FIAP

FIAP

STATISTICAL COMPUTING WITH R

AULA 4

Prof. Reinaldo Borges Júnior

São Paulo - 2021

ESTATISTICA

Medidas de tendência central para dados agrupados: média, moda e mediana na tabela de frequência

- Para aprendermos como determinar as **medidas de tendência central (MÉDIA, MODA, MEDIANA)** por meio de uma tabela de frequência **(DADOS AGRUPADOS)** vamos utilizar o método de aprendizagem por resolução de problemas (PBL). Para tanto, considere o problema abaixo:

Qual o valor da média aritmética da distribuição de frequências a seguir?

Valores	Frequência
0 |– 2	10
2 |– 4	15
4 |– 6	40
6 |– 8	25
8 |–| 10	10

a) 4,9.
b) 5,2.
c) 5,3.
d) 5,5.
e) 6,1.

ESTATISTICA

Medidas de tendência central para dados agrupados: média, moda e mediana na tabela de frequência

Resolução do problema:

- Aqui temos uma tabela com intervalos de classes. A fórmula de cálculo da média aritmética ponderada será:

$$\bar{X} = \frac{\sum_{i=1}^{n}(\bar{x}.f_i)}{\sum_{i=1}^{n} f_i}$$

- Como vimos, a aplicação desta fórmula é simples, porém, devemos criar a tabela de frequência.

Tabela de distribuição de frequência

i	Valores (classes)	f_i	$\bar{x}$	$\bar{x}.f_i$	Fac_i
1	0 ├─ 2	10	1	10	10
2	2 ├─ 4	15	3	45	25
3	4 ├─ 6	40	5	200	65
4	6 ├─ 8	25	7	175	90
5	8 ├─┤ 10	10	9	90	100
Total		100		520	

$$\bar{X} = \frac{\sum_{i=1}^{n}(x.f_i)}{\sum_{i=1}^{n} f_i} = \frac{520}{100} = \mathbf{5,2}$$

ESTATISTICA

Medidas de tendência central para dados agrupados: média, moda e mediana na tabela de frequência

- **<u>Considerando o mesmo problema apresentado</u>.** Vamos, agora, vamos ampliar nossos estudos e conhecer a fórmula para moda para dados agrupados:

MODA: $$\mathbf{M_O = l_{Mo} + \left(\frac{d_1}{d_1+d_2}\right).\,h_{Mo}}$$

- Onde: a classe modal é a de maior frequência (em nosso problema, temos i = 3, ou seja, terceira classe)
- l_{Mo} = limite inferior da classe modal
- h_{Mo} = amplitude da classe
- $d_1 = f_i - f_{i-1}$ (Frequência da classe modal menos a frequência da classe anterior).
- $d_2 = f_i - f_{i+1}$ (Frequência da classe modal menos a frequência da classe posterior).

Substituindo na fórmula: $M_O = \cancel{4 +} \left(\frac{(40-15)}{(40-15)+(40-25)}\right).2 = 4 + \left(\frac{25}{25+15}\right).2 = 5{,}25$

Tabela de distribuição de frequência

i	Valores (classes)	f_i	$\overline{x}$	$\overline{x}.f_i$	Fac_i
1	0 ⊢— 2	10	1	10	10
2	2 ⊢— 4	15	3	45	25
3	4 ⊢— 6	40	5	200	65
4	6 ⊢— 8	25	7	175	90
5	8 ⊢⊣ 10	10	9	90	100
Total		100		520	

ESTATISTICA

Medidas de tendência central para dados agrupados: média, moda e mediana na tabela de frequência

- **MEDIANA:** para determinar a mediana em uma tabela de classes, primeiro vamos determinar **a posição** da mediana. Para isso, basta dividir o número de elementos “n” por 2, ou seja:

$$\mathbf{Pos_{Med}} = \frac{\mathbf{n}}{\mathbf{2}}$$

- A equação que calcula a mediana em dados agrupados é:

$$\mathbf{M_d} = \mathbf{l_{Md}} + \left(\frac{\frac{\mathbf{n}}{\mathbf{2}} - \mathbf{Fac_{i-1}}}{\mathbf{f_{Md}}} \right) . \mathbf{h_{Md}}$$

- Onde:
- $\mathbf{l_{Md}}$ = limite inferior da classe modal da mediana.
- Fac_{i-1} = frequência acumulada da classe anterior.
- f_{Md} = frequência (f_i) da classe em que se encontra a mediana.
- h_{Mo} = amplitude da classe onde está a mediana.

Tabela de distribuição de frequência

i	Valores (classes)	f_i	$\overline{x}$	$\overline{x}.f_i$	Fac_i
1	0 ⊢ 2	10	1	10	10
2	2 ⊢ 4	15	3	45	25
3	4 ⊢ 6	40	5	200	65
4	6 ⊢ 8	25	7	175	90
5	8 ⊢⊣ 10	10	9	90	100
Total		100		520	

ESTATISTICA

Medidas de tendência central para dados agrupados: média, moda e mediana na tabela de frequência

Resolução do problema:

- Já que tem-se 100 elementos, com isso, determinamos a posição da mediana:

$$\mathrm{Pos}_{\mathrm{Med}} = \frac{100}{2} = 50$$

- Podemos determinar a classe da mediana: **i = 3**
- Basta verificar a **frequência acumulada**: **Fac_i = 65** (significa que a mediana se encontra nessa Classe)
- Cálculo da Mediana (Md), fazemos a substituição dos dados obtidos, através da tabela, na sua equação:

$$M_{\mathrm{d}} = 4 + \left(\frac{50 - 25}{40}\right) . 2 = \mathbf{5,25}$$

- Veja, agora temos as três medidas de tendência central para nossa tabela de frequência:
- **Média: 5,2**
- **Moda: 5,25**
- **Mediana: 5,25**

- Observe que nossa distribuição de frequência é praticamente **simétrica**.

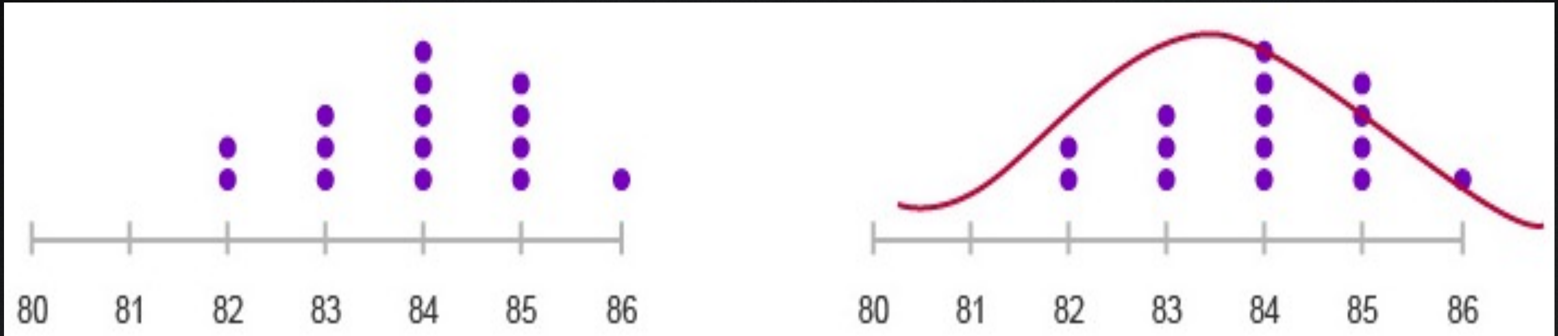
ESTATISTICA

Relação entre as medidas de tendência central: Gráfico de pontos e curva Gaussiana

- Existe um tipo de gráfico que será tratado agora, pois era preciso apresentar as **medidas de tendência central** para que ele fizesse sentido, é chamado de **Gráfico de Pontos (Diagrama de Pontos)** (Figura 1).

- Para sua construção se utiliza a **frequência (absoluta)** e com ele é possível criar **outro tipo de Gráfico em Curva (curva normal ou Gaussiana)** – teremos um momento específico para aprofundar nesse caso em particular, da distribuição normal dos dados) (Figura 2).

- Neste momento se faz necessário entender a relação entre a **tabela de frequência,** as **medidas de tendência central** e a **curva Gaussiana**.

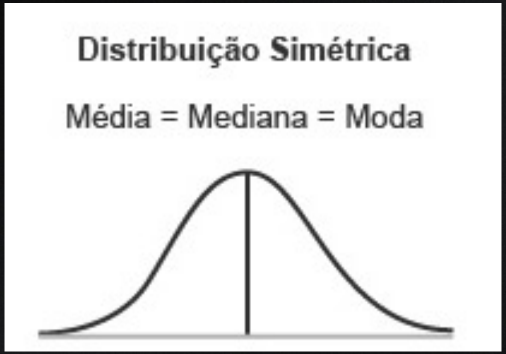


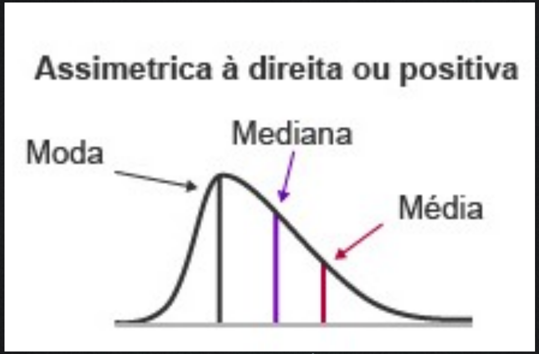
(1) **(2)**

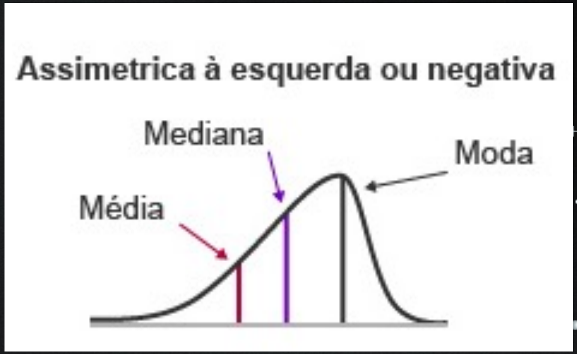
ESTATISTICA

Medidas de tendência central para dados agrupados: média, moda e mediana na tabela de frequência

- Desse modo, quando representarmos essas medidas de tendência central (MÉDIA, MODA e MEDIANA) em forma de uma curva, pode-se classificar em três tipos, como mostra a figura a seguir:







ESTATISTICA

Medidas de Posição: Separatrizes

- A mediana, como apresentado, permite localizar o elemento que divide os dados exatamente ao meio, ou seja, a mediana também se comporta como uma separatriz, abaixo dela, 50% dos valores e acima 50%.

- As separatrizes ou quartis são medidas de posição que permitem calcularmos valores da variável que dividem ou separam a distribuição em partes iguais, entre elas: a mediana (que é também uma medida de tendência central); os quartis (quartas partes ou 25%); os decis (décimas partes ou 10%); e os percentis (centis ou percentual em 100 partes iguais, ou seja, 1%).

- Quartis (Qi): dividem um conjunto de dados em quatro partes iguais, assim:

- **Q1: 1º quartil, organiza-se de modo a sua posição garantir que abaixo dele se encontram 25% dos elementos.**
- **Q2: 2º quartil, 50% dos elementos e coincide com a mediana.**
- **Q3: 3º quartil, 75% dos elementos antecedem o seu valor.**
- **Q4: 4º quartil, todos os elementos antecedem o seu valor.**

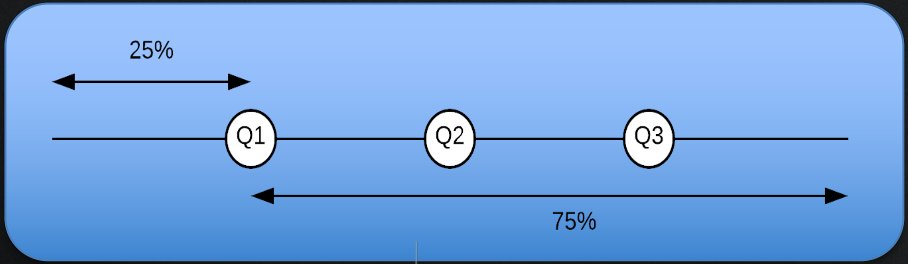
ESTATISTICA

Medidas de Posição: Separatrizes

- A posição de uma separatriz é dada por:

$$P_i = \frac{i.n}{100}$$

- Deste modo, é possível localizar quaisquer separatrizes no conjunto, o processo para os demais quantis é o mesmo, sejam os decis (em partes de 10%) e os centis ou percentis (em partes de 1%).



ESTATISTICA

Medidas de Posição: Separatrizes

EXERCÍCIO: Dado um conjunto de dados: 7, 13, 5, 12, 16, 4, 9, 15, 6. Calcule os Q1; Q2 e Q3.

- Escrevo os números em ordem crescente :

4 5 6 7 9 12 13 15 16

- Então:

Q1 = (5+6)/2 = 5,5

Q2 = 9

Q3 = (13+15)/2 = 14

ESTATISTICA

Medidas de Posição: Separatrizes

EXERCÍCIO: Calcular o Q1 da série de dados: 1, 2, 5, 5, 5, 8, 10, 11, 12, 12, 13, 15.

- Os números já estão em ordem crescente, então:

- Sabe-se que o primeiro quartil Q1 é o elemento que ocupa a posição 25%, logo:

$$Q_1 = P_{25} = \frac{25.12}{100} = 3 \rightarrow Q1 = 5$$

ESTATISTICA

Medidas de Posição: Separatrizes

- Se os dados estão apresentados na forma de variável discreta, eles já estão ordenados.

- Identificamos a posição da medida, como fizemos com a mediana na seção anterior, e determino o percentil correspondente Pi.

- Após, com o auxílio da frequência acumulada da série, localizamos o elemento que ocupa esta posição. O valor deste elemento é a separatriz desejada.

- A coluna da frequência acumulada se torna a principal aliada para determinar a classe na qual a separatriz pertence.

ESTATISTICA

Medidas de Posição: Separatrizes

Avaliação	Média de notas (M)
Excelente	$9 < M \leq 10$
Bom	$7 \leq M \leq 9$
Regular	$5 \leq M \leq 7$
Ruim	$3 \leq M \leq 5$
Péssimo	$M < 3$

PROBLEMA

A avaliação de rendimento de alunos de um curso universitário baseia-se na média ponderada das notas obtidas nas disciplinas pelos respectivos números de créditos, como mostra o quadro acima. Quanto melhor a avaliação de um aluno em determinado período letivo, maior sua prioridade na escolha de disciplinas para o período seguinte.

Determinado aluno sabe que se obtiver avaliação “Bom” ou “Excelente” conseguirá matrícula nas disciplinas que deseja.

Ele já realizou as provas de 4 das 5 disciplinas em que está matriculado, mas ainda não realizou a prova da disciplina I, conforme o quadro.

Para que atinja seu objetivo, a nota mínima que ele deve conseguir na disciplina I é?

Disciplinas	Notas	Número de créditos
I		12
II	8,00	4
III	6,00	8
IV	5,00	8
V	7,50	10

ESTATISTICA

Medidas de Posição: Separatrizes

Avaliação	Média de notas (M)
Excelente	$9 < M \leq 10$
Bom	$7 \leq M \leq 9$
Regular	$5 \leq M \leq 7$
Ruim	$3 \leq M \leq 5$
Péssimo	$M < 3$

RESOLUÇÃO DO PROBLEMA

M = (12.x + 4.8 + 8.6 + 8.5 + 10.7,5) / 42
M = (12x + 32 + 48 + 40 + 75) / 42

Para atingir o objetivo M = 7

7 = (12x + 195) /42
12x = 294 – 195
x = 99/12
x = 8,25

Disciplinas	Notas	Número de créditos
I		12
II	8,00	4
III	6,00	8
IV	5,00	8
V	7,50	10

OBRIGADO

/reinaldoborgesjunior