Probabilidade e estatística

Medidas Estatísticas

Tendência central, posição ou separatrizes

UESC – CiC – 2016.2

Graduandos: Eberty Alves e Felipe Oliveira
Professor: José Cláudio Faria

Tendência Central

- Medidas que orientam quanto aos **valores centrais**.
- Representam os fenômenos pelos seus valores médios, em torno dos quais tendem a se concentrar os dados.
- Também chamados de **centro da distribuição**.

---

1. **Média**
2. **Moda**
3. **Mediana**

Média

- **Média Aritmética**
  - **Dados Não Agrupados**
    - Seja $Y = (y_1, y_2, \ldots, y_n)$ e (N,n) = Quantidade de variáveis em Y, temos que:

$$\text{Parâmetro: } \bar{Y} \ ou\ \mu = \frac{\sum y}{N}$$

$$\text{Estimativa: } \bar{y} \ ou\ m = \frac{\sum y}{n}$$

  - **Exemplo:** considerando {3, 7, 8, 10, 11} como uma amostra

$$\bar{y} = \frac{3+7+8+10+11}{5} = 7,8$$

Média

- **Média Aritmética**
  - **Dados Agrupados**
    - Quando os dados estão agrupados em uma distribuição de frequência temos:
  - $\mathbf{Y=(y_1,\ y_2,\ \ldots,\ y_n)}$ ponderados pelas respectivas frequências absolutas $\mathbf{F=(F_1,\ F_2,\ \ldots,\ F_n)}$.

$$\bar{y} \text{ ou } \boldsymbol{m} = \frac{\sum \boldsymbol{y} \, . \, \boldsymbol{F}}{\sum \boldsymbol{F}}$$

Média

- **Média Aritmética**
  - **Dados Agrupados**
    - Exemplo:

Y	**F_i**	**$Y*F_i$**
1	1	1
2	3	6
3	5	15
4	1	4
Σ	**10**	**26**

$$\overline{y} \text{ ou } m = \frac{\sum y \,.\, F}{\sum F} = \frac{26}{10} = 2{,}6$$

Média

- **Média Aritmética**
  - **Dados Agrupados**
    - Exemplo:

Idade	F_i	Y	$Y*F_i$
02 ⊢ 04	5	3	15
04 ⊢ 06	10	5	50
06 ⊢ 08	14	7	98
08 ⊢ 10	8	9	72
10 ⊢ 12	3	11	33
Σ	**40**		**268**

$$\boxed{\bar{y} \text{ ou } m = \frac{\sum y \,.\, F}{\sum F} = \frac{268}{40} = 6{,}7}$$

Média

- **Média Geral**
  - Sejam $y_1, y_2, \ldots, y_k$ as **estimativas das médias aritméticas** de K séries.
  - Sejam $n_1, n_2, \ldots, n_k$ os **números de termos destas séries**, respectivamente.

- A média aritmética da série formada pelos termos da K séries é dada pela fórmula:

$$\boldsymbol{\bar{y}} \; ou \; \boldsymbol{m} = \frac{n_1\bar{y}_1 + n_2\bar{y}_2 + \ldots + n_k\bar{y}_k}{n_1 + n_2 + \ldots + n_k}$$

Média

- **Média Geral**
  - **Exemplo**
    - **Dadas as Séries:**

{4, 5, 6, 7, 8}	$n_1 = 5$	$\bar{Y}_1 = 6$
{1, 2, 3}	$n_2 = 3$	$\bar{Y}_2 = 2$
{9, 10, 11, 12, 13}	$n_3 = 5$	$\bar{Y}_3 = 11$

$$\bar{Y} = \frac{n_1\bar{y}_1 + n_2\bar{y}_2 + \ldots + n_k\bar{y}_k}{n_1 + n_2 + \ldots + n_k} = \frac{5.6 + 3.2 + 5.11}{5 + 3 + 5} = 7$$

Média

- **Média Geométrica**
  - Usada para médias proporcionais de crescimento quando uma **medida subsequente depende de medidas prévias**.

- Sejam $\mathbf{y_1, y_2, \ldots, y_n}$, valores de Y associados às respectivas frequências absolutas $\mathbf{F_1, F_2, \ldots, F_n}$. **A média geométrica de Y** é definida por:

$$\boldsymbol{MG\ ou\ mg = \sqrt[n]{y_1^{f_1} \,.\, y_2^{f_2} \,.\, \ldots \,.\, y_n^{f_n}}}$$

Média

- **Média Geométrica**
  - Exemplo:

Amostra: {3, 6, 12, 24, 48}

$$\boldsymbol{mg = \sqrt[5]{3\,.\,6\,.\,12\,.\,24\,.\,48} = \sqrt[5]{248\,832} = 12}$$

Média

- **Média Harmônica**
  - Usada para médias de crescimento e proporções de velocidade.

    - Sejam $\mathbf{y_1, y_2, \ldots, y_n}$, valores de Y
    - Sejam $\mathbf{F_1, F_2, \ldots, F_n}$, frequências absolutas.

- A **média harmônica de Y** é definida por:

$$\boldsymbol{MH} \text{ ou } \boldsymbol{mh} = \frac{\boldsymbol{n}}{\frac{\boldsymbol{F_1}}{\boldsymbol{y_1}} + \frac{\boldsymbol{F_2}}{\boldsymbol{y_2}} + \ldots + \frac{\boldsymbol{F_n}}{\boldsymbol{y_n}}} = \frac{\boldsymbol{n}}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \frac{\boldsymbol{F_i}}{\boldsymbol{y_i}}}$$

Média

- **Média Harmônica**
  - Exemplo:

Amostra: {2, 5, 8}

$$mh = \frac{n}{\frac{F_1}{y_1} + \frac{F_2}{y_2} + \frac{F_3}{y_3}} = \frac{3}{\frac{1}{2} + \frac{1}{5} + \frac{1}{8}} = \frac{3}{\frac{20+8+5}{40}} =$$

$$\frac{3}{\frac{33}{40}} = \frac{3}{1} \cdot \frac{40}{33} = 3,63$$

Média

Vantagens

- Fácil de compreender e calcular
- Utiliza todos os valores da série
- É um valor único
- É fácil de ser incluída em expressões matemáticas
- Pode ser determinada nas escalas: intervalar e proporcional.

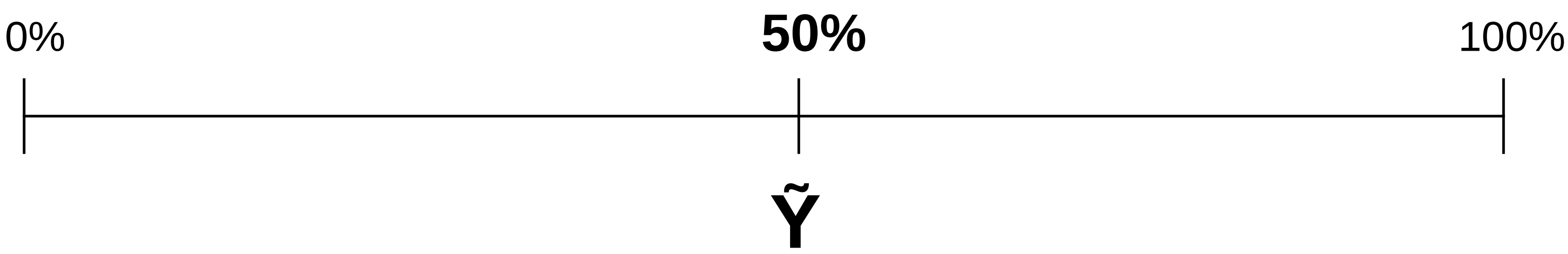
Desvantagens

- Muito afetada por valores extremos
- Necessário conhecer todos os valores da série.

Mediana

- Medida de tendência muito usada quando o interesse é a **determinação do valor que separa a série de dados em duas partes iguais**, 50% situados acima e 50% situados abaixo da medida.

  - Notação adotada: **($\tilde{Y}$ ou MD)** para o parâmetro e **($\tilde{y}$ ou md)** para a estimativa.



Mediana

- **Cálculo da mediana para variável discreta**

Se n for ímpar:

- A mediana é o elemento central.

$$\frac{n+1}{2}$$

Se n for par:

- A mediana é a média dos dois elementos centrais.

$$\text{Média } \left[\frac{n}{2}, \frac{n}{2}+1\right]$$

Mediana

- **Cálculo da mediana para variável discreta**
  - **Exemplo**

Y_i	F_i	Fac
1	1	1
2	3	4
3	5	9
4	2	11
Σ	**11**	

$$n = 11$$

$$\frac{11+1}{2}$$

$$\frac{11+1}{2} = 6°$$

$$\tilde{y} = 3$$

1	2	3	4	5	6	7	8	9	10	11
1	**2**	**2**	**2**	**3**	**3**	**3**	**3**	**3**	**4**	**4**

Mediana

- **Cálculo da mediana para variável discreta**
  - **Exemplo**

Yi	Fi	Fac
82	5	5
85	10	15
87	15	30
89	8	38
90	4	42
Σ	**42**	

$$\boldsymbol{n = 42}$$

Média $\left[\frac{\boldsymbol{n}}{\boldsymbol{2}}, \frac{\boldsymbol{n}}{\boldsymbol{2}} + \boldsymbol{1}\right]$

$$\frac{42}{2}, \frac{42}{2} + 1$$

(elementos 21º e 22º)

21º=87 22º=87

ỹ = **87**

Mediana

- **Cálculo da mediana para variável contínua**

1. Calcular $\frac{n}{2}$
2. Usar a ***Fac*** para identificar a classe que contém a mediana *(classe md)*
3. $\tilde{y} = l_{md} + \frac{\left(\frac{n}{2} - \sum f\right).h}{F_{md}}$

- l_{md} = Limite inferior da classe md
- n = Tamanho da série
- $\sum f$ = Soma das frequências anteriores à classe md
- h = Amplitude da classe md
- F_{md} = Frequência da classe md

Mediana

- **Cálculo da mediana para variável contínua**
  - **Exemplo**

Classe	Fi	Fac
35 ⊢ 45	5	5
45 ⊢ 55	12	17
55 ⊢ 65	18	35
65 ⊢ 75	14	49
75 ⊢ 85	6	55
85 ⊢ 95	3	58
Σ	**58**	**268**

1. $\frac{\mathbf{58}}{\mathbf{2}} = \mathbf{29°}$

2. Classe md = 3°

$l_{md} = 55$

$\mathrm{n} = 58$

$\sum f = 17$

$\mathrm{h} = 10$

$F_{md} = 18$

3. $\tilde{y} = 55 + \frac{\left(\frac{58}{2} - 17\right).10}{18} = \mathbf{61,67}$

Mediana

Vantagens

- Fácil de compreender e aplicar
- Não é afetada por valores extremos
- É um valor único
- Pode ser determinada nas escalas: ordinal, intervalar e proporcional.

Desvantagens

- É difícil de ser incluída em expressões matemáticas
- Não usa todos os valores da série.

Moda

- Medida de tendência central muito usada quando o interesse é o **valor mais frequente da série**.

- Notação adotada: (MO) para o parâmetro e (mo) para a estimativa.

  - Série sem moda: **Série amodal**
  - Mais de uma moda: **Série multimodal**

Moda

- **Exemplo**
  - Distribuição sem agrupamento de classes:

y_i	253	245	248	251	307
F_i	7	17	23	20	8

mo: 248

Moda

- **Exemplo**
  - Distribuição com agrupamento de classes

1. Identifica-se a classe modal
2. Aplique a fórmula de Czuber

$$\mathbf{mo} = \boldsymbol{l} + \frac{\boldsymbol{\Delta_1}}{\boldsymbol{\Delta_1} + \boldsymbol{\Delta_2}} \,.\, \boldsymbol{h}$$

- $l = Limite\ inferior\ da\ classe\ mo$
- $\Delta_1 =$ D$iferen$ç$a\ entre\ a\ frequ$ê$ncia\ da\ classe\ modal\ e\ a\ imediatamente\ anterior$
- $\Delta_2 =$ D$iferen$ç$a\ entre\ a\ frequ$ê$ncia\ da\ classe\ modal\ e\ a\ imediatamente\ posterior$
- $h\ = Amplitude\ da\ classe$

Moda

- **Exemplo**
  - Distribuição com agrupamento de classes:

$Classes$	$0 \vdash 1$	$1 \vdash 2$	$2 \vdash 3$	$3 \vdash 4$	$4 \vdash 5$	Σ
F_i	3	10	17	8	5	43

1. Classe modal = 3ª ($2 \vdash 3$)
2. $\mathrm{m}o = 2 + \frac{7}{7+9} \,.\, 1 = 2{,}44$

Moda

Vantagens

- Fácil de compreender e calcular
- Não é afetada por valores extremos
- Pode ser aplicada em todas as escalas: nominal, ordinal, intervalar e proporcional.

Desvantagens

- Pode estar afastada do centro dos valores
- É difícil de ser incluída em expressões matemáticas
- Não usa todos os valores da série
- A variável pode ter mais de uma moda (bimodal ou multimodal)
- Algumas variáveis não possuem moda.

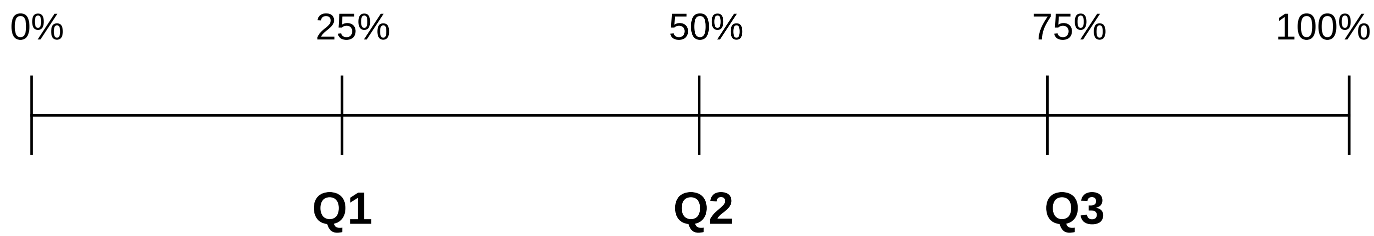
Medidas de Posição ou Separatrizes

- Genericamente denominadas **quantis**.
- Orientam quanto à **posição na distribuição**.
- Permitem determinar **valores que particionam a série** de n observações em partes iguais.

1. **Quartis**
2. **Decis**
3. **Percentis**

Quartis

- **Dividem uma série em 4 partes iguais**



- **Notação adotada:** (Q) para o parâmetro e (q) para a estimativa.

Quartis

$$\boldsymbol{q_i = l_{q_i} + \frac{\left(\frac{i.n}{4} - \sum f\right).h}{F_{q_i}}}$$

- l_{q_i} = limite inferior da classe q_i (i = 1, ..., 3)
- i = 1 para q_1, ..., 3 para q_3
- n = tamanho da série
- Σ f = soma das frequências anteriores à classe q_i
- h = amplitude da classe q_i
- F_{q_i} = frequência da classe q_i

Quartis

- **Quartis para dados agrupados em classes - Exemplo**

Classe	Fi	Fac
50 ⊢ 54	4	4
54 ⊢ 58	9	13
58 ⊢ 62	11	24
62 ⊢ 66	8	32
66 ⊢ 70	5	37
70 ⊢ 74	3	40
Σ	**40**	

$$q_i = l_{q_i} + \frac{\left(\frac{i.n}{4} - \sum f\right).h}{F_{q_i}}$$

Para q1 temos 1*40/4 = 10

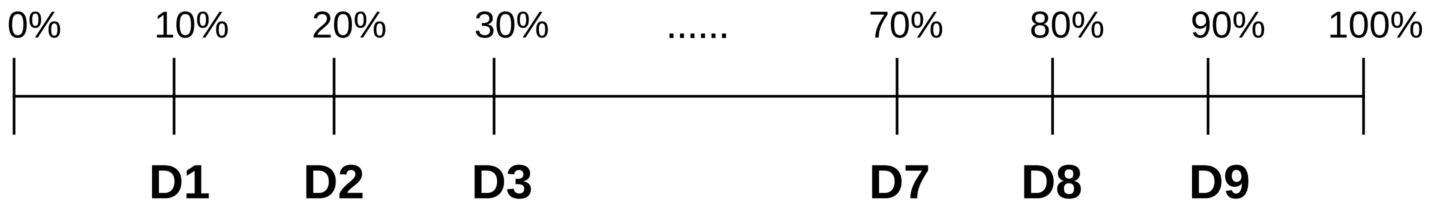
q1 = 54 + [(10 - 4) x 4] / 9 = 54 + 2,66 = 56,66

Para q2 temos 2*40/4 = 20

q2 = 58 + [(20 - 13) x 4] / 11 = 58 + 28/11 = 60,54

Decis

- **Dividem uma série em 10 partes iguais**



- **Notação adotada:** (D) para o parâmetro e (d) para a estimativa.

Decis

$$d_i = l_{d_i} + \frac{\left(\frac{i.n}{10} - \sum f\right).h}{F_{d_i}}$$

- l_{d_i} = limite inferior da classe d_i (i = 1, ..., 9)
- i = 1 para d_1, ..., 9 para d_9
- n = tamanho da série
- Σ f = soma das frequências anteriores à classe d_i
- h = amplitude da classe d_i
- F_{d_i} = frequência da classe d_i

Decis

- **Exemplo**

Classe	Fi	Fac
50 ⊢ 54	4	4
54 ⊢ 58	9	13
58 ⊢ 62	11	24
62 ⊢ 66	8	32
66 ⊢ 70	5	37
70 ⊢ 74	3	40
Σ	**40**	

$$d_i = l_{d_i} + \frac{\left(\frac{i.n}{10} - \sum f\right).h}{F_{d_i}}$$

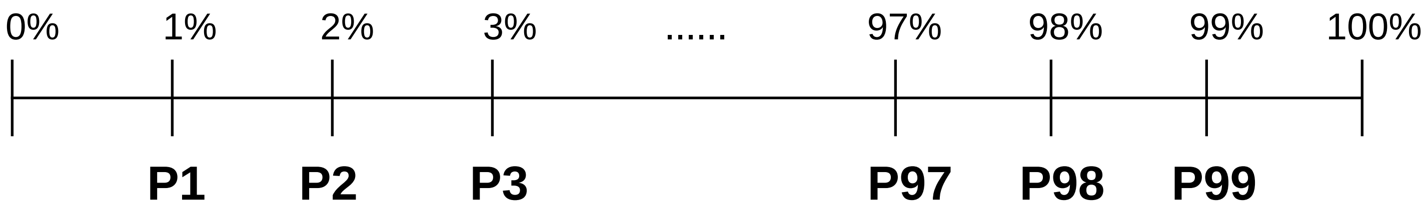
Calcular o 3º decil:

3*40/10 = 12 que corresponde a 2ª Classe

d3 = 54 + [(12 - 4) x 4] / 9 = 54 + 3,55 = **57,55**

Percentis

- **Dividem uma série em 100 partes iguais**



- **Notação adotada:** (P) para o parâmetro e (p) para a estimativa.

Percentis

$$p_i = l_{p_i} + \frac{\left(\frac{i.n}{100} - \sum f\right).h}{F_{p_i}}$$

- l_{p_i} = limite inferior da classe p_i (i = 1, ..., 99)
- i = 1 para p_1, ..., 99 para p_{99}
- n = tamanho da série
- Σ f = soma das frequências anteriores à classe $\mathrm{p}d_i$
- h = amplitude da classe p_i
- F_{p_i} = frequência da classe p_i

Percentis

- **Exemplo**

Classe	Fi	Fac
50 ⊢ 54	4	4
54 ⊢ 58	9	13
58 ⊢ 62	11	24
62 ⊢ 66	8	32
66 ⊢ 70	5	37
70 ⊢ 74	3	40
Σ	**40**	

$$p_i = l_{p_i} + \frac{\left(\frac{i.n}{100} - \sum f\right).h}{F_{p_i}}$$

Calcular o 8º centil:

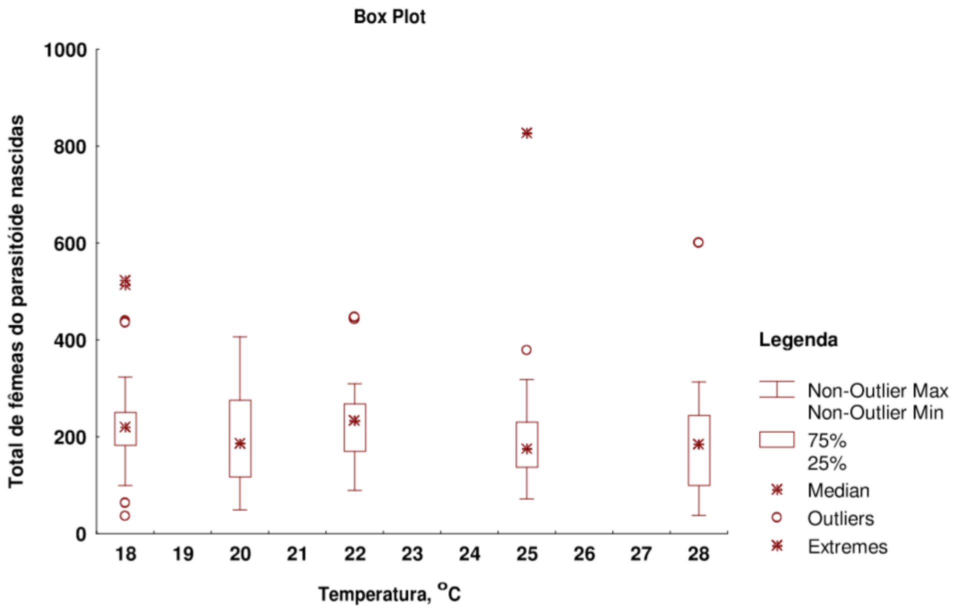
8*40/100 = 3,2 que corresponde a 1ª Classe

p8 = 50 + [(3,2 - 0) x 4] / 4 = 50 + 3,2 =

53,2

Medidas de posição ou separatrizes

- **Situações de uso mais comum dessas medidas**
  - Diagrama de caixa ("box plot")



- Diagrama de caixa do total de fêmeas do parasitóide nascidas

Probabilidade e estatística

Medidas Estatísticas

Tendência central, Posição ou separatrizes

UESC – CiC – 2016.2

Graduandos: Eberty Alves e Felipe Oliveira
Professor: José Cláudio Faria