ Outras médias: a média geométrica

Medidas de localização central

- ✍ Quando se tem um conjunto de dados é normal verificar-se uma concentração maior dos dados em torno de um certo valor «central»
- ✍ Muitas vezes esse valor «central» é usado como «indicador» do conjunto de dados, é usado para caracterizar os dados.
- ✍ Há várias medidas de tendência central. Tudo depende da forma como «centro» é definido.

Média

- ✍ A palavra **média** faz parte da linguagem comum:
 - ✍ O rendimento médio em Portugal é 74% do rendimento médio na EU.
 - ✍ A esperança de vida média em Portugal é de 80 anos
 - ✍ A média das notas na disciplina de Metodologia de Diagnóstico em 2003 foi 14.
- ✍ A média é muitas vezes usada como valor «representativo» de um conjunto de dados.

Média

A média aritmética (ou simplesmente média) obtêm-se somando todas as observações de um conjunto de dados e dividindo essa soma pelo nº de observações.

Exemplo – nota média em Metodologias

Aluno	Nota	Notação
1	12	x_1
2	15	x_2
3	11	x_3
4	18	x_4
5	17	x_5
6	14	x_6
7	12	x_7
8	13	x_8
Total	112	? x_i

Variável – nota
Nº de observações - 8

Média é:

$$\frac{12+15+11+18+17+14+12+13}{8} = 14$$

Média

Se tivermos um conjunto de n observações, com valores $x_1, x_2, \dots, x_n$ a média é dada por:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

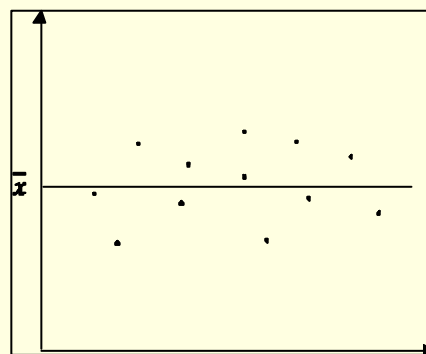
ou, usando notação de somatório

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

- ➔ Média leva em conta todas as observações
- ➔ Média é sensível à presença de valores extremos (outliers)
Neste caso, média pode não ser «representativa»

Média

- ✍ A média é o centro de «gravidade» da distribuição
- ✍ A média é o valor tal que há «equilíbrio» entre valores abaixo e valores acima da média.



Desvios em relação à média

Aluno	Nota	Desvio em relação à média
1	12	-2
2	15	+1
3	11	-3
4	18	+4
5	17	+3
6	14	0
7	12	-2
8	13	-1
Soma	122	0

Os desvios positivos e negativos compensam-se.

A soma dos desvios é zero.

A média é centro de gravidade

Propriedades da média

- ✍ Se todas as observações tiverem o mesmo valor, a média é igual a esse mesmo valor. Ou seja, a média de uma constante é a própria constante.
- ✍ Se somarmos uma constante **k** a cada um dos valores, a média também aumenta **k**
- ✍ Se multiplicarmos cada um dos valores por uma constante **k**, a média também vem multiplicada por **k**
- ✍ A média da soma de duas variáveis é igual à soma das médias

Cálculo das médias com dados agregados

- ✍ Se tivermos os dados originais, a média deve ser calculada com esses dados.
- ✍ Mas se só tivermos acesso ao dados agrupados (em classes) podemos calcular um valor aproximado da média.
- ✍ A ideia é usar o ponto médio de cada classe como valor representativo daquela classe. Por exemplo, se a classe é entre 10 e 20, o ponto médio é 15. O ponto médio é igual a $(\text{limite inferior da classe} + \text{limite superior})/2$.

Cálculo das médias com dados agregados

Idade	Nº Trabalhadores
[16,20]	27
[21,25]	41
[26,30]	24
[31,35]	19
[36,40]	4
Total	115

Média usando 115 observações originais é 25.11

Média usando dados agregados dá só um valor aproximado da verdadeira média

25.04

$$\frac{18 \times 27 + 23 \times 41 + 28 \times 24 + 33 \times 19 + 38 \times 4}{115}$$

Mediana

- ✍ Definição não muito rigorosa: a mediana é o valor da variável que tem tantas observações com valores mais baixos como observações com valores mais elevados (50% das observações têm valores mais baixos que a mediana, 50% das observações tem valores mais altos que a mediana).
- ✍ Para encontrar a mediana temos de começar por ordenar por ordem crescente as observações
- ✍ Grosso modo, a mediana fica no centro das observações (já ordenadas). Por isso se diz que a mediana é o «centro posicional».

Mediana – n° ímpar de observações

Dados originais

Aluno	Nota
1	12
2	15
3	11
4	18
5	17
6	14
7	12

Dados ordenados

Aluno	Nota
3	11
1	12
7	12
6	14
2	15
5	17
4	18

A nota mediana é 14, há 3 notas mais baixas que 14 e 3 notas mais altas que 14.

A mediana não é sensível a valores extremos.

Mediana – n° par de observações

Dados originais

Aluno	Nota
1	12
2	15
3	11
4	18
5	17
6	14
7	12
8	13

Dados ordenados

Aluno	Nota
3	11
1	12
7	12
8	13
6	14
2	15
5	17
4	18

Os dois valores centrais são 13 e 14. Fazendo a média desses valores centrais obtemos a mediana.

↓
13.5

Mediana – dados classificados

- ✍ Recordemos a noção de frequência acumulada. $F(c)$ é a frequência relativa de valores inferiores ou iguais a c . Ou seja, $F(c)$ indica-nos a proporção de observações que tem um valor inferior ou igual a c .
- ✍ A mediana é o valor M tal que $F(M) = 0.5$
- ✍ Se na tabela tivermos calculado frequências relativas acumuladas é fácil identificar a classe mediana.

Mediana – dados classificados

Idade	Frequência relativa	Frequência Relativa acumulada
[16,20]	0.235	23.5
[21,25]	0.356	59.1
[26,30]	0.209	80.0
[31,35]	0.165	96.5
[36,40]	0.035	100.0

A proporção de trabalhadores com 20 anos ou menos é 0.235.
A proporção de trabalhadores com 25 anos ou menos é 0.591.



A mediana é maior que 20, mas menor que 25. A mediana fica algures na classe [21,25]. Esta é a **classe mediana**.

Mediana – dados classificados

A proporção de trabalhadores com 20 anos ou menos é 0.235, a proporção de trabalhadores com 25 anos ou menos é 0.591.

Qual é o valor **M** tal que a proporção de trabalhadores com **M** anos ou menos é 0.5?

$$\frac{0.5 - 0.235}{M - 20} = \frac{0.591 - 0.235}{25 - 20} \quad \Rightarrow \quad M = 23.72$$

Cálculo admite que a proporção de trabalhadores na classe [21,25] se distribui de forma uniforme nessa classe.

Estatísticas de ordem

✍ A mediana é uma estatística de ordem, mas há mais...

✍ Quartis:

- ✍ 1º quartil – é o valor da variável que tem 25% das observações com valores mais baixos.
- ✍ 2º quartil – é o valor da variável que tem 50% das observações com valores mais baixos. (onde é que já ouvi isto???)
- ✍ 3º quartil – é o valor da variável que tem 75% das observações com valores mais baixos.

✍ Decis

- ✍ 1º decil – é o valor da variável que tem 10% das observações com valores mais baixos.

✍ ...

✍ Percentis

✍ ...

Moda

✍ A moda é o valor mais frequente da variável (o que é observado mais vezes)

✍ É mais interessante quando nº de observações é grande.
Com **n** pequeno pode nem haver valores repetidos.

✍ Como se identifica? Basta olhar para as frequências absolutas ou relativas e ver qual é o valor da variável com maior frequência.

✍ Pode haver mais do que uma moda

✍ No caso de dados classificados é fácil identificar a **classe modal**

Moda - exemplos

Nº pessoas no agregado familiar	Frequência absoluta n_i
1	1138
2	2748
3	2304
4	2082
5	848
6 ou mais	520

Moda = 2

Idade	Frequência relativa	Frequência Relativa acumulada
[16,20]	0.235	23.5
[21,25]	0.356	59.1
[26,30]	0.209	80.0
[31,35]	0.165	96.5
[36,40]	0.035	100.0

Classe modal é [21,25]

21

Medidas de localização central

✍ Qual das medidas é melhor?

- ✍ A média tem a vantagem de levar em conta todas as observações, não ignora informação.
- ✍ A média tem a desvantagem de ser sensível à presença de outliers.

✍ A «melhor medida» depende do nosso objectivo (que informação queremos obter com a medida utilizada)

- ✍ Exemplo: para o professor a nota média pode ser interessante. Para o aluno pode ser mais interessante a nota mediana (saber se está nos 50% melhores alunos, ou nos 50% piores)
- ✍ Exemplo: concurso com 50 vagas e 100 candidatos, a mediana é muito mais importante que a média.

Outra média – a média geométrica

- ✍ A média geométrica é aplicável quando as variáveis têm um efeito multiplicativo
 - ✍ Taxas de crescimento
 - ✍ Taxas de juro (quando há capitalização)
 - ✍ Taxa de inflação
- ✍ Como se deve calcular a taxa de crescimento médio? Ou taxa de juro média?

Taxas de crescimento

Ano	Nº de Acidentes	Crescimento	Taxa de crescimento anual
2001	20000		
2002	24000	4000	$4000/20000 = 0.2$ (20%)
2003	25000	1000	$1000/24000 = 0.041$ (4.1%)
2004	29000	4000	$4000/25000 = 0.16$ (16%)

De 2001 para 2002 o nº de acidentes aumentou 20%
Notar que $2400 = 2000 \times (1+0.2)$

Taxa de crescimento média

As taxas de crescimento anuais foram 20%, 4.1% e 16%. Qual é a taxa de crescimento média?

É a taxa g tal que se o nº de acidentes aumentasse em cada ano a essa taxa, teríamos o mesmo nº acidentes no final (29000).

$$2000(1+g)(1+g)(1+g) = 29000 \Leftrightarrow 2000(1+g)^3 = 29000$$

$$\Leftrightarrow (1+g)^3 = \frac{29000}{2000} \Leftrightarrow g = \sqrt[3]{\frac{29000}{2000}} - 1$$

$g = 0.131 \rightarrow$ A taxa de crescimento média anual foi de 13.1%

Taxa de crescimento média

No caso geral: se for V_0 o valor inicial e V_n o valor ao fim de n períodos, a taxa de crescimento média é dada por:

$$g = \sqrt[n]{\frac{V_n}{V_0}} - 1$$

Ou doutra maneira: se for g_1 a taxa de crescimento no primeiro período, g_2 a taxa de crescimento no segundo período,... A taxa de crescimento médio é:

$$g = \sqrt[n]{(1+g_1)(1+g_2)\dots(1+g_n)} - 1$$