

Princípios básicos da Análise Exploratória de Dados UFCD: 10807

CARGA HORÁRIA: 25 horas

ALEXANDRA CAMPOS VIDAL DE SOUZA
FORMADORA



Conteúdos

Objetivos e tarefas da estatística e análise de dados

Mentiras, ambiguidades e más utilizações da estatística

Variáveis estatísticas e tipos de dados

Métodos de análise de dados para uma variável

Distribuições de frequências e histogramas

Medidas de tendência central

Medidas de dispersão

Medidas de localização

Outros indicadores

Métodos de análise de dados para duas variáveis

Covariância e correlação

R de Spearman

Medidas de concentração

Números índice

Princípios básicos sobre probabilidades

Princípios básicos sobre amostragem e metodologia de recolha de dados

Princípios da análise de componentes principais

Projeto de análise exploratória de dados



Conceitos básicos: População, Amostra e Unidade Amostral

- **População:** É qualquer conjunto, não necessariamente de pessoas, que constituem todo o universo de informações de que se necessita.
- **Amostra:** corresponde a um grupo representativo da população
- **Unidade Amostral:** corresponde ao indivíduo (ao elemento) da população amostral sobre o qual a medida interessa e será observada.



População



A Venn diagram consisting of three concentric circles. The outermost circle is green and labeled 'População'. Inside it is a blue circle labeled 'Amostra'. Inside the blue circle is a yellow circle labeled 'UC'.

Amostra

UC



Aplicação prática

1. Considere uma população de 70 formandos de uma turma que compareceram a aula em um determinado dia. Suponha que foram selecionados aleatoriamente 14 formandos para participarem de uma pesquisa de satisfação sobre uma determinada aula.
2. Considere ainda que o aluno “Tiago” participou da pesquisa supracitada.

Com base nessa situação hipotética. Temos que:

- Os 70 formandos que compareceram a aula representam: População
- Os 14 alunos selecionados para participar da pesquisa representam: Amostra
- E o Tiago representa: UC



Distribuição de frequencia

Dados Estatísticos são todas as informações levantadas coletadas que servirão como base para estudo e análise.

E Dados brutos, o que são?



Dados brutos são dados coletados que não foram organizados, podem ser coletados de um CRM, ERP, pode ser de uma infinidade de origens.



Exemplo prático

- Escores de superação de pacientes de acidente vascular cerebral e ataque cardíaco.

Acidente Vascular Cerebral		Ataque Cardíaco	
39	27	27	27
26	1	29	23
26	25	27	26
9	23	27	35
14	23	27	35
28	40	22	32
21	9	29	32
26	13	23	22
23	13	29	25
18	21	30	30

**COMO VOCÊ
DESCREVERIA OS
DADOS A UM AMIGO
QUE NÃO PODE VÊ-
LOS?**



Distribuição de frequencia

- **ROL:** Arranjo de dados brutos em ortdem crescente ou decrescente.
- Exemplo:
3, 2, 5, 0, 1, 9, 10 ,3 (dados brutos)
0, 1, 2, 3, 3, 5, 9, 10 (rol)



O que é Estatística

Estatística é a ciência que coleta, analisa e interpreta dados, utilizando métodos para calcular e extrair insights que auxiliam na compreensão das informações.



A Estatística pode ser dividida em tres áreas principais





Univariada

Analisa cada variável de forma independente, uma única variável por vez.



Bivariada

Analisa duas variáveis ao mesmo tempo, buscando relacionamento entre elas.

Estatística Descritiva

Multivariada

Trabalha com três ou mais variáveis. Avaliar como esse conjunto de variáveis culminam num desfecho.



A Estatística descritiva pode ser dividida em tres grupos: Univariada, Bivariada e Multivariada.



O que é estatística descritiva

Ela está voltada a

**Organizar, Resumir, e apresentar os dados
(tabelas, gráficos e medidas-resumos)**

**Compõe a análise exploratória de dados, com o objetivo de
observar determinados aspectos relevantes e começar a
delinear suas hipóteses**

**Não há
conclusões
nessa etapa**

**Observar alguns
aspectos
importantes e
delinear algumas
hipóteses.**

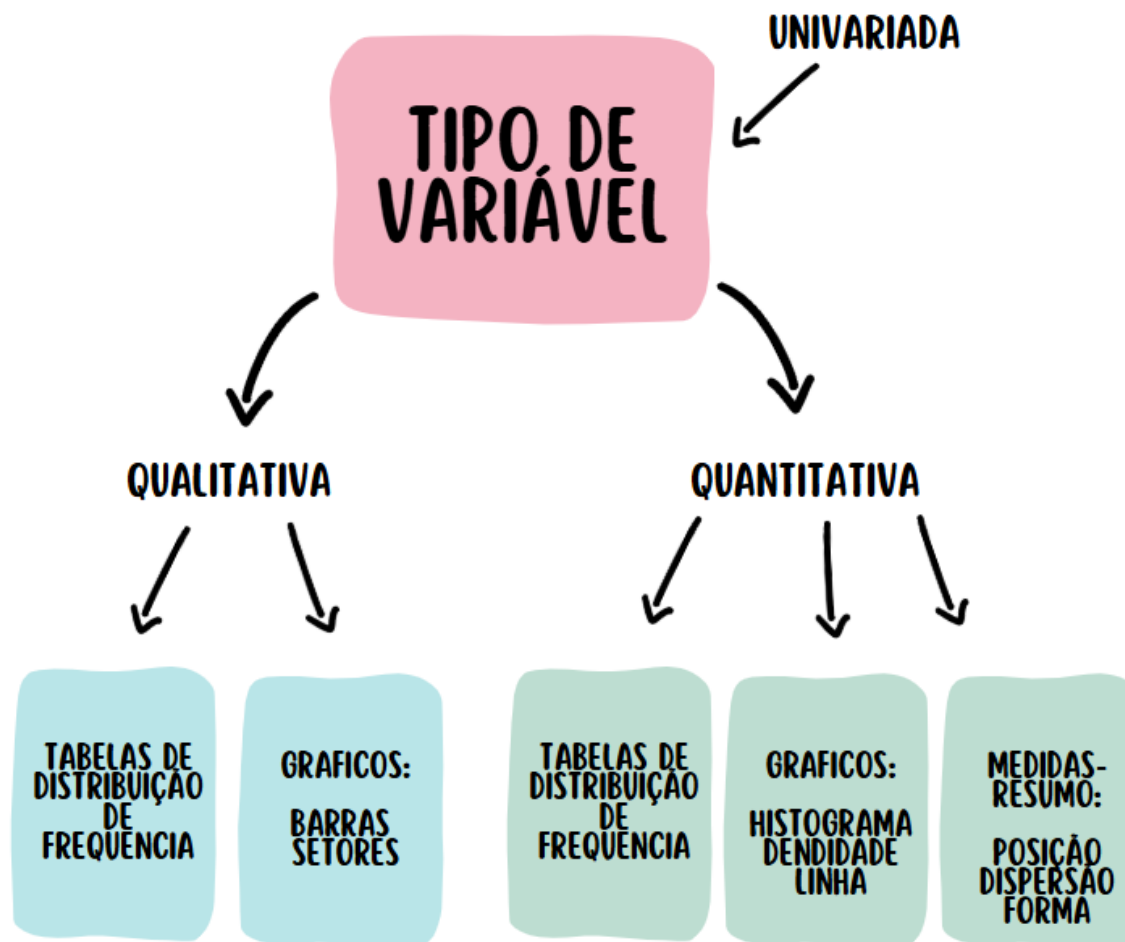


A Estatística Descritiva é a etapa fundamental para a análise de dados. Nela, organizamos, resumimos e representamos os dados por meio de tabelas, gráficos e medidas numéricas facilitando sua interpretação.

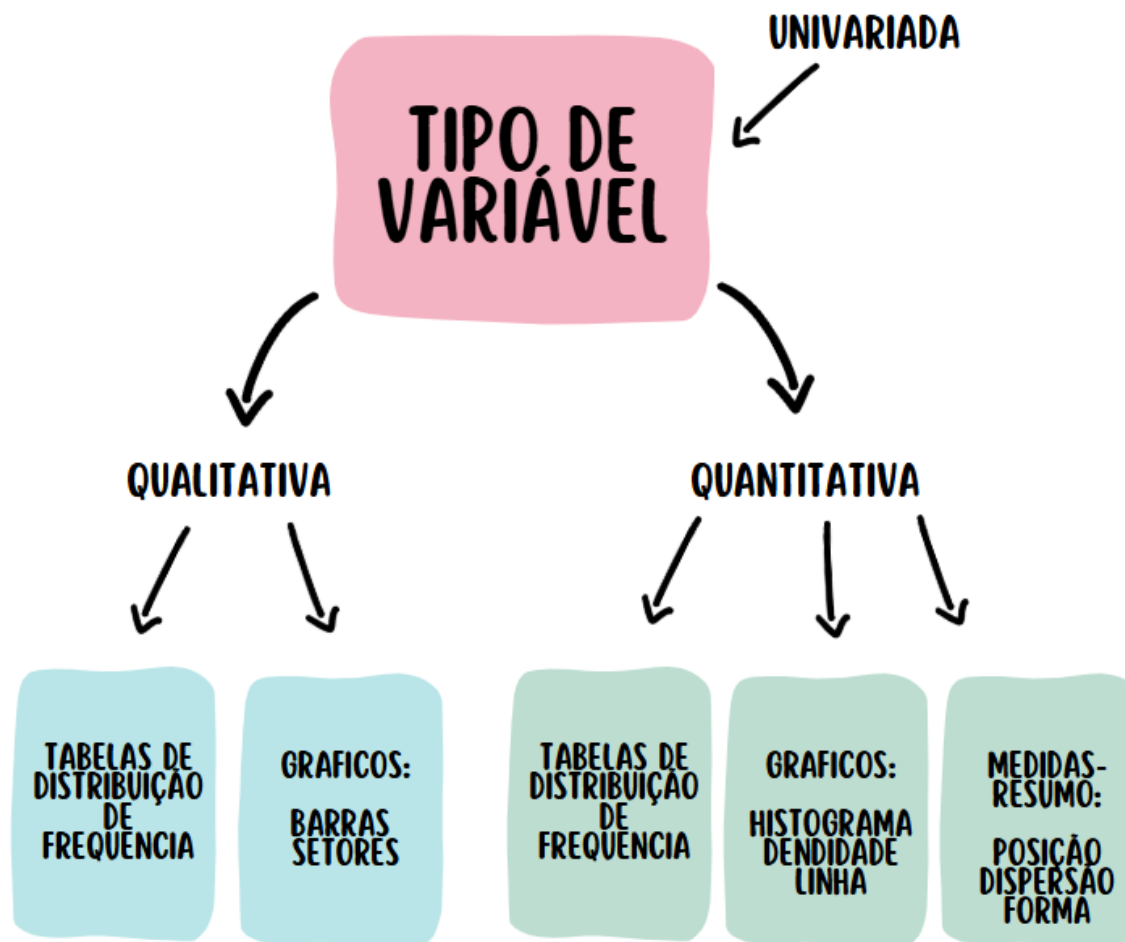
A análise é baseada no conjunto de dados em estudo (de uma amostra), sem tirar quaisquer conclusões.



Estatística Descritiva Univariada



Estatística Descritiva Univariada



Tipo de Variável Qualitativa

Uma variável qualitativa é aquela que expressa uma característica, categoria ou atributo, sem valores numéricos **que representem quantidades**. Ela pode ser classificada em diferentes categorias, como cor, gênero, estado civil ou nível de satisfação.

Exemplo: Tipo de cabelo (liso, ondulado, cacheado, crespo) ou opinião sobre um serviço (ruim, regular, bom, excelente),
tipo sanguíneo...



Tabela de distribuição de frequências

A **tabela de frequência** serve para organizar e resumir um conjunto de dados, facilitando sua interpretação. Ela apresenta a distribuição dos valores observados, indicando quantas vezes cada categoria ou intervalo de dados ocorre.

Frequência absoluta

Frequência relativa

Frequência acumulada

Frequência relativa acumulada

Tipo Sanguíneo	F_i	$Fr_i (\%)$	F_{ac}	$Fr_{ac}(\%)$
A+	15	25	15	25
A-	2	3,33	17	28,33
B+	6	10	23	38,33
B-	1	1,67	24	40
AB+	1	1,67	25	41,67
AB-	1	1,67	26	43,33
O+	32	53,33	58	96,67
O-	2	3,33	60	100
Total	60	100		



Gráfico de barras

Representa por meio de barras as frequências absolutas ou relativas de cada possível categoria.

Cada entidade de variável categorica é representada como uma barra

O tamanho da barra representa seu valor numerico.



Muito utilizada com varriáveis ordinais

Dica: organizar as categorias ordinais da esquerda para a direita para serem visualizadas em sequencia



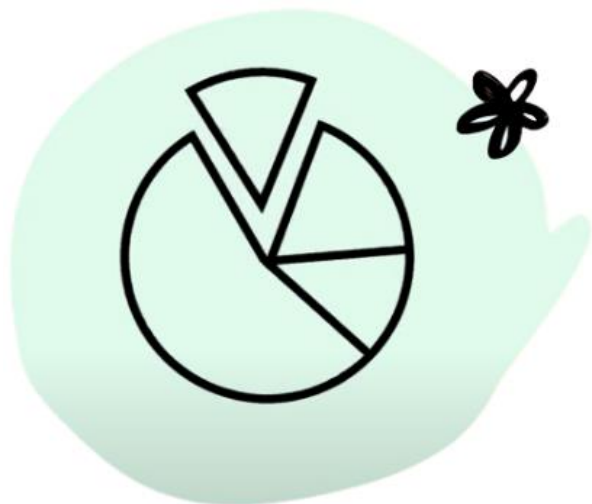
Muito usado quando os nomes das categorias são extensos



Como representamos a Estatística Qualitativa univariada

Grafico de Pizza ou Circular: Representa as frequencias relativas de cada possivel categoria.

É frequentemente usado para mostrar percentagem, em que a soma dos setores é igual a 100%.



Mais chamado de
grafico de pizza ou
grafico circular



Gráfico de setores

No grafico ao lado tente descobrir qual grupo é o maior e tente ordená-los por valor.

Você provavelmente terá dificuldades para fazer isso.

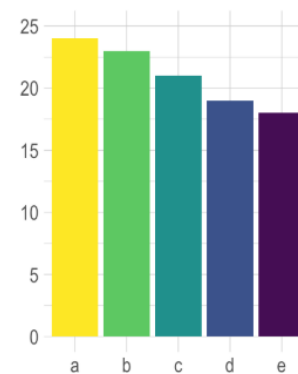
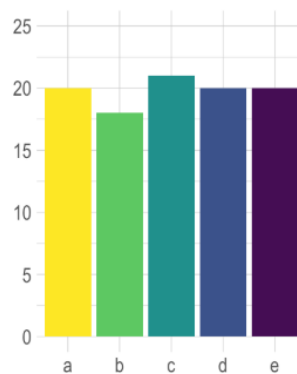
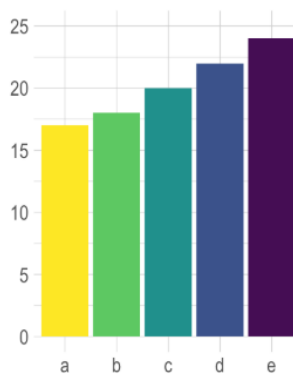
E é por esse motivo que os graficos de pizza devem ser evitados.



Gráfico de setores

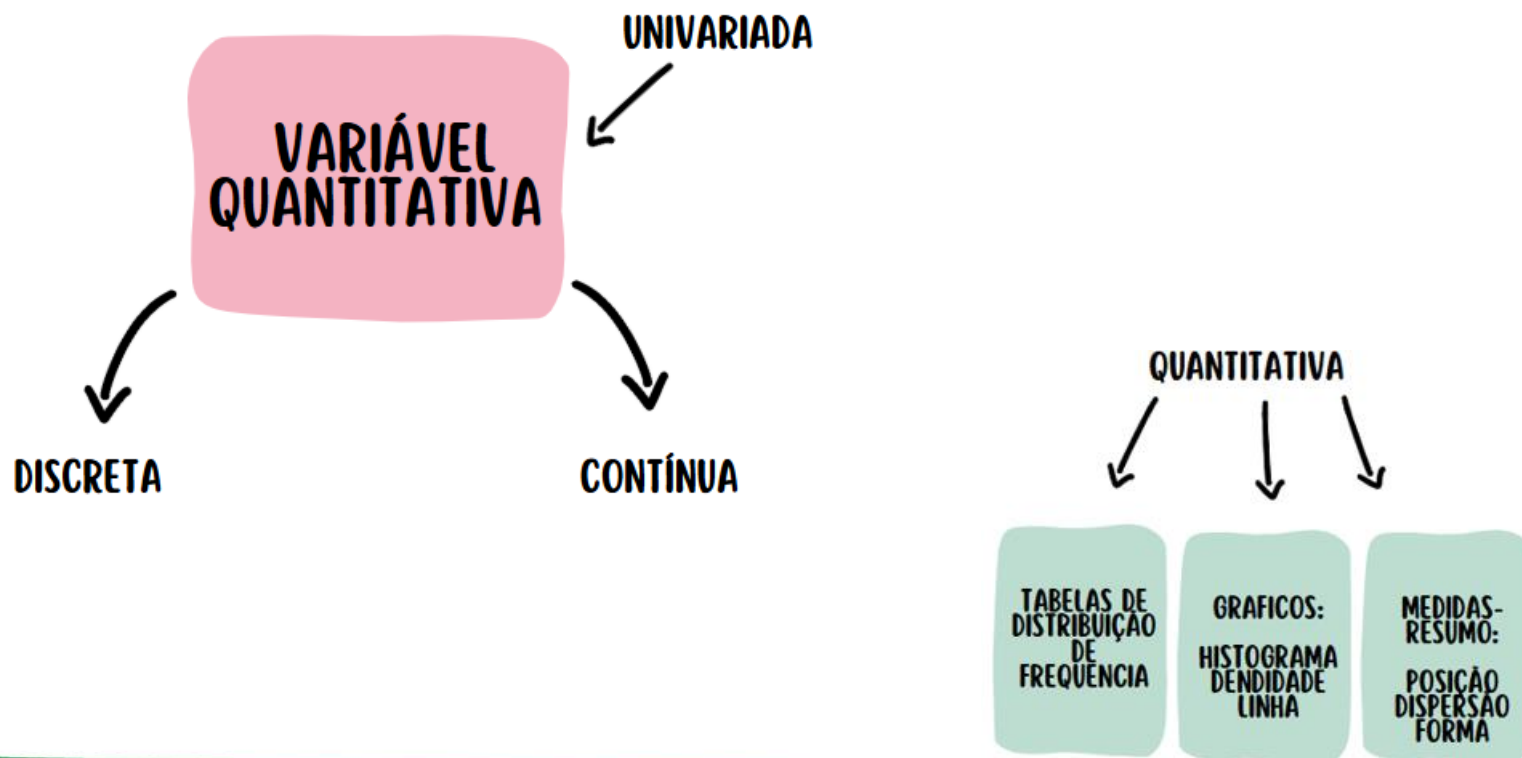
Qua é a evolução do valor entre os grupos?

Qual grupo tem o maior valor nesses tres graficos?



Tipo de Variável Quantitativa

As **variáveis quantitativas** são aquelas que representam valores numéricos e podem ser medidas ou contadas. Elas se dividem em dois tipos principais.



Variável quantitativa

Uma variável quantitativa é aquela que representa valores numéricos e pode ser medida ou contada. Ela expressa quantidades e permite operações matemáticas, como soma e média.

Tipos de variáveis quantitativas:

Discreta: Assume valores inteiros e contáveis. (Exemplo: número de clientes atendidos por dia.)

Contínua: Pode assumir qualquer valor dentro de um intervalo. (Exemplo: tempo de atendimento em minutos.)



Quantitativa discreta

Número de Filhos	F_i	Fr_i (%)	F_{ac}	Fr_{ac} (%)
0	5	25	5	25
1	7	35	12	60
2	6	30	18	90
3	2	10	20	100
Total	20	100		

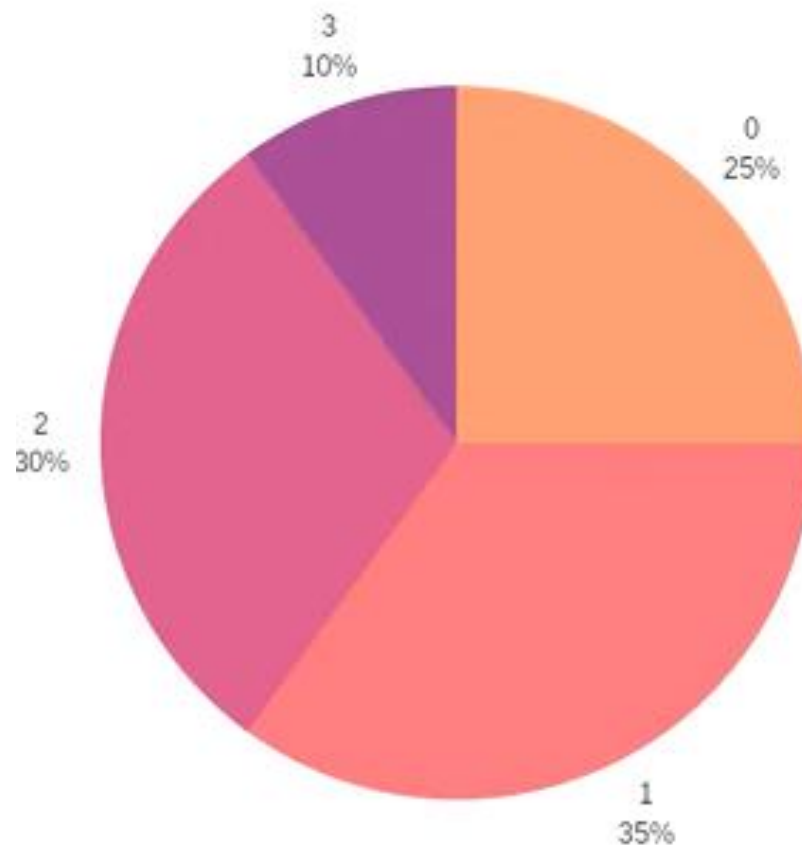
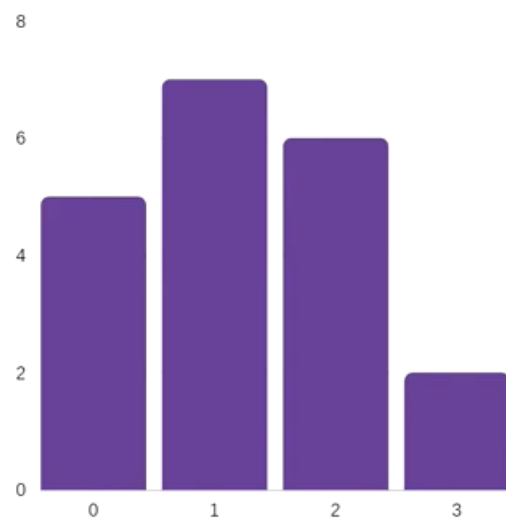


Tabela de distribuição de frequencia

ORDENAÇÃO

Notas dos 30 alunos na disciplina de Estatística.

4,2	3,9	5,7	6,5	4,6	6,3	8,0	4,4	5,0	5,5
6,0	4,5	5,0	7,2	6,4	7,2	5,0	6,8	4,7	3,5
6,0	7,4	8,8	3,8	5,5	5,0	6,6	7,1	5,3	4,7

Dados anteriores ordenados de forma crescente.

3,5	3,8	3,9	4,2	4,4	4,5	4,6	4,7	4,7	5
5	5	5	5,3	5,5	5,5	5,7	6	6	6,3
6,4	6,5	6,6	6,8	7,1	7,2	7,2	7,4	8	8,8



Quantitativa contínua

Frequência absoluta

Frequência relativa

Frequência acumulada

Frequência relativa acumulada

Classe	F_i	$F_i(\%)$	F_{ac}	$F_{ac}(\%)$
[3,5;4,5)	5	16,67	5	16,67
[4,5;5,5)	9	30	14	46,67
[5,5;6,5)	7	23,33	21	70
[6,5;7,5)	7	23,33	28	93,33
[7,5;8,5)	1	3,33	29	96,67
[8,5;9,5)	1	3,33	30	100
Soma	30	100		



Estatística Inferencial

Estatística

Estatística Inferencial usa métodos para tirar conclusões sobre uma população com base em uma amostra, permitindo previsões, estimativas e testes de hipóteses com um grau de confiança. Para isso, utiliza técnicas como testes de hipóteses, intervalos de confiança e análise de regressão, entre outras.



Objetivos da Estatística

- **1.1. Descrever e Resumir Dados**

- **Objetivo:** Organizar e apresentar informações de forma compreensível, utilizando tabelas, gráficos e medidas descritivas.
- **Exemplo:** Apresentar a média e a dispersão dos salários em uma empresa.

- **1.2. Identificar Padrões e Relações**

- **Objetivo:** Detectar tendências, associações ou correlações nos dados.
- **Exemplo:** Identificar se há relação entre o aumento das vendas e uma campanha publicitária específica.

- **1.3. Fazer Previsões**

- **Objetivo:** Usar dados passados para prever eventos ou comportamentos futuros.
- **Exemplo:** Estimar a demanda de produtos com base nas vendas históricas.



Objetivos da Estatística

- **1.4. Apoiar a Tomada de Decisão**

- **Objetivo:** Oferecer informações baseadas em evidências para decisões estratégicas ou operacionais.
- **Exemplo:** Decidir a alocação de recursos em projetos com maior retorno esperado.

- **1.5. Avaliar Hipóteses**

- **Objetivo:** Testar suposições ou teorias sobre um conjunto de dados.
- **Exemplo:** Verificar se a mudança no layout de uma loja aumentou o número de vendas.



A principal diferença entre a **Estatística Descritiva** e a **Estatística Inferencial** está no propósito e na forma como tratam os dados:

	Propósito	Foco	Técnicas	Aplicação
Estatística Descritiva	Resumir e descrever os dados disponíveis (população ou amostra) de forma clara e compreensível.	Trabalha exclusivamente com os dados coletados.	Utiliza tabelas, gráficos, medidas de tendência central (média, mediana, moda), medidas de dispersão (variância, desvio padrão, amplitude), entre outras.	Não tira conclusões além dos dados analisados. É usada, por exemplo, para descrever as vendas de uma empresa em um determinado mês.
Estatística Inferencial	Fazer generalizações, estimativas ou previsões sobre uma população com base em uma amostra dos dados.	Vai além dos dados observados para tirar conclusões sobre o conjunto maior (população).	Utiliza métodos como testes de hipóteses, intervalos de confiança, análises de regressão, entre outros.	Baseia-se na teoria das probabilidades para medir o grau de incerteza das conclusões. É usada, por exemplo, para prever o comportamento de todos os clientes de uma empresa com base em uma pesquisa realizada com uma amostra.



Conceitos

Análise de dados

Estatística e Análise de dados estão relacionadas, mas não representam o mesmo conceito.

- Enquanto a Estatística é um ramo da Matemática que lida com a coleta, análise, interpretação, apresentação e organização de dados, a Análise de Dados é um campo interdisciplinar que utiliza técnicas e métodos estatísticos, programação de computadores e conhecimento de domínio (área de negócio) específico para extrair informações e conhecimentos a partir de dados brutos.



Conceitos

Análise de dados

Podemos dizer que a estatística está contida na análise de dados?



Conceitos

Análise de dados

Sim, podemos dizer que a **Estatística está contida na Análise de Dados**, pois ela é um dos pilares fundamentais dessa área. A Análise de Dados utiliza a Estatística como uma ferramenta essencial para:

- Organizar e resumir dados (Estatística Descritiva);
- Realizar inferências e previsões a partir de amostras (Estatística Inferencial);
- Validar modelos e hipóteses por meio de testes estatísticos.



Conceitos

Análise de dados

Enquanto a Estatística fornece a base **teórica e metodológica** para lidar com dados, a Análise de Dados expande esse escopo para explorar **tecnologias e estratégias práticas**, permitindo extrair valor dos dados em um mundo digital, integrando outras disciplinas como:

- **Ciência da Computação:** Para processar grandes volumes de dados (Big Data);
- **Visualização de Dados:** Para comunicar os resultados de forma clara e compreensível;
- **Conhecimento de Domínio:** Para aplicar as análises a contextos específicos, como negócios, saúde, engenharia, entre outros;
- **Aprendizado de Máquina e IA:** Para construir modelos preditivos mais complexos e adaptativos.



Principais Tarefas da Análise de Dados

- **1. Coleta de Dados**

- Recolher dados relevantes de diversas fontes, como bases de dados, planilhas, sistemas empresariais, redes sociais, APIs, entre outros.
- Garantir que os dados sejam representativos do problema ou questão em análise.

- **2. Preparação de Dados (Pré-processamento):**

- **Limpeza dos dados:** Identificar e corrigir dados inconsistentes, incompletos ou duplicados.
- **Transformação dos dados:** Ajustar formatos, normalizar escalas ou converter categorias para facilitar a análise.
- **Integração de dados:** Combinar dados provenientes de diferentes fontes num único conjunto coeso.



Tarefas da Análise de Dados

- **3. Exploração de Dados:**

- Identificar padrões, tendências e relações nos dados.
- Utilizar visualizações iniciais (gráficos e tabelas) para entender o comportamento e a distribuição dos dados.

- **4. Análise Estatística e Computacional:**

- Aplicar técnicas estatísticas (descritiva ou inferencial) para obter métricas como médias, variações, correlações, entre outras.
- Utilizar algoritmos ou modelos de machine learning para aprofundar a análise e prever resultados.

- **5. Criação de Modelos:**

- Construir modelos descritivos (como clustering) ou preditivos (como regressões ou redes neurais).
- Ajustar os modelos para melhorar a sua precisão e relevância.



Tarefas da Análise de Dados

- **6. Visualização de Dados:**

- transformar os resultados em gráficos, dashboards e relatórios que sejam claros e compreensíveis.
- Usar ferramentas como Power BI, Tableau ou Python para criar representações visuais que destaquem os principais insights.

- **7. Interpretação dos Resultados:**

- Traduzir os dados e os resultados das análises em insights significativos para os stakeholders.
- Garantir que os resultados estejam alinhados com os objetivos do negócio ou da investigação.

- **8. Tomada de Decisão e Ação:**

- Usar os insights para propor soluções, apoiar decisões estratégicas ou otimizar processos.
- Implementar recomendações baseadas nos resultados obtidos.



Tarefas da Análise de Dados

- **9. Monitorização e Atualização:**
 - Monitorizar os indicadores e métricas continuamente para verificar a eficácia das ações implementadas.
 - Reavaliar o processo de análise caso novos dados ou situações surjam.



Principais Tarefas

Análise de Dados	Estatística
Coleta de Dados	Descrever e Resumir Dados
Preparação de Dados (Pré-processamento)	Identificar Padrões e Relações
Exploração de Dados	Fazer Previsões
Análise Estatística e Computacional	Avaliar Hipóteses
Criação de Modelos	Apoiar a Tomada de Decisão
Visualização de Dados	
Interpretação dos Resultados	
Tomada de Decisão e Ação	



Obrigado!

