

# Bioestatística

Cláudio Roberto Thiersch  
Monica Fabiana Bento Moreira Thiersch

Monitor: Lucas Kröhling Bernardi

PPGPUR - Programa de Pós-Graduação em  
Planejamento e Uso de Recursos Renováveis

Universidade Federal de São Carlos - Campus Sorocaba

# Introdução

## O que é estatística?

"A estatística é um conjunto de métodos e técnicas que auxiliam a tomada de decisão sob a presença de incerteza."

## Estatística

"A estatística é um conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados em qualquer área de conhecimento."

## Áreas da estatística:

- Estatística descritiva  
É a etapa inicial da análise utilizada para descrever e resumir os dados.
- Probabilidade  
A teoria de probabilidades nos permite descrever os fenômenos aleatórios, ou seja, aqueles em que está presente a incerteza.
- Inferência Estatística  
É o estudo de técnicas que possibilitam a extrapolação, a um grande conjunto de dados, das informações e conclusões obtidas a partir da **amostra**.

## População

É o conjunto de elementos com pelo menos uma característica em comum. Essas características comuns devem delimitar inequivocamente quais os elementos que pertencem a população e quais os que não pertencem.

## Amostra

Qualquer subconjunto da população. É necessariamente finita pois todos os elementos serão examinados para efeito da realização do estudo estatístico desejado.



## Variável

Característica pela qual deseja-se que a população seja descrita.

## Tipos de variáveis

- Variáveis qualitativas: Correspondem a atributos ou categorias.
  - Nominal: Os atributos não são passíveis de ordenação.
  - Ordinal: Os atributos são passíveis de ordenação.
- Variáveis quantitativas: Correspondem a números resultantes de contagem ou medidas
  - Discreta: São próprias dos dados de contagem.
  - Contínua: As observações são resultantes de uma medida que pode assumir qualquer valor real entre os dois extremos.

## Dado ou observação

É o valor que assume a variável para um elemento em particular.

## Notação

- Tamanho da população (**N**): Número de elementos da população.
- Tamanho da amostra (**n**): Número de elementos da amostra.
- Variável (**X, Y, Z, etc**): São letras maiúsculas e geralmente as últimas do alfabeto.

## Técnicas de somatório

- Notação por índice

O símbolo  $X_j$  (leia-se X índice "j") representa qualquer um dos "n" valores de  $X_1, X_2, \dots, X_n$  assumidos por uma variável  $X$ .

- Notação de somatório

$\sum$  é a letra grega sigma maiúscula que é utilizada para representar a soma de todos os valores de  $X_j$ , desde  $X_1$  até  $X_n$ .

$$\sum_{j=1}^n X_j = X_1 + X_2 + \dots + X_n$$



## Técnicas de somatório

- Propriedades ( $a$ ,  $b$  e  $k$  são constantes e  $X$  e  $Y$  variáveis)
  - $\sum_{j=1}^n aX_j = aX_1 + aX_2 + \dots + aX_n = a \sum_{j=1}^n X_j$
  - $\sum_{j=1}^n X_j Y_j = X_1 Y_1 + X_2 Y_2 + \dots + X_n Y_n$
  - $\sum_{j=1}^n (aX_j + bY_j) = a \sum_{j=1}^n X_j + b \sum_{j=1}^n Y_j$
  - $\sum_{j=1}^n k = nk$

## Estatística descritiva

Preocupa-se com a organização, apresentação e análise dos dados quantitativos ou qualitativos, sejam eles amostrais ou populacionais, utilizando alguma representação que pode ser tabular ou gráfica.

### • Dados brutos

Tabela 1: Informações sobre estado civil, grau de instrução, número de filhos, salário (expresso como fração do salário mínimo), idade (medida em anos e meses) e procedência de 36 empregados da seção de orçamentos da Companhia MB.

Nº	Estado civil	Grau de instrução	Nº de filhos	Salário (× sal. mín.)	Idade		Região de procedência
					anos	meses	
1	Solteiro	ensino fundamental	—	4	26	3	Interior
2	Casado	ensino fundamental	1	4,56	32	10	Capital
3	Casado	ensino fundamental	2	5,25	36	5	Capital
4	Solteiro	ensino médio	—	5,73	21	10	Outra
5	Solteiro	ensino fundamental	—	6,26	41	7	Outra
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
36	Casado	superior	3	23,3	42	2	Interior

Fonte: Bussab e Morettin (2002)

- **Dados agrupados**

Tabela 2: Frequências e porcentagens dos 36 empregados da seção de orçamentos da Companhia MB segundo o grau de instrução

Grau de instrução	Frequência ( $n_i$ )	Proporção ( $fr_i$ )	Porcentagem ( $100fr_i$ )
Fundamental	12	0,3333	33,33
Médio	18	0,5000	50,00
Superior	6	0,1667	16,67

Fonte: Tabela 1 - Bussab e Morettin(2002)

Tabela 3: Frequências e porcentagens dos 36 empregados da seção de orçamentos da Companhia MB por faixa de salário

Classe de salários	Frequência ( $n_i$ )	Proporção ( $fr_i$ )	Porcentagem ( $100fr_i$ )
[ 4, 00; 8, 00)	10	0,2778	27,78
[ 8, 00; 12, 00)	12	0,3333	33,33
[12, 00; 16, 00)	8	0,2222	22,22
[16, 00; 20, 00)	5	0,1389	13,89
[20, 00; 24, 00)	1	0,0278	2,78
Total	36	1,0000	100,00

Fonte: Tabela 1 - Bussab e Morettin(2002)

- Gráficos para variáveis qualitativas

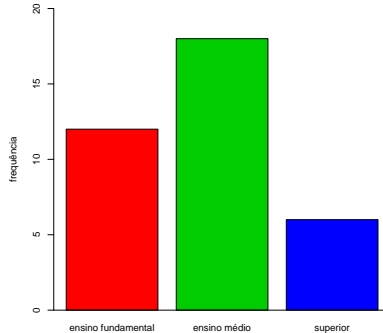


Figura 1: Gráfico em barras para a variável grau de instrução

Figura 2: Gráfico em setores para a variável grau de instrução

- Gráficos para variáveis quantitativas

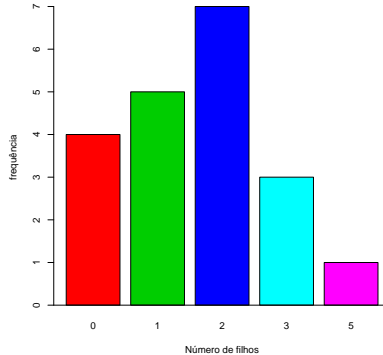


Figura 3: Gráfico em barras para a variável número de filhos

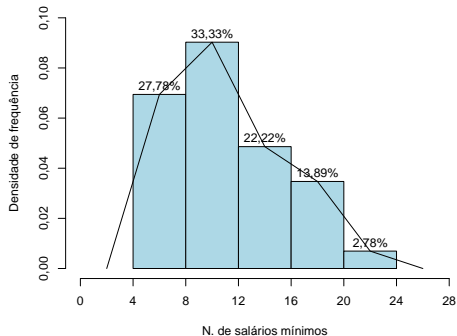


Figura 4: Histograma e polígono de frequências da variável salários



## Medidas de posição e variabilidade

Tipo da medida	Medida	Simbologia	
		Amostra (n)	População (N)
Posição	Média	$\bar{x}$	$\mu$
	Mediana	md	Md
	Moda	mo	Mo
Variabilidade	Variância	$s^2$	$\sigma^2$
	Desvio Padrão	s	$\sigma$
	Coeficiente de Variação	cv	CV

# Medidas de posição

- Média Aritmética

Se  $x_1, \dots, x_n$  são os  $n$  valores da variável  $X$ , a média de  $X$  pode ser escrita como:

$$\bar{x} = \frac{x_1 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

Se  $f_i = \frac{n_i}{n}$  representar a frequência relativa da observação  $x_i$ , então a média pode ser escrita como

$$\bar{x} = \sum_{i=1}^k f_i x_i$$

- a) Calcular a média aritmética dos valores: 82, 89, 90, 102.
- b) Calcule o número médio de filhos que os funcionários da empresa MB possuem.

Número de filhos	Frequência ( $n_i$ )
0	4
1	5
2	7
3	3
5	1
Total	20

- Mediana

Mediana de um conjunto de dados é o valor ao qual metade dos dados são iguais ou inferiores e metade dos dados são iguais ou superiores.

Ex.: Encontre a mediana para os dados abaixo:

i) 3,4,10,6,7

ii) 3,11,7,9,4,10

- Moda

Moda de um conjunto de dados é o valor que ocorre com maior frequência.

Ex.: Encontre a moda para os dados abaixo:

i) 3,4,7,7,9

ii) 3,3,5,7,10,10,11

iii) 1,2,7,10

- Média Ponderada

Dado um conjunto de  $n$  valores observados  $(x_1, \dots, x_n)$  e conhecidos os respectivos fatores de ponderação  $(q_1, \dots, q_n)$ , a média ponderada de  $x$  é

$$w = \frac{\sum_{i=1}^n x_i q_i}{\sum_{i=1}^n q_i}$$

Ex.: Calcular a média final da disciplina Estatística dado que os pesos das 3 avaliações são 25%, 45% e 30%. Suponhamos que as notas de um determinado aluno tenha sido: 80, 30, 50, respectivamente. Qual foi o resultado deste aluno?

# Medidas de dispersão

- Variância

A variância amostral de um conjunto de dados,  $x_1, x_2, \dots, x_n$  é definida por

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

e para dados de frequência

$$S^2 = \sum_{i=1}^k f_i (x_i - \bar{x})^2$$

- Desvio Padrão

$$S = \sqrt{\text{Variância}} = \sqrt{S^2}$$

- Coeficiente de Variação

$$CV(\%) = \frac{S}{\bar{x}} 100$$

Ex.: Calcule a variância, o desvio padrão e o coeficiente de variação dos dados abaixo:

i) 3, 4, 6, 7, 10

ii) -20, 5, 15, 24.

- Percentil

O  $p$ -ésimo percentil é um valor tal que pelo menos  $p$  por cento das observações são menores ou iguais a esse valor e pelo menos  $(100 - p)$  por cento das observações são maiores ou iguais a esse valor.

Para calcular o  $p$ -ésimo percentil temos:

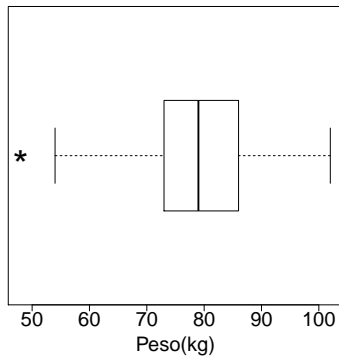
- 1) Organize os dados em ordem crescente;
- 2) Calcule um índice  $i$ ,  $i = \frac{p}{100}n$ , onde  $p$  é o percentil procurado e  $n$  é o número de observações.
- 3) Se  $i$  for um número inteiro, o  $p$ -ésimo percentil será a média dos valores que ocupam as posições  $i$  e  $i + 1$ . Caso contrário, arredonde  $i$  para cima.



Ex.: Determinar o 85º percentil e a mediana dos dados de salários iniciais:

3450, 3550, 3650, 3480, 3355, 3310, 3490, 3730, 3540, 3925, 3520, 3480.

# Box plot



## Box plot (Desenho Esquemático)

Para a construção do gráfico são necessárias as estatísticas:  
Mediana,  $Q_1$ ,  $Q_3$  e amplitude interquartil dada por  $A/Q = Q_3 - Q_1$ .

Passos:

- 1) Desenhe um retângulo em que suas extremidades são  $Q_1$  e  $Q_3$ ;
- 2) Desenhe uma linha no retângulo na posição da mediana;
- 3) Desenhe os **limites**. Eles estão a  $1,5A/Q$  abaixo de  $Q_1$  e  $1,5A/Q$  acima de  $Q_3$ . Dados fora desses limites são considerados dados fora da curva.
- 4) Desenhe as **costeletas**;
- 5) Desenhe todos os pontos fora da curva utilizando um símbolo \*.

# Box plots - Exemplo

Tabela 4: Salários mensais iniciais de uma amostra de 12 graduados da escola de administração.

Graduado	Salário (US\$)	Graduado	Salário (US\$)
1	2850	7	2890
2	2950	8	3130
3	3050	9	2940
4	2880	10	3325
5	2755	11	2920
6	2710	12	2880

## Bibliografia

ANDERSON, D. R., SWEENEY, D. J. e WILLIAMS, T. A.  
**Estatística Aplicada à Administração e Economia**. São Paulo,  
Pioneira Thomson Learning, 2002.

BUSSAB, W. O. ; MORETTIN, P. A. **Estatística Básica**. Editora  
Saraiva, 2007.

COCHRAN, W.G. **Técnicas de Amostragem (Tradução)**. Ed.  
Fundo de Cultura. Rio de Janeiro. 555p.1965.

COSTA NETO, P. L. de O., 1939-. **Estatística**. São Paulo:  
Edgard Blücher, 2005. 266 p.

KISH, L. **Survey sampling**. Wiley, Nova York. 1965.