



ESTATÍSTICA ORIENTADA À CIÊNCIA DE DADOS



INFORMAÇÕES

- Aulas às sextas-feiras, das 19h às 22h10
- Códigos e materiais (<https://github.com/aulasUnisenai/ES-2025>)
- Atividades e plano de aula: AVA (Moodle)





Apresentação - professor

- **Doutor** em Gestão da Informação (UFPR/ 2023);
- **Mestre** em Gestão da Informação (UFPR/ 2019);
- **Bacharel** em Gestão da Informação (UFPR/ 2017);
- Experiência profissional em: **desenvolvimento de sistemas, *machine learning*; análise quantitativa de dados; consultoria.**





Interesses de pesquisa

- **Análise quantitativa de dados**
- **Recuperação de informação musical**
- **Desenvolvimento de sistemas**
- **Robótica inteligente de baixo custo**
- **Processamento multimodal de informação (áudio, imagem, texto)**
- **Inteligência artificial**



No UniSENAI

- Design de software aplicado à Engenharia (algoritmos)
- Projeto de sistemas orientados a objetos
- Projeto de banco de dados não relacionais
- *Web frontend technologies*
- *Backend development*
- Jornada de Aprendizagem – Interface de Programação de Aplicações



No UniSENAI

- Estatística orientada à ciência de dados
- Estatística aplicada à contabilidade
- Inteligência artificial aplicada
- *Data Science*
- Jornada de Aprendizagem – Otimização de sistemas e IoT





Conteúdos formativos

- Estimativas de localização (média, moda, mediana e estimativa robusta);
- Estimativas de variabilidade (variância, desvio padrão, coeficiente de variação);
- Distribuição de dados (percentis e boxplot);
- Gráficos de dispersão, explorando duas ou mais variáveis, aplicações;



Conteúdos formativos

- Amostragem aleatória, viés de amostra, viés de seleção;
- Distribuição de amostragem;
- Intervalos de confiança;
- Testes estatísticos (teste A/B, testes de hipótese, testes t, testagem múltipla);
- Análise de variância (ANOVA);
- Correlação, regressão linear, regressão linear múltipla.

APS



- **APS 1:**
 - Estatística descritiva e análise e exploração de gráficos e tabelas
- **APS 2:**
 - Testes de diferenças entre grupos.

Avaliações

- **1ª avaliação:**
 - 26/09/2025
- **2ª avaliação:**
 - 28/11/2025
- **Exame final:**
 - 12/12/2025

Linguagem R

- A disciplina será conduzida com auxílio da linguagem R
- Trata-se de uma linguagem voltada ao ensino de Estatística
- IDE recomenda: RStudio (suporte total para as bibliotecas e aplicações)
- Para apoio, atividades e exemplos serão conduzidos em Excel, para facilitar a compreensão





Introdução à Estatística: conceitos básicos





Estatística

- Dedicar-se a **coletar, organizar, analisar e interpretar dados**;
- O principal objetivo é **extrair informações relevantes dos dados**, permitindo **tomar decisões**;



Estatística

- Aplicada em diversas áreas, como ciências sociais, economia, saúde, ciências naturais, computação, entre outras;
- Importante na pesquisa científica, auxiliando na formulação de hipóteses, análise de resultados e validação de conclusões;



Estatística

- Utiliza técnicas específicas, como a **estatística descritiva**, que **resume e descreve os dados**, e a **estatística inferencial**, que faz **inferências e generalizações** a partir de dados amostrais para uma população maior;



Estatística

- É conduzida por meio de **gráficos, tabelas, medidas de resumo, testes de hipóteses, intervalos de confiança e modelos de regressão**, entre outros métodos estatísticos.



Estatística descritiva

Área da estatística que se dedica a organizar, resumir e descrever conjuntos de dados;

Seu principal objetivo é tornar os dados mais compreensíveis, permitindo que sejam analisados e interpretados de forma clara e concisa;



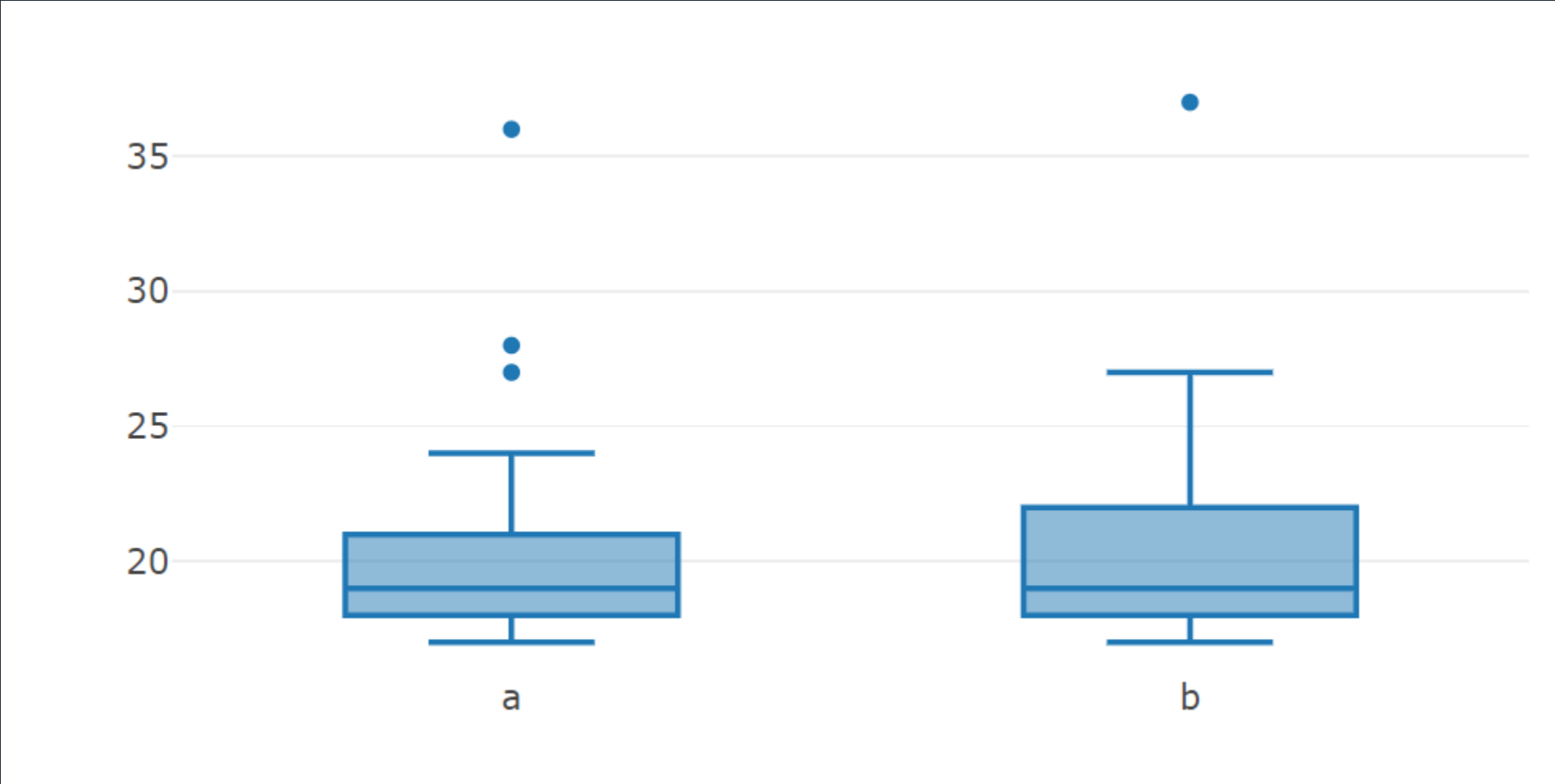
Estatística descritiva

Ao aplicar técnicas estatísticas descritivas, **pode-se obter uma visão geral dos padrões e características dos dados, facilitando a identificação de tendências, variações e relações entre variáveis.**



Estatística descritiva

Exemplo: calcular a média e o desvio padrão da altura de um grupo de estudantes universitários para obter uma compreensão geral da distribuição da idade nesse grupo específico.





Turma	n	Mínimo	Média	Desvio padrão	Máximo
A	34	17	20,32	3,78	36
B	31	17	20,58	4,09	37



Estatística inferencial

Área da estatística que se preocupa com a análise e interpretação de dados amostrais a fim de fazer inferências ou generalizações sobre a população a partir da qual a amostra foi obtida.



Estatística inferencial

Seu objetivo é tirar conclusões sobre a população com base nos resultados observados na amostra;

Utiliza técnicas como testes de hipóteses, intervalos de confiança e regressão, entre outras, para fazer inferências sobre parâmetros populacionais.



Estatística inferencial

Exemplo: conduzir um teste de hipótese para verificar se a média de idade de **todos os estudantes universitários (população)** é diferente de uma determinada média de referência, com base nos dados de idade coletados de uma **amostra de estudantes**.



Estatística inferencial

A média de idade das turmas de Software é 20 anos?

Resposta: supondo que a Turma A seja representativa de todas as turmas de Software e sob condições de normalidade, não há evidências ($\alpha=0.05$) para afirmar que a média populacional difere de 20 anos ($t=0.49$, $p\text{-valor} > 0.05$).

Conceito	Estatística descritiva	Estatística inferencial
Definição	Descreve e resume dados disponíveis	Faz inferências e generalizações sobre a população a partir de uma amostra específica
Objetivo	Apresentar informações do conjunto de dados	Tirar conclusões sobre a população com base em dados amostrais
Dados	Lida apenas com os dados disponíveis	Usa dados amostrais para fazer projeções sobre a população
Técnicas	Utiliza gráficos, tabelas e medidas de resumo	Utiliza testes de hipóteses, intervalos de confiança, modelos de regressão, entre outros
Exemplo	Calcular média, mediana, moda e desvio padrão de um conjunto de notas	Testar se a média de salários em uma amostra é representativa de toda a empresa

Conceito	Estatística Descritiva	Estatística Inferencial
Uso	Resumo de dados, identificação de padrões e tendências	Tomada de decisões, previsões, validação de hipóteses em pesquisas
População	Não faz inferências sobre a população	Faz inferências e generalizações sobre a população
Amostra	Pode trabalhar com dados de uma população ou amostra	Requer uma amostra representativa para fazer inferências
Nível de “Certitude”	Fornece informações sobre os dados disponíveis	Estima a probabilidade de que as conclusões se apliquem à população

População e amostra



População

A população refere-se ao conjunto completo de elementos ou indivíduos que compartilham características específicas de interesse em um estudo.



Amostra

A amostra é um subconjunto representativo selecionado da população, usado para fazer inferências e generalizações sobre a população como um todo;

A escolha de uma amostra adequada garante que os resultados obtidos sejam confiáveis e aplicáveis à população de interesse.

Característica	População	Amostra
Definição	Conjunto completo de elementos ou indivíduos de interesse no estudo	Subconjunto selecionado da população para análise
Objetivo	Avaliar características da totalidade dos elementos	Inferir características da população a partir de um grupo representativo
Tamanho	Geralmente grande, incluindo todos os elementos	Menor do que a população, mas representativo dela
Seleção	Não é selecionada; é o conjunto total de interesse	Selecionada usando técnicas de amostragem

Característica	População	Amostra
Importância	Fornece informações sobre o universo de interesse	Permite generalizações sobre a população
Viés	Não está sujeita a viés de seleção	Pode conter viés se a amostra não for representativa
Exemplo	Todos os alunos de uma escola formam a população	Uma amostra aleatória de alunos é selecionada para uma pesquisa



Amostra

Probabilística

Não
Probabilística

Aleatória
Simples

Estratificada

Sistemática

Conglomerado

Conveniência

Julgamento

Intencional

Bola de neve



Probabilística

Envolve a seleção de elementos da população de forma aleatória e com probabilidade conhecida



Não Probabilística

A escolha dos elementos é feita de forma não aleatória, com base em critérios subjetivos ou disponibilidade

Amostragem probabilística

Tipo	Definição	Vantagens	Desvantagens
Aleatória simples	Cada elemento da população tem a mesma probabilidade de ser selecionado para a amostra.	Garante representatividade; Fácil de implementar; Cálculo do erro amostral.	Exige lista completa da população; Inválida para populações amplas ou dispersas.
Estratificada	Divide a população em subgrupos (estratos) e seleciona amostras aleatórias em cada um.	Representa heterogeneidade; Reduz erro amostral; Garante inclusão de todos estratos.	Requer informações prévias sobre a população.
Sistemática	Seleciona elementos em intervalos regulares após escolher aleatoriamente o primeiro.	Mais eficiente que a AAS; Fácil de implementar em populações ordenadas.	Sensível a padrões periódicos na população.
Por conglomerados	Divide a população em grupos; Seleciona aleatoriamente alguns grupos e inclui todos.	Eficiente para populações dispersas; Facilita logística de amostragem.	Potencialmente menos representativa; Complexidade na análise dos conglomerados.

Tipo	Aleatória Simples	Estratificada	Sistemática	Por Conglomerados
Exemplo	Em uma escola com 500 alunos, utilizar um gerador de números aleatórios para selecionar 50 alunos para uma pesquisa acadêmica.	Em uma pesquisa eleitoral, dividir a população em grupos por faixa etária (18-29, 30-49, 50+ anos) e selecionar aleatoriamente nas faixas.	Em um supermercado com 100 produtos, selecionar a cada 5º produto na prateleira após sortear o número inicial.	Em uma pesquisa nacional de saúde, selecionar aleatoriamente algumas cidades e incluir todos os indivíduos nessas cidades na amostra.

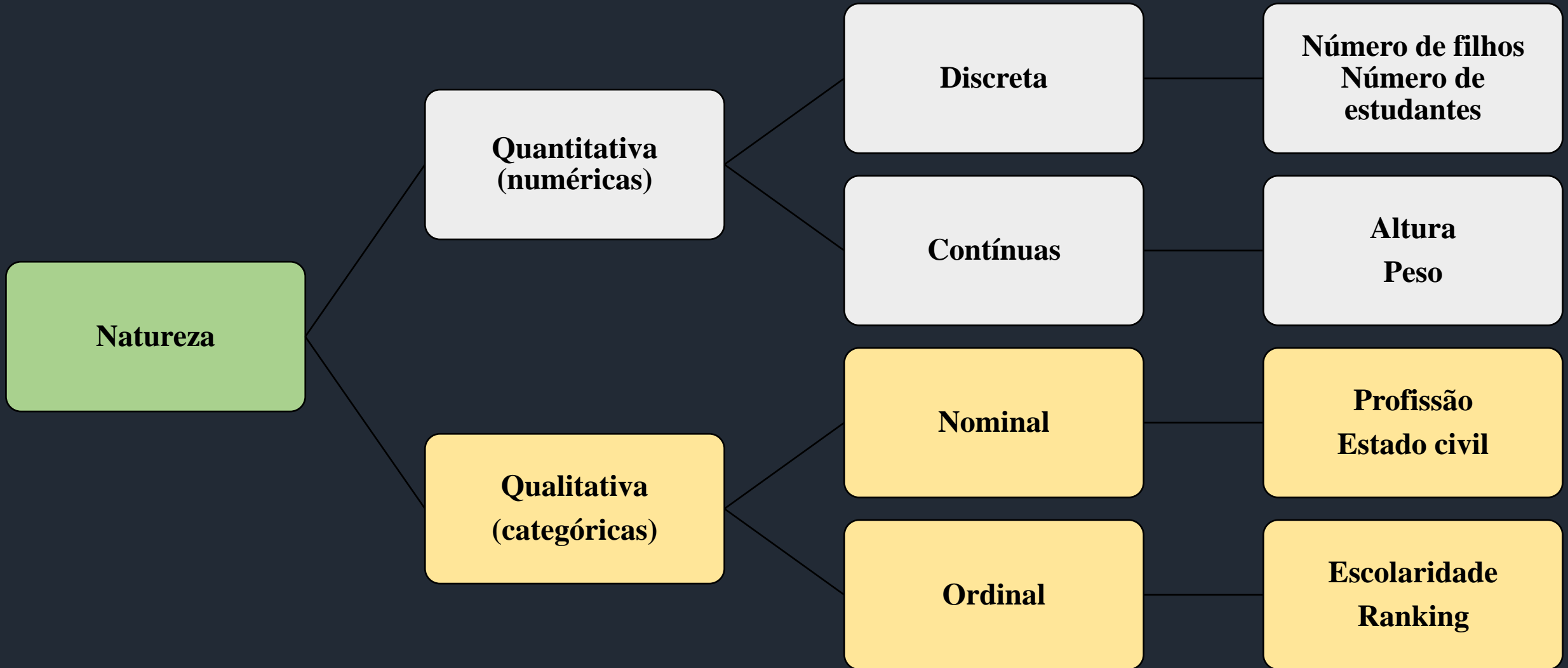
Amostragem não probabilística

Tipo	Definição	Vantagens	Desvantagens
Por Conveniência	Elementos são selecionados com base na disponibilidade, facilidade de acesso ou conveniência do pesquisador.	Rápida e econômica; Fácil de implementar; Adequada para estudos exploratórios.	Amostra pode não ser representativa da população; Pode conter vieses e não refletir a diversidade do grupo estudado.
Julgamento (Quota)	O pesquisador seleciona manualmente elementos da amostra com base em critérios subjetivos ou cotas pré-definidas.	Fácil de implementar; Permite representar subgrupos específicos da população.	Suscetível a vieses e subjetividade na escolha dos elementos; Não garante representatividade da população em geral.

Tipo	Definição	Vantagens	Desvantagens
Intencional (Propósito)	O pesquisador seleciona deliberadamente elementos que considera mais relevantes ou representativos.	Útil para estudos qualitativos ou quando a representatividade não é a principal preocupação.	Não garante representatividade; Pode não ser adequada para estudos que requerem inferências estatísticas generalizáveis.
Bola de Neve	Usada quando a população é difícil de alcançar ou identificar; os participantes indicam outros participantes.	Útil para grupos de difícil acesso; Ajuda a identificar participantes ocultos ou pouco conhecidos; Pode obter informações valiosas.	Não garante representatividade; Pode levar a amostras pequenas e limitadas; Suscetível a vieses e falta de diversidade.

Tipo	Por Conveniência	Julgamento (Quota)	Intencional (Propósito)	Bola de Neve
Exemplo	Entrevistar alunos de uma universidade que estão próximos ao local da pesquisa para obter respostas rapidamente.	Entrevistar 50 homens e 50 mulheres para representar igualmente as duas categorias de gênero na amostra de uma pesquisa.	Entrevistar especialistas em determinado campo para obter opiniões qualificadas sobre um tema específico de estudo.	Em um estudo sobre uma comunidade marginalizada, o pesquisador entrevista alguns membros que indicam outros membros para entrevista.

Tipos de variáveis/dados



Variáveis qualitativas

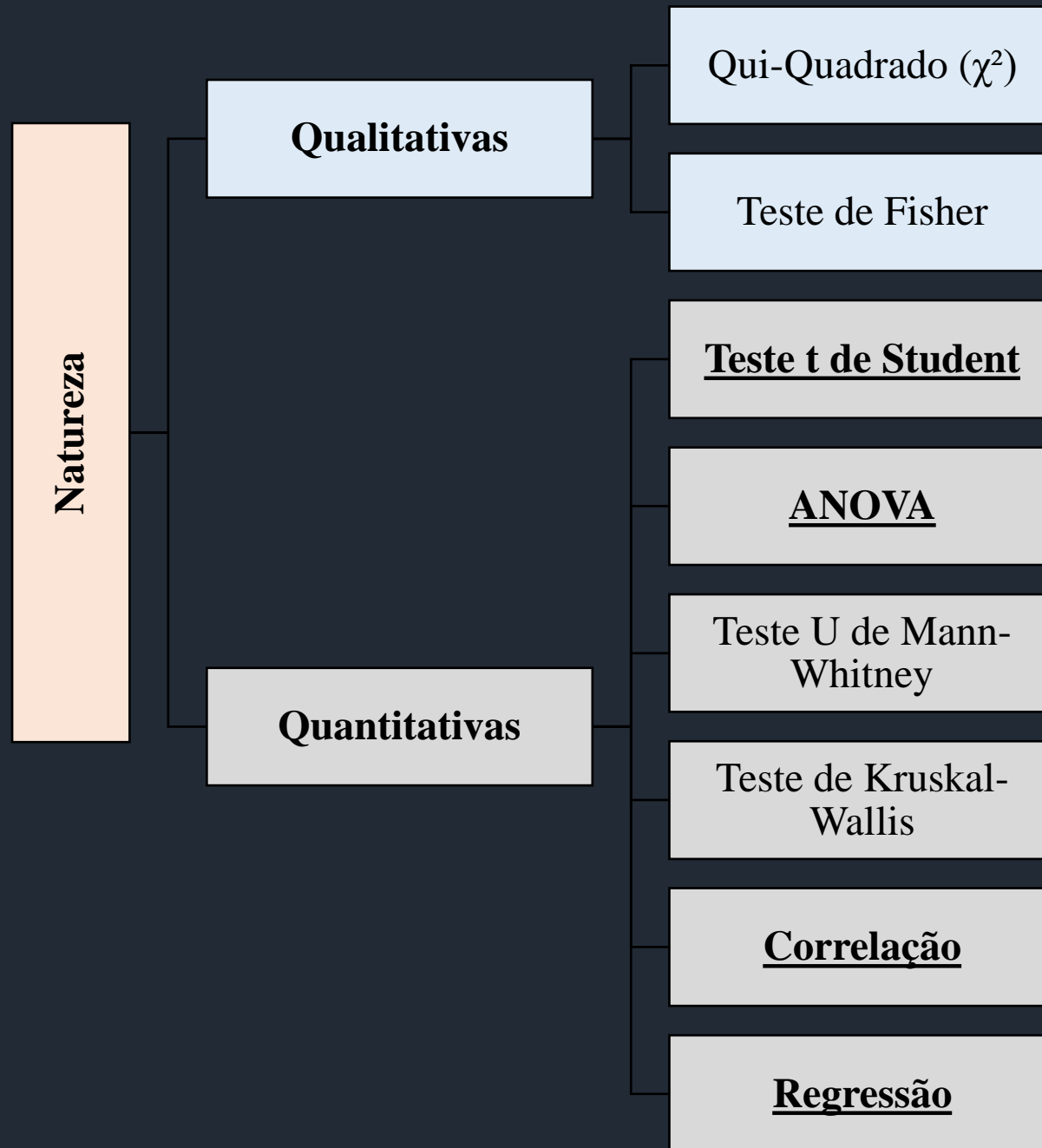
Representam **características, atributos ou categorias**, e **não podem ser medidas em uma escala numérica contínua**;

1. **Variáveis nominais:** categorias sem ordem específica.
2. **Variáveis ordinais:** categorias com uma ordem específica, mas não possuem igual intervalo entre os valores.

Variáveis quantitativas

Representam quantidades ou valores numéricos que podem ser submetidos a cálculos matemáticos;

1. **Variáveis discretas:** valores individuais são contados ou enumerados.
2. **Variáveis ordinais:** medem quantidades com intervalos iguais entre os valores, mas não possuem um zero absoluto.



Introdução à linguagem R e ao ambiente RStudio