

# Disciplina: 5170 - Probabilidade e Estatística

## Aula 04

**Prof. George Lucas M. Pezzott**  
**glmpezzott2@uem.br**

Departamento de Estatística - UEM  
Sala: 222 - Bloco: E-90

Segunda-feira - 19:30 ~ 21:10 - D67 - Sala 108

Sexta-feira - 21:20 ~ 23:00 - D67 - Sala 108

## Apresentação dos Dados

## O que fazer com os dados coletados?

- **Etapla inicial:** Estatística Descritiva (Análise Exploratória): é o ramo que trata da organização, resumo e apresentação dos dados. Ela consiste em calcular medidas descritivas, gráficos e tabelas.

# Tabelas e gráficos

# Organização dos Dados

- **Objetivo principal:**

- Facilitar a leitura dos dados.
- Compreensão dos dados.
- Obter informações sobre a população.
- Fazer comparações.

# Tabela de frequências

- Uma das formas de organizar e **resumir** a informação contida em dados observados é por meio de **tabelas de frequências**.
- A **frequência** de um valor da variável é o número de vezes que este valor ocorre no conjunto de dados.
- **Tabela de frequências** é uma tabela com diferentes valores de uma variável (ou **intervalo** de valores) e suas respectivas frequências ( $f_i$ )

# Configuração de gráficos e tabelas

- **Título:** colocado na **parte superior da tabela** e deve ser completo respondendo as questões:
  - O que se está estudando?
  - Onde foi feito o estudo?
  - Quando este estudo foi realizado?\*\*\*
- **Fonte:** se a fonte dos dados não é do próprio autor, ela deve ser indicada abaixo da tabela.
- **Notas:** Utilizadas para fazer esclarecimentos gerais. São numeradas e colocadas abaixo da tabela.
- **Valores totais:** os totais e subtotais devem ser bem destacados em tabelas.

# Configuração de gráficos e tabelas

- **Valores decimais:** usar a mesma quantidade de casas decimais para os algarismos em tabelas.
- **Margens:** as tabelas não devem ser fechadas lateralmente.
- **Valores ausentes em tabelas:**
  - Hífen (-) valor nulo pela própria natureza do fenômeno
  - Reticências (...) não se tem informação sobre o valor.
  - Interrogação (?) existe dúvida sobre a exatidão da informação.

## Organização dos Dados

## Representação dos Dados

- Representação na forma de tabelas:
  - Tabela de Frequência Simples;
  - Tabela de Frequência por classes.



- Quando observamos as respostas referentes a uma variável, podemos verificar que vários valores aparecem mais de uma vez.
- Chamamos de **frequência absoluta** ( $f_i$ ) o número de vezes que determinado valor aparece na série.
- O número total de elementos da amostra  $n$  (tamanho da amostra) pode ser obtido por  $n = \sum_{i=1}^k f_i$ , em que  $k$  é o número de valores diferentes na amostra (“classes”).

- **Provedor usado pelos visitantes do site [www.exemplo.com](http://www.exemplo.com) no dia 14/04/2019 (50 observações), em ordem de coleta:**

---

A	A	A	A	A	A	C	B	A	A
A	A	A	B	A	A	A	A	A	A
B	B	B	A	A	A	B	C	B	B
A	A	B	A	A	B	B	B	B	C
A	A	B	A	A	A	B	B	A	A

---

**Tabela de Frequência Simples:**

Tabela: Distribuição de frequência do provedor usado pelo visitante do site [www.exemplo.com](http://www.exemplo.com) no dia 14/04/2019.

Provedor	Frequência ( $f_i$ )
A	31
B	16
C	2
D	1
Total	50

Fonte: Elaborada pelo professor.

**Outras medidas a serem incluídas em uma tabela:**

- Frequência relativa ( $fr_i$ ):

$$fr_i = \frac{f_i}{\sum_{i=1}^k f_i} = \frac{f_i}{n} \quad \text{ou} \quad fr_i(\%) = \frac{f_i}{n} \times 100.$$

- Frequência acumulada ( $F_i$ ).

$F_i$  = Número observações “menores ou iguais” à classe  $i$ .

- Frequência acumulada relativa ( $Fr_i$ ).

$$Fr_i = \frac{F_i}{n} \quad \text{ou} \quad Fr_i(\%) = \frac{F_i}{n} \times 100.$$

Tabela: Distribuição de frequência do provedor usado pelo visitante do site [www.exemplo.com](http://www.exemplo.com) no dia 14/04/2019.

Provedor	Frequências Simples		Frequências Acumuladas	
	Absoluta	Relativa	Absoluta	Relativa
	$f_i$	$fr_i(\%)$	$F_i$	$F_{r_i}(\%)$
A	31	62	31	62
B	16	32	47	94
C	2	4	49	98
D	1	2	50	100
Total	50	1,00	-	-

Fonte: Elaborada pelo professor

## Qual a necessidade de construir uma distribuição de frequência por classes?

- **Exemplo: Tempos (em segundos) para carga de um aplicativo num sistema compartilhado (60 observações).:**

37	48	53	20	28	21	47	26	37	27
40	62	38	34	73	31	54	56	49	74
51	35	52	34	40	63	68	35	71	31
70	21	66	17	29	72	52	63	31	62
23	42	58	26	65	50	75	48	65	64
18	17	39	67	37	72	23	29	26	45

## Organização dos Dados

## Representação Tabular

Tabela: Distribuição de frequências dos tempos (em segundos) para carga de um aplicativo num sistema compartilhado .

Tempo	Freq.	Tempo	Freq.	Tempo	Freq.	Tempo	Freq.	Tempo	Freq.
11	2	24	1	40	2	58	2	68	2
12	2	25	2	43	1	59	1	70	1
13	1	27	1	46	1	60	1		
14	1	28	2	47	1	61	2		
15	2	29	1	50	1	62	2		
17	1	31	2	51	1	63	1		
18	1	33	1	52	1	64	3		
20	1	34	2	53	1	65	1		
21	1	36	3	54	2	66	1		
22	1	39	1	55	1	67	2		

## Para construir uma tabela de frequência por classes:

- **Passo 1:** Ordenação dos dados  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ ;
- **Passo 2:** Calcular a amplitude do Rol:  $A = x_{(n)} - x_{(1)}$ ;
- **Passo 3:** Calcular o número de classes (aproximadamente)  $k$ ;
- **Passo 4:** Calcular a amplitude das classes  $A_c \approx A/k$ ;
- **Passo 5:** Calcular os limites inferior (LI) e superior (LS) de cada classe;
- **Passo 6:** Contagem das observações dentro de cada classe.



- **Número de classes:**

- Regra de Sturges:

$$k \approx 1 + 3,3 \log_{10}(n)$$

- Regra do quadrado:

$$k = \sqrt{n}.$$

**Ob.:**  $k$  ideal deve estar entre 5 e 20. Isso acarreta em desvantagem na regra do quadrado quando  $n$  é muito grande.

## Tabela de Frequência por Classes:

Tabela: Distribuição de frequências dos tempos (em segundos) para carga de um aplicativo num sistema compartilhado.

Tempos Classes $c_i$	Freq. Absoluta $f_i$
11 ┤ 21    [11, 21)	11
21 ┤ 31    [21, 31)	9
31 ┤ 41    [31, 41)	11
41 ┤ 51    [41, 51)	4
51 ┤ 61    [51, 61)	10
61 ┤ 71    [61, 71)	15
Total	60

Fonte: Elaborada pelo autor.

## Organização dos Dados

## Representação Tabular

Tabela: Distribuição de frequências dos tempos (em segundos) para carga de um aplicativo num sistema compartilhado.

Tempos (Classes) $c_i$	Frequências				Ponto Médio $Pm_i$
	Absoluta $f_i$	Relativa $fr_i$	Acumulada $F_i$	Acum. Relativa $Fr_i$	
11 ┤ 21	11	0,1833	11	0,1833	16
21 ┤ 31	9	0,1500	20	0,3333	26
31 ┤ 41	11	0,1833	31	0,5166	36
41 ┤ 51	4	0,0667	35	0,5833	46
51 ┤ 61	10	0,1667	45	0,7500	56
61 ┤ 71	15	0,2500	60	1,0000	66
Total	60	1,0000	-	-	-

Fonte: Elaborada pelo autor

## Medidas Resumo × Tabela de Frequência Simples

- Média:

$$\bar{x} = \frac{c_1 f_1 + c_2 f_2 + \dots + c_k f_k}{f_1 + \dots + f_k} = \sum_{i=1}^k \frac{c_i f_i}{n},$$

em que  $n = \sum_{i=1}^k f_i$ .

- Variância:

$$s^2 = \sum_{i=1}^k \frac{f_i (c_i - \bar{x})^2}{n - 1}$$

**Tabela de Frequência Simples:**

Tabela: Número de acidentes diários em uma rodovia X no mês de abril/2019.

Acidentes	Frequência ( $f_i$ )
0	13
1	10
2	5
3	2
Total	30

Fonte: Elaborada pelo professor.

## Medidas Resumo × Tabela de Frequência por Classes

- Média:

$$\bar{X} = \frac{Pm_1 f_1 + Pm_2 f_2 + \dots + Pm_k f_k}{f_1 + \dots + f_k} = \sum_{i=1}^k \frac{Pm_i f_i}{n},$$

em que  $n = \sum_{i=1}^k f_i$ .

- Variância:

$$s^2 = \sum_{i=1}^k \frac{f_i (Pm_i - \bar{X})^2}{n - 1}$$

## Organização dos Dados

## Representação Tabular

Tabela: Distribuição de frequências dos tempos (em segundos) para carga de um aplicativo num sistema compartilhado.

Tempos (Classes) $c_i$	Frequências				Ponto Médio $Pm_i$
	Absoluta $f_i$	Relativa $fr_i$	Acumulada $F_i$	Acum. Relativa $Fr_i$	
11 ┤ 21	11	0,1833	11	0,1833	16
21 ┤ 31	9	0,1500	20	0,3333	26
31 ┤ 41	11	0,1833	31	0,5166	36
41 ┤ 51	4	0,0667	35	0,5833	46
51 ┤ 61	10	0,1667	45	0,7500	56
61 ┤ 71	15	0,2500	60	1,0000	66
Total	60	1,0000	-	-	-

Fonte: Elaborada pelo autor

- Quantis:

- Encontrar a posição:  $P = p \cdot n$ ,  $0 < p < 1$ .
- Identificar a classe que contém a  $P$ -ésima observação:  $j$ ,  $1 \leq j \leq k$ .

$$q(p) = \frac{Ac_j(p - Fr_{\{j-1\}})}{fr_j} + Ll_j \quad \text{ou} \quad q(p) = \frac{Ac_j(P - F_{\{j-1\}})}{f_j} + Ll_j$$

em que  $Ac_j = LS_j - Ll_j$ .