

Estatística Descritiva I

Ciências Contábeis - FEA - Noturno

1º Semestre 2025

Profs. Leonardo T. Rolla e Nikolai Kolev

(baseado em material previamente desenvolvido
pelo Prof. Gilberto Alvarenga Paula – CC BY-SA 4.0)

- 1 Informações sobre esta unidade
- 2 O que é Estatística
- 3 Variável
- 4 Distribuição de frequências
- 5 Gráficos
- 6 Medidas de posição
- 7 Medidas de dispersão
- 8 Exercícios

Compreender os conceitos fundamentais da Estatística, incluindo tipos de variáveis, distribuição de frequências, representações gráficas e medidas resumo, para descrever e interpretar conjuntos de dados de forma eficaz.

- 1 Introdução à Estatística: definição e aplicações
- 2 Tipos de variáveis: qualitativas e quantitativas
- 3 Distribuição de frequências: construção e interpretação
- 4 Representação gráfica de dados: tabelas, histogramas e gráficos de setores
- 5 Medidas de posição: média, mediana, moda e percentis
- 6 Medidas de dispersão: intervalo interquartil, desvio médio, variância e desvio padrão

Depois desta unidade, o aluno será capaz de:

- 1 Compreender o papel da Estatística na análise de dados.
- 2 Diferenciar tipos de variáveis e classificá-las corretamente.
- 3 Construir e interpretar distribuições de frequências.
- 4 Utilizar representações gráficas adequadas para descrever dados.
- 5 Calcular e interpretar medidas de posição e dispersão.

O aluno pode fazer os exercícios postados no Moodle e posteriormente comparar suas respostas com as soluções apresentadas, identificando possíveis erros, e eventualmente refazer os exercícios.

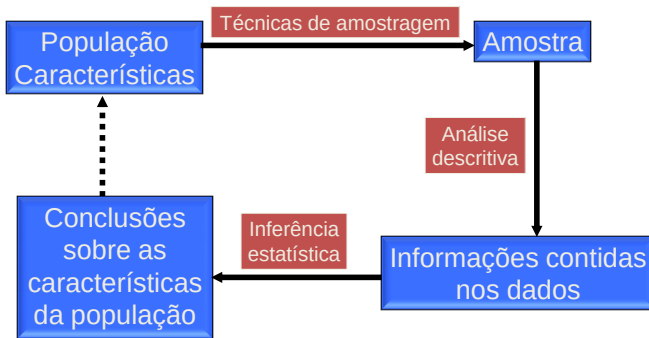
Além disso, é recomendável refletir sobre as dificuldades encontradas e revisar os conceitos conforme necessário.

- 1 Informações sobre esta unidade
- 2 O que é Estatística**
- 3 Variável
- 4 Distribuição de frequências
- 5 Gráficos
- 6 Medidas de posição
- 7 Medidas de dispersão
- 8 Exercícios

Estatística

Metodologia para obtenção, organização e análise de dados oriundos das mais variadas áreas do conhecimento, visando auxiliar na **tomada de decisões em situações de incerteza**.

Etapas da Análise Estatística



1

O que é Estatística

Amostragem

Técnicas para a obtenção de amostras convenientemente extraídas da população de interesse.

Estatística Descritiva

Etapa inicial da análise utilizada para descrever, organizar e resumir os dados coletados.

Probabilidade

Auxilia na modelagem de fenômenos aleatórios.
É uma ferramenta fundamental para a inferência estatística.

Inferência Estatística

Conjunto de técnicas que permite, a partir de dados amostrais, tirar conclusões sobre a população de interesse, controlando erros.

- 1 Informações sobre esta unidade
- 2 O que é Estatística
- 3 Variável**
- 4 Distribuição de frequências
- 5 Gráficos
- 6 Medidas de posição
- 7 Medidas de dispersão
- 8 Exercícios

Definição

Defini-se como variável qualquer característica associada a uma população. Há dois tipos de variável: **Qualitativa** e **Quantitativa**.

Variável Qualitativa

- **Nominal**: gênero, estado civil, cor dos olhos
- **Ordinal**: classe social, grau de instrução, faixa etária

Variável Quantitativa

- **Contínua**: peso, altura, salário, temperatura
- **Discreta**: número de filhos, número de sinistros

- 1 Informações sobre esta unidade
- 2 O que é Estatística
- 3 Variável
- 4 Distribuição de frequências**
- 5 Gráficos
- 6 Medidas de posição
- 7 Medidas de dispersão
- 8 Exercícios

Nível de escolaridade

Classe	n_i	f_i	%
Fundamental	12	0,333	33,3
Médio	18	0,500	50,0
Superior	6	0,167	16,7
Total	36	1,000	100,0

- Frequência absoluta
- Proporção
- Porcentagem

Variável quantitativa contínua

- Não há dados repetidos
- Agrupar dados
- Como? Resumir ou reter informação?

(também é necessário agrupar variáveis discretas que assumem muitos valores diferentes)

Exemplo: número de interjeições na fala

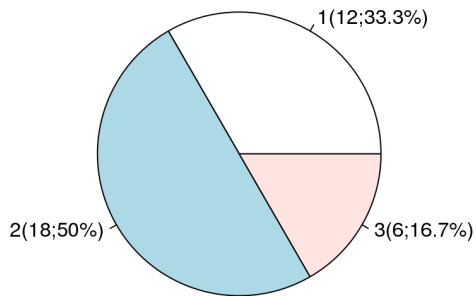
Frequências

Classe	n_i	f_i
0 ┊ 5	360	0,606
5 ┊ 10	165	0,278
10 ┊ 15	47	0,079
15 ┊ 20	17	0,029
20 ┊ 25	5	0,008
Total	594	1,000

- 1 Informações sobre esta unidade
- 2 O que é Estatística
- 3 Variável
- 4 Distribuição de frequências
- 5 Gráficos**
- 6 Medidas de posição
- 7 Medidas de dispersão
- 8 Exercícios

- Pizza
Ideal para variável qualitativa nominal
- Barras
Adequado para variável qualitativa em geral
- Barras contíguas
Ideal para variável qualitativa ordinal

Gráfico de pizza



1=Fundamental, 2=Médio, 3=Superior

Gráfico de barras

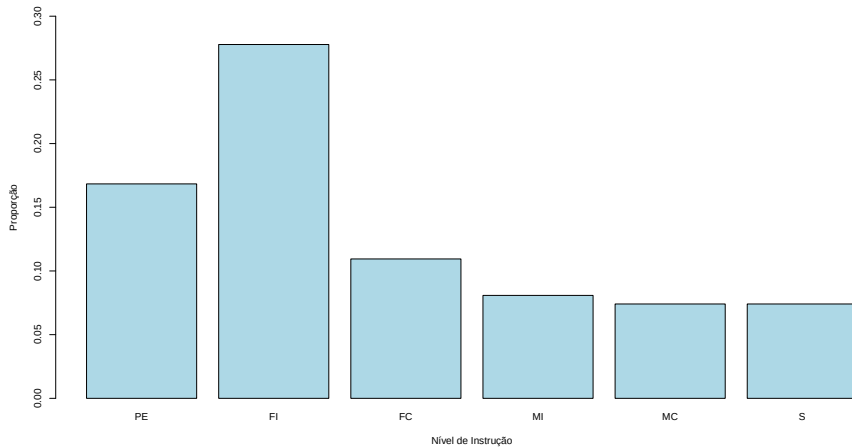
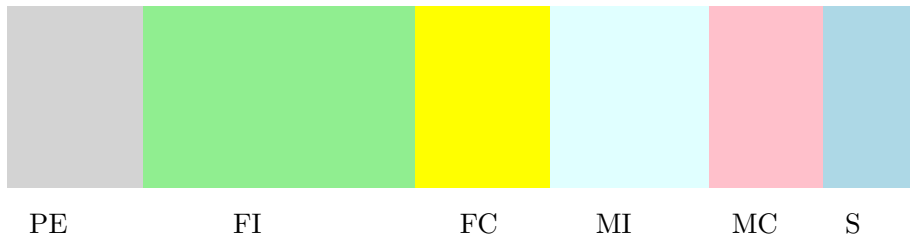


Gráfico de barras contíguas



Considerações

- Necessário agrupar dados
(exceto variáveis discretas com poucos valores)
- Pode-se usar o gráfico de barras
- Ideal usar o histograma

Histograma

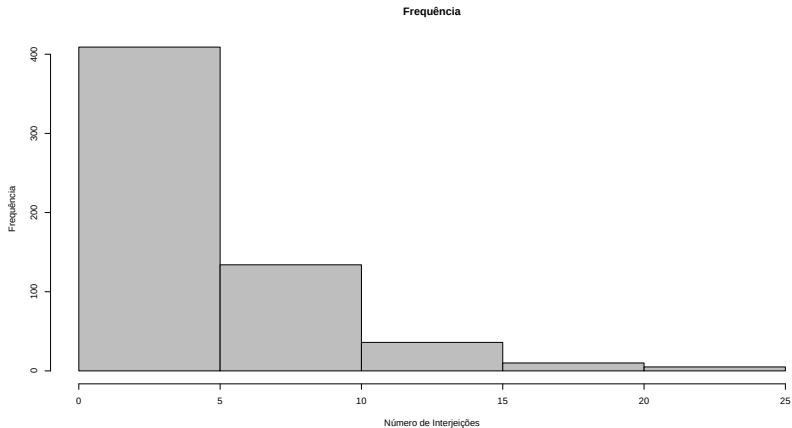
- Parecido ao gráfico de barras, porém mais versátil
- Não há um vão entre as barras

Exemplo: número de interjeições na fala

Frequências

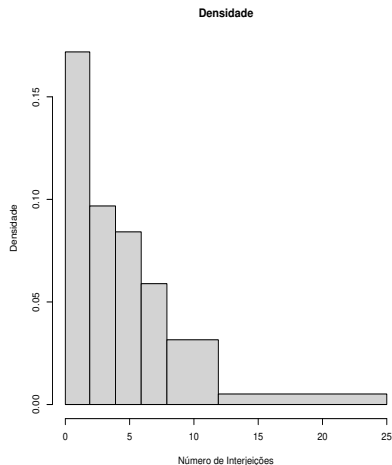
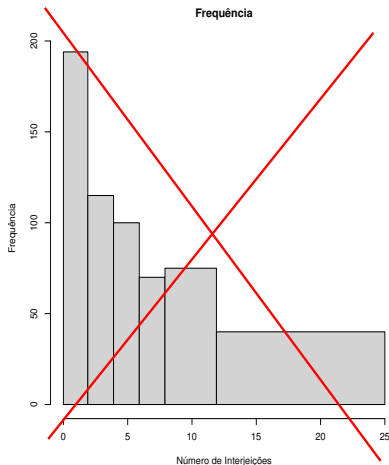
Classe	n_i	f_i
0 ┊ 5	360	0,606
5 ┊ 10	165	0,278
10 ┊ 15	47	0,079
15 ┊ 20	17	0,029
20 ┊ 25	5	0,008
Total	594	1,000

Exemplo: número de interjeições na fala



O gráfico é o mesmo se usamos frequência absoluta, proporção ou porcentagem, diferindo apenas na escala do eixo vertical.

Intervalos de tamanho variável



Frequência corresponde à **área** da barra e não à **altura**!

- 1 Informações sobre esta unidade
- 2 O que é Estatística
- 3 Variável
- 4 Distribuição de frequências
- 5 Gráficos
- 6 Medidas de posição**
- 7 Medidas de dispersão
- 8 Exercícios

Moda

Valor (ou atributo) que ocorre com maior frequência

Exemplo

Dados: 4, 5, 4, 6, 5, 8, 4.

Moda: $M_o = 4$

Média Amostral

Observando-se os valores $x_1, \dots, x_n$ para uma variável quantitativa, a média amostral (média aritmética) é definida por

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

Exemplo

Dados: 2, 5, 3, 7, 8.

$$\bar{x} = \frac{2 + 5 + 3 + 7 + 8}{5} = 5.$$

Notação de somatório

$\sum_{i=1}^n x_i$ é a soma de x_i com i variando de 1 a n .

Mediana

A mediana é o valor da variável que ocupa a posição central de um conjunto de **n dados ordenados**, ou seja, ocupa posição

$$k = \frac{n + 1}{2}.$$

Exemplo

Dados: **2, 3, 6, 7, 8**. Temos que $n = 5$, $k = \frac{5+1}{2} = 3$,

$$\text{Md} = 6.$$

Exemplo

Dados: **1, 2, 4, 6, 8, 9**. Temos que $n = 6$, $k = \frac{6+1}{2} = 3,5$,

$$\text{Md} = \frac{4 + 6}{2} = 5.$$

- 1 Informações sobre esta unidade
- 2 O que é Estatística
- 3 Variável
- 4 Distribuição de frequências
- 5 Gráficos
- 6 Medidas de posição
- 7 Medidas de dispersão**
- 8 Exercícios

Exemplo

Considere as notas de um teste de 3 grupos de alunos:

- Grupo 1: 3, 4, 5, 6, 7
- Grupo 2: 1, 3, 5, 7, 9
- Grupo 3: 5, 5, 5, 5, 5

Problema

- $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 5$
- $Md_1 = Md_2 = Md_3 = 5$

Medida de dispersão

Um valor que resuma a variabilidade de um conjunto de dados.

Percentil

O percentil de ordem $p \times 100$, em que $0 < p < 1$, em um conjunto de n dados ordenados, é o valor que ocupa a posição

$$k = p \times (n + 1).$$

Casos Particulares

- percentil 10 = primeiro decil
- percentil 25 = primeiro quartil (Q1)
- percentil 50 = mediana ou segundo quartil (Md ou Q2)
- percentil 75 = terceiro quartil (Q3)

Quartil Q1

O primeiro quartil (percentil 25) é o valor de um conjunto de **n dados ordenados** que ocupa a posição

$$0,25 \times (n + 1) = \frac{1}{4} \times (n + 1) = \frac{n + 1}{4}.$$

Quartil Q3

O terceiro quartil (percentil 75) é o valor de um conjunto de **n dados ordenados** que ocupa a posição

$$0,75 \times (n + 1) = \frac{3}{4} \times (n + 1) = \frac{3(n + 1)}{4}.$$

Exemplo

Dados

0,9 1,0 1,7 2,9 3,1 5,3 5,5 12,2 12,9 14,0 33,6. $n=11$.

Mediana

É o valor que está na posição $\frac{11+1}{2} = 6$. Portanto, $Md = 5,3$.

Quartil Q1

É o valor que está na posição $\frac{(11+1)}{4} = 3$. Portanto, $Q1 = 1,7$.

Quartil Q3

É o valor que está na posição $\frac{3(11+1)}{4} = 9$. Portanto, $Q3 = 12,9$.

Outro exemplo

Dados

1,9 2,0 2,1 2,5 3,0 3,1 3,3 3,7 6,1 7,7 (Temos $n = 10$)

Mediana

$$Md = (3,0 + 3,1)/2 = 3,05.$$

Quartil Q1

O primeiro quartil é o valor que está na posição $k = \frac{10+1}{4} = 2,75$. Tomaremos a média entre os valores nas posições 2 e 3. Assim,

$$Q1 = \frac{2,0 + 2,1}{2} = 2,05.$$

Quartil Q3

Analogamente, $k = \frac{3}{4} \times 11 = 8,25$, logo $Q3 = \frac{3,7+6,1}{2} = 4,9$.

Excel usa $k = p \times (n + 1)$

Google e R usam $k = p \times (n - 1) + 1$

Outra opção é $k = p \times n + \frac{1}{2}$

Pode-se arredondar k para o inteiro mais próximo, interpolar os dados, ou tirar a média simples. Google, Excel e R interpolam

Obs.: quando n é grande e p está longe de 0 e 1, esses métodos dão praticamente o mesmo resultado

Nesta disciplina, vamos tirar a média simples

Intervalo Interquartil

O intervalo interquartil é definido por

$$d_q = Q3 - Q1.$$

Exemplo

Do exemplo anterior: 1,9 2,0 2,1 2,5 3,0 3,1 3,3 3,7 6,1 7,7.

- $Q1 = 2,05$
- $Q3 = 4,9$
- $d_q = 4,9 - 2,05 = 2,85$

Desvio médio

Observando-se os valores $x_1, \dots, x_n$ para uma variável X quantitativa, o desvio médio é definido por

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

Cálculo do desvio médio para os grupos do início desta seção

Grupo 1

$$DM = \frac{|3-5| + |4-5| + |5-5| + |6-5| + |7-5|}{5} = \frac{6}{5} = 1,2.$$

Grupo 2

$$DM = \frac{|1-5| + |3-5| + |5-5| + |7-5| + |9-5|}{5} = \frac{12}{5} = 2,4.$$

Grupo 3

$$DM = \frac{|5-5| + |5-5| + |5-5| + |5-5| + |5-5|}{5} = 0.$$

Variância amostral

Observando-se os valores $x_1, \dots, x_n$ para uma variável X quantitativa, a variância amostral (mais utilizada) é definida por

$$s^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}.$$

A divisão da variância amostral por $(n - 1)$ leva a uma estimativa **não tendenciosa** da variância populacional.

Desvio Padrão Amostral

O desvio padrão amostral é definido por

$$s = \sqrt{s^2}.$$

Variância amostral dos grupos do início desta seção

Grupo 1

$$\frac{s^2 = (3-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2}{4} = \frac{10}{4} = 2,5$$
$$\Rightarrow s = \sqrt{2,5} \cong 1,58.$$

Grupo 2

$$\frac{s^2 = (1-5)^2 + (3-5)^2 + (5-5)^2 + (7-5)^2 + (9-5)^2}{4} = \frac{40}{4} = 10$$
$$\Rightarrow s = \sqrt{10} \cong 3,16.$$

Grupo 3

$$\frac{s^2 = (5-5)^2 + (5-5)^2 + (5-5)^2 + (5-5)^2 + (5-5)^2}{4} = 0$$
$$\Rightarrow s = \sqrt{0} = 0.$$

Fórmula prática para cálculo da variância

Fórmula prática para cálculo da variância

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{1}{n} \times (\sum_{i=1}^n x_i)^2}{n - 1}.$$

Exemplo

Do exemplo anterior (Grupo 1). Dados: 3, 4, 5, 6, 7

$$\sum x_i^2 = 3^2 + 4^2 + 5^2 + 6^2 + 7^2 = 9 + 16 + 25 + 36 + 49 = 135.$$

Além disso, $\sum_i x_i = 25$.

Portanto

$$s^2 = \frac{135 - \frac{1}{5} \times 25^2}{5 - 1} = \frac{10}{4} = 2,5.$$

Definição alternativa

Em alguns textos, usa-se uma fórmula ligeiramente diferente para variância amostral.

Definição alternativa

A definição abaixo também é utilizada para variância amostral:

$$\hat{\sigma}^2 = \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}.$$

Neste caso, o desvio padrão amostral é definido por

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}.$$

Aviso

Vamos usar a definição com $n - 1$ no denominador, e não com n

Coeficiente de variação

O coeficiente de variação é uma medida de dispersão relativa que elimina o efeito da magnitude dos dados e exprime a variabilidade em relação à média.

O coeficiente de variação amostral é definido por

$$CV = \frac{s}{\bar{x}} \times 100\%.$$

Exemplo

Altura (em cm) de uma amostra de recém-nascidos e de uma amostra de adolescentes

Grupo	Média	D. Padrão	CV
Recém-Nascidos	50	6	12%
Adolescentes	160	16	10%

Conclusão

Em relação às médias, as alturas dos adolescentes e dos recém-nascidos apresentam variabilidades muito parecidas.

- 1 Informações sobre esta unidade
- 2 O que é Estatística
- 3 Variável
- 4 Distribuição de frequências
- 5 Gráficos
- 6 Medidas de posição
- 7 Medidas de dispersão
- 8 Exercícios**

Exercício 1

Classifique cada uma das variáveis abaixo em qualitativa (nominal / ordinal) ou quantitativa (discreta / contínua):

- (a) Ocorrência de apneia em bebês internados numa UTI neonatal.
- (b) Número de quedas no sistema de computação de certa empresa durante o ano de 2020.
- (c) Tempo semanal despendido por funcionários para a atenção de reclamações (em horas).
- (d) Grau de satisfação dos paulistanos com relação ao trabalho do prefeito (valores de 0 a 5, com 0 indicando totalmente insatisfeito e 5 totalmente satisfeito).
- (e) Intensidade do uso de bebida alcoólica por universitários (baixa, moderada ou alta).
- (f) Variação do peso de atletas entre 35 e 40 anos com hipotireoidismo depois do uso por três meses de levotiroxina sódica (em kg).
- (g) Aumento percentual nas vendas de varejo durante o último ano.

Exercício 2

Os salários (em número de salários mínimos) de cinco funcionários do departamento pessoal de uma certa firma são:

2 4 7 3 4

- (a) Calcule a média, a variância e o desvio padrão dos salários;
- (b) Se for concedido um abono de 2 salários mínimos para todos os 5 funcionários, haverá alteração na média? E na variância? E no desvio padrão? Justifique.
- (c) Se for concedido um aumento de 100% no salário dos 5 funcionários, haverá alteração na média? E na variância? E no desvio padrão? Justifique.

Exercício 3

Uma agência de publicidade deseja verificar o grau de satisfação de seus clientes. Para tanto, escolheu aleatoriamente clientes que fizeram uso da agência classificando-os como sendo de tipo A, B e C e solicitou que um questionário fosse preenchido pela pessoa responsável do setor. Os questionários foram devidamente codificados, a fim de fornecer um índice de satisfação que varia de 1 (totalmente insatisfeito) a 5 (totalmente satisfeito). A tabela abaixo apresenta um resumo dos resultados.

Classe	A	B	C
Média	3,8	2,4	3,4
Mediana	3,8	2,6	3,9
Desvio Padrão	0,3	0,3	1,4

A que conclusão podemos chegar?