MAT02018 - Estatística Descritiva

Medidas de tendência central

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística

Porto Alegre, 2021

Introdução

Introdução

- As **medidas de tendência central** (ou **de locação** ou também **de posição**) são **valores** calculados com o objetivo de representar os dados de uma forma ainda mais condensada do que usando uma tabela.
- Quando o desejo é representar, por meio de um valor único, determinado conjunto de informações que variam, parece razoável escolher um **valor central**, mesmo que esse valor seja uma abstração.

Introdução

- Assim, se um aluno realizou quatro provas objetivas com 20 questões cada uma, e o número de acertos, em cada prova, foi

$$(14, 19, 16, 17)$$

o professor poderá querer registrar o desempenho do aluno por um valor "central", a **média aritmética**, por exemplo.

- A média aritmética de tais valores é $16,5$.
- Embora seja uma quantidade de acertos que **não pode ocorrer na realidade**, ela está mostrando que o aluno **apresentou em geral um bom desempenho**.

Introdução

- Há várias medidas de tendência central.
- As mais utilizadas em análise estatística são a **média aritmética**, a **mediana** e a **moda**.
- Nestas notas apresentamos a definição destas medidas, e através de exemplos, como calculá-las e interpretá-las.

Média aritmética

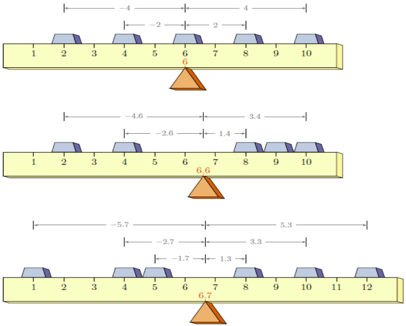
Média aritmética

- A **média aritmética**, ou simplesmente **média**, é a medida de tendência central mais utilizada.
- Isto se deve, em parte, por ser uma medida fácil de calcular, que tem interpretação familiar e propriedades estatísticas que a tornam muito útil nas comparações entre populações.
- Quando calculada para a população, também é chamada de **valor esperado** da variável, ou **esperança matemática**.
- Pode-se também imaginar a média como **centro de gravidade** de uma distribuição.
- Assim, na figura a seguir, os blocos azuis representam os valores observados para uma certa variável[1].

[1] Pense que estes blocos possuem um peso, de tal forma que a barra amarela se desequilibra conforme os blocos se movem em cima dela.

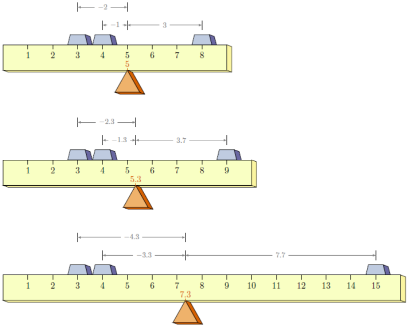
Média aritmética

- Note que conforme estes valores estão posicionados (distribuídos), a média (triângulo em vermelho), se desloca, de tal forma que este valor mantém o equilíbrio da distribuição (representada pela barra amarela).



Média aritmética

- Se um valor se afasta dos demais valores da distribuição, como ocorre na figura abaixo, a média também se deslocada para manter o equilíbrio da distribuição[2].



[2] **Para pensar:** neste caso, a média é uma boa representação sumária da distribuição?

Média aritmética

- Agora vamos definir a média aritmética.
- Seja x uma variável de interesse e $x_1, x_2, \ldots, x_n$ observações de x em um conjunto de dados (amostra ou população) de tamanho n.
- A média amostral é denotada por $\overline{x}$, enquanto que a média populacional é, geralmente, denotada por μ[3].
- Para dados que **não estão** agrupados, a média é simplesmente a soma de todos os valores observados da variável dividido pelo número de observações[4]

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

---

[3] Na população, o número de elementos é presentado por N.

[4] A letra grega Σ (sigma) é usada como símbolo para indicar uma soma. Assim, $\sum_{i=1}^{n} x_i$ indica o "somatório de x índice i, i variando de 1 a n". Ou seja, estamos somando os n elementos do conjunto $\{x_i; i = 1, \ldots, n\}$. Cada x_i representa o valor da variável x com respeito ao i-ésimo indivíduo. Por exemplo, se "Harry" apresentou o valor 19 anos para a variável idade (representada por x), e Harry foi o entrevistado de número 48 em um certo inquérito, temos que $x_{48} = 19$.

Média aritmética (exemplo)

- Suponha que, ao estudar a quantidade de albumina no plasma de pessoas com determinada doença, um pesquisador obtenha, em 25 indivíduos, os seguintes valores (em g/100 ML):

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

Média aritmética (exemplo)

- Considere que x representa a variável **quantidade de albumina no plasma**, e portanto, $x_1 = 5,1; x_2 = 4,9; \ldots ; x_{25} = 5,0$. A média, conforme a sua definição, é calculada

$$\overline{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = \frac{5,1 + 4,9 + \ldots + 5,0}{25} = \frac{126,5}{25} = 5,06.$$

Média aritmética (exemplo)

- Assim, temos que o valor médio da **quantidade de albumina no plasma** neste grupo de indivíduos é $5,06$ **g/100 mL**.
- Note que a média é uma característica da distribuição, e não de um ou outro indivíduo/observação em particular.
- As observações estão distribuídas em torno deste valor[5].
- Note que a média **é influenciada pelos valores extremos** da distribuição.
  - Se no lugar de $x_{20} = 5,5$ **(hipoteticamente)** observássemos o valor $6,8$, então teríamos que média seria $5,11$.

[5] Faça um diagrama de pontos com os valores do exemplo, e avalie se a média é o ponto de equilíbrio esta distribuição

Média aritmética (exercício)

Sua vez! (largura da pétala de íris): calcule a média para o seguinte conjunto de dados.

1,3	1,3	0,2	1,8	0,2
0,4	1,9	0,6	2,0	0,1

Cálculo da média para dados em agrupamento simples

Média para dados em agrupamento simples

- Para dados que são apresentados em agrupamentos simples[6],
- calcula-se a média do seguinte modo:

$$\overline{x} = \frac{\sum_i n_i x_i}{\sum_i n_i},$$

em que n_i é a frequência absoluta do valor x_i.

[6] Imagine que você não possui acesso a planilha de dados brutos, logo você não tem a informação a nível da observação. Mas, digamos que você tenha acesso a uma tabela de frequências (uma forma já resumida/agregada) da variável de interesse.

Média para dados em agrupamento simples

- Nota-se que, no caso de haver um agrupamento de dados, cada valor de x deve ser multiplicado pelo **número de vezes em que ele ocorre**, para depois se obter a soma.
- Podemos utilizar uma calculadora, fórmulas de planilhas eletrônicas, ou ainda funções de um *software* estatístico para nos auxiliar neste procedimento.

Média para dados em agrupamento simples

- **Exemplo (número de carburadores):** suponha que, em uma amostra de 32 diferentes modelos de automóveis, coletamos a informação com respeito ao número de carburadores e organizamos os dados em uma tabela de frequência.

Número de carburadores (x_i)	Frequência (n_i)
1	7
2	10
3	3
4	10
6	1
8	1
Total	32

Média para dados em agrupamento simples

- Multiplicando a primeira coluna (valores da variável x) pela segunda coluna (frequências de x) obtemos os elementos $n_i x_i$.

Número de carburadores (x_i)	Frequência (n_i)	$n_i x_i$
1	7	7
2	10	20
3	3	9
4	10	40
6	1	6
8	1	8
Total	32	90

Média para dados em agrupamento simples

- E assim temos que o número médio de carburadores neste grupo de automóveis estudado é

$$\overline{x} = \frac{(7 \times 1 + 10 \times 2 + \ldots + 8 \times 1)}{(7 + 10 + \ldots + 8)} = \frac{(7 + 20 + \ldots + 8)}{(7 + 10 + \ldots + 8)} = \frac{90}{32} = 2,8.$$

- Note que o número de carburadores, em cada carro, é um valor inteiro, porém, a média não precisa ser necessariamente um número inteiro[7].

[7] Lembramos novamente aqui que a média é uma característica da distribuição (do grupo e não de um indivíduo/unidade).

Média para dados em agrupamento simples

- Uma forma alternativa para calcular a média, no caso de agrupamentos simples, consiste em notar que $1/\sum_i n_i$ é uma constante, e portanto, podemos reescrever

$$\bar{x} = \frac{1}{\sum_{i=1} n_i} \sum_{i=1} n_i x_i = \sum_{i=1} x_i \frac{n_i}{\sum_{i=1} n_i}.$$

- Como $\sum_i n_i = n$[8] o cálculo da média é expresso em função dos seus valores e suas frequências relativas (f_i)

$$\bar{x} = \frac{1}{\sum_{i=1} n_i} \sum_{i=1} n_i x_i = \sum_{i=1} x_i \frac{n_i}{\sum_{i=1} n_i} = \sum_{i=1} x_i \frac{n_i}{n} = \sum_{i=1} x_i f_i.$$

[8] Salve o caso em temos dados ausentes. Neste caso n é o número de elementos com dados observados.

Média para dados em agrupamento simples

- Utilizando esta forma alternativa para o caso do exemplo dos carburadores, calculamos as frequências relativas, e em seguida multiplicamos as colunas x_i e f_i.
- A soma da coluna resultante é o valor da média do número de carburadores.

Número de carburadores (x_i)	Frequência (n_i)	f_i	$f_i x_i$
1	7	0,22	0,22
2	10	0,31	0,62
3	3	0,09	0,28
4	10	0,31	1,25
6	1	0,03	0,19
8	1	0,03	0,25
Total	32	1,00	2,81

Cálculo da média para dados em agrupamento por intervalo de classe

Média para dados em agrupamento por intervalo de classe

- Quando os dados estão organizados em uma **tabela com intervalos de classe**, é preciso haver um valor que represente cada intervalo.
- Este valor é o **ponto médio** do intervalo de classe, denotado por M[9].
- **Relembrando:** o ponto médio é a média aritmética dos dois extremos de classe

$$M = \frac{\text{limite inf. do int. de classe} + \text{limite sup. do int. de classe}}{2}.$$

[9] Lembre que já utilizamos o ponto médio (ou ponto central) do intervalo de classe na construção do **polígono de frequências**.

Média para dados em agrupamento por intervalo de classe

- A média é calculada da mesma forma que no caso do agrupamento simples, substituindo x_i por M_i:

$$\overline{x} = \frac{\sum_i n_i M_i}{\sum_i n_i} = \sum_i M_i \frac{n_i}{n} = \sum_i M_i f_i,$$

em que M_i, n_i e f_i representam, respectivamente, o **ponto médio**, a **frequência absoluta** e a **frequência relativa** da i-ésima classe.

Média para dados em agrupamento por intervalo de classe

Exemplo (milhas por galão): considere mais uma vez o exemplo do levantamento estatístico de 32 modelos de automóveis. Os dados referentes a variável *milhas por galão*, que é uma medida do desempenho do carro, foram organizados na tabela de frequências de intervalos de classe apresentada a seguir.

Milhas por galão	n_i	f_i
10,4 $\vdash$ 14,3	4	0,12
14,3 $\vdash$ 18,2	10	0,31
18,2 $\vdash$ 22,1	9	0,28
22,1 $\vdash$ 26,1	4	0,12
26,1 $\vdash$ 30,0	1	0,03
30,0 $\vdash$ 33,9	4	0,12
Total	32	1,00

Média para dados em agrupamento por intervalo de classe

- Para obter a média da distribuição da variável **milhas por galão**, calculamos os pontos médios (M_i) e multiplicamos estes pelas suas respectivas frequências relativas (f_i).
- Logo após, somamos esta coluna ($M_i f_i$).
  - O resultado é a média de milhas por galão neste conjunto de 32 automóveis.

Milhas por galão	n_i	f_i	M_i	$M_i f_i$
10,4 $\vdash$ 14,3	4	0,12	12,35	1,54
14,3 $\vdash$ 18,2	10	0,31	16,25	5,08
18,2 $\vdash$ 22,1	9	0,28	20,15	5,67
22,1 $\vdash$ 26,1	4	0,12	24,10	3,01
26,1 $\vdash$ 30,0	1	0,03	28,05	0,88
30,0 $\vdash$ 33,9	4	0,12	31,95	3,99
Total	32	1,00	-	20,17

Média para dados em agrupamento por intervalo de classe

Sua vez! (Teste de Desempenho Escolar): calcule a média para os dados agrupados em intervalos de classe para o exemplo de desempenho escolar (TDE) referente a 27 alunos.

Classe TDE	Frequência (n_i)
$7 \vdash 29$	4
$29 \vdash 51$	1
$51 \vdash 73$	0
$73 \vdash 95$	6
$95 \vdash 117$	15
Total	27

Para casa

1. Resolver os exercícios 1 a 3 do Capítulo 8.5 do livro **Fundamentos de Estatística**[10] (disponível no Sabi+).
2. Para o seu levantamento estatístico, calcule médias, modas e medianas, de acordo com a classificação das variáveis. Compartilhe no Fórum Geral do Moodle.

[10] Vieira, S. **Fundamentos de Estatística**, Atlas, 2019, p. 135-136.

Próxima aula

- Mediana e moda.
- Médias ponderada, geométrica e harmônica.
- ComplementaR.

Por hoje é só!

Bons estudos!