

Estatística para Farmácia

Unidade I -

Análise Exploratória de Dados.

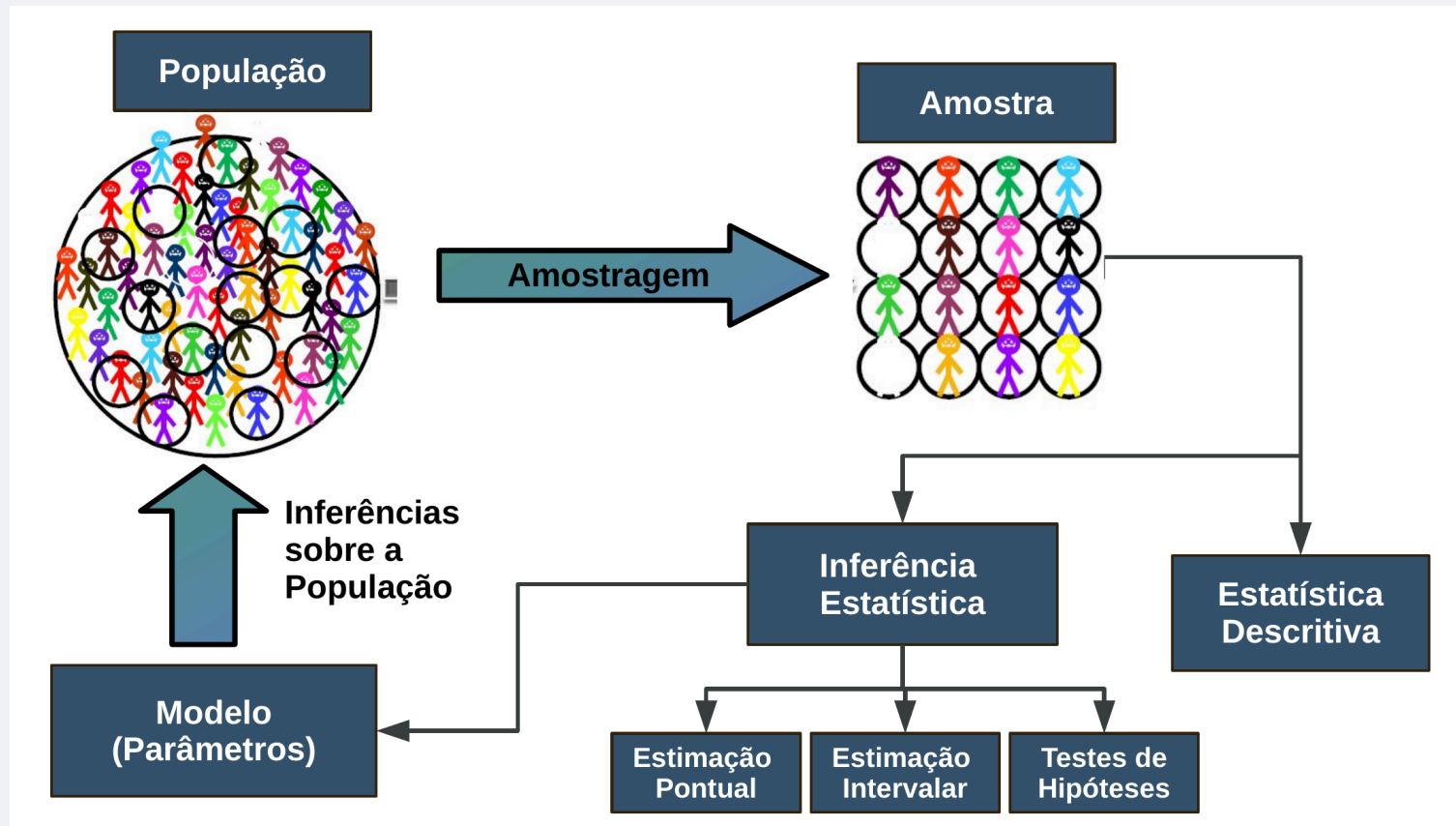
Prof. Thiago A. N. De Andrade

Universidade Federal de Santa Maria
Departamento de Estatística

Aviso aos estudantes

- Este é um material novo e atualizado, elaborado especialmente para nosso curso **Estatística para Farmácia - UFSM 2024.2**. Entretanto, **não se configura em conteúdo original**. É apenas uma compilação resumida de conteúdos presentes nas referências citadas. Em resumo: é indispensável consultar as referências indicadas.
- As imagens não são autorais e os respectivos créditos são reservados aos autores.
- Este material foi integralmente produzido em R Markdown, utilizando o pacote `xaringan`, que possibilita a criação de apresentações **ninja**.

Professor, o que é essa tal de Estatística? 🤔 🤔 🤔 🤔



Unidade 000: Conceitos básicos e motivações

1. **População (N):**

É a coleção completa de todos os elementos que possuem em comum uma certa característica de interesse para o estudo.

2. **Amostra (n):** É um subconjunto de elementos da população, coletada segundo critérios estatísticos.

- Na maioria das pesquisas científicas é impossível avaliar todos os elementos que compõem uma população de interesse de estudo.
- Isto se deve principalmente ao custo e tempo necessário para coletar
- Com a finalidade de estudar **população**, retiramos desta uma parte representativa que chamamos de **amostra**, e coletamos dados apenas desta amostra.

Exemplo

1. Uma pesquisa para determinar a eficácia de um novo medicamento antiviral envolve o recrutamento de pacientes diagnosticados com o vírus em um hospital universitário.

R:

População: Todos os pacientes diagnosticados com o vírus.

Amostra: Os pacientes diagnosticados que foram recrutados no hospital universitário.

Exemplo

2. Uma indústria farmacêutica conduz testes de controle de qualidade em lotes de vacinas contra a gripe antes de distribuí-los. Seleciona aleatoriamente frascos de diferentes lotes para análise.

R:

População: Todos os frascos de vacina contra a gripe produzidos pela indústria.

Amostra: Os frascos selecionados aleatoriamente para os testes de controle de qualidade.

Exemplo

3. Um estudo de mercado para entender as preferências dos consumidores em relação a suplementos alimentares. Os pesquisadores coletam dados por questionário online enviado a clientes de farmácias.

R:

População: Todos os consumidores de suplementos alimentares.

Amostra: Os clientes que responderam ao questionário online.

Estatísticas (amostra) X Parâmetros (população)

Estatísticas (amostra):

A estatística resume uma característica de uma amostra. É uma estimativa do parâmetro correspondente. Representada geralmente por uma letra romana.

Parâmetros (população):

O parâmetro resume uma característica de uma população. Representado geralmente por uma letra grega.

	Estatísticas	Parâmetros
Média	\overline{X}	μ
Desvio Padrão	s	σ
Variância	s^2	σ^2
Número de elementos	n	N
Proporção	\hat{p}	p
Correlação	r_{xy}	ρ_{xy}

Organização de dados quanto à sua origem

Quanto à sua origem, os dados podem ser obtidos de:

- **Estudos observacionais:** Quem aplica a pesquisa não tem controle intencional sobre os fatores que influenciam as respostas.
- **Estudos experimentais:** Quem aplica o experimento tem controle intencional sobre os fatores que influenciam as respostas. Estudos experimentais são geralmente randomizados, agrupando as unidades amostrais por acaso.

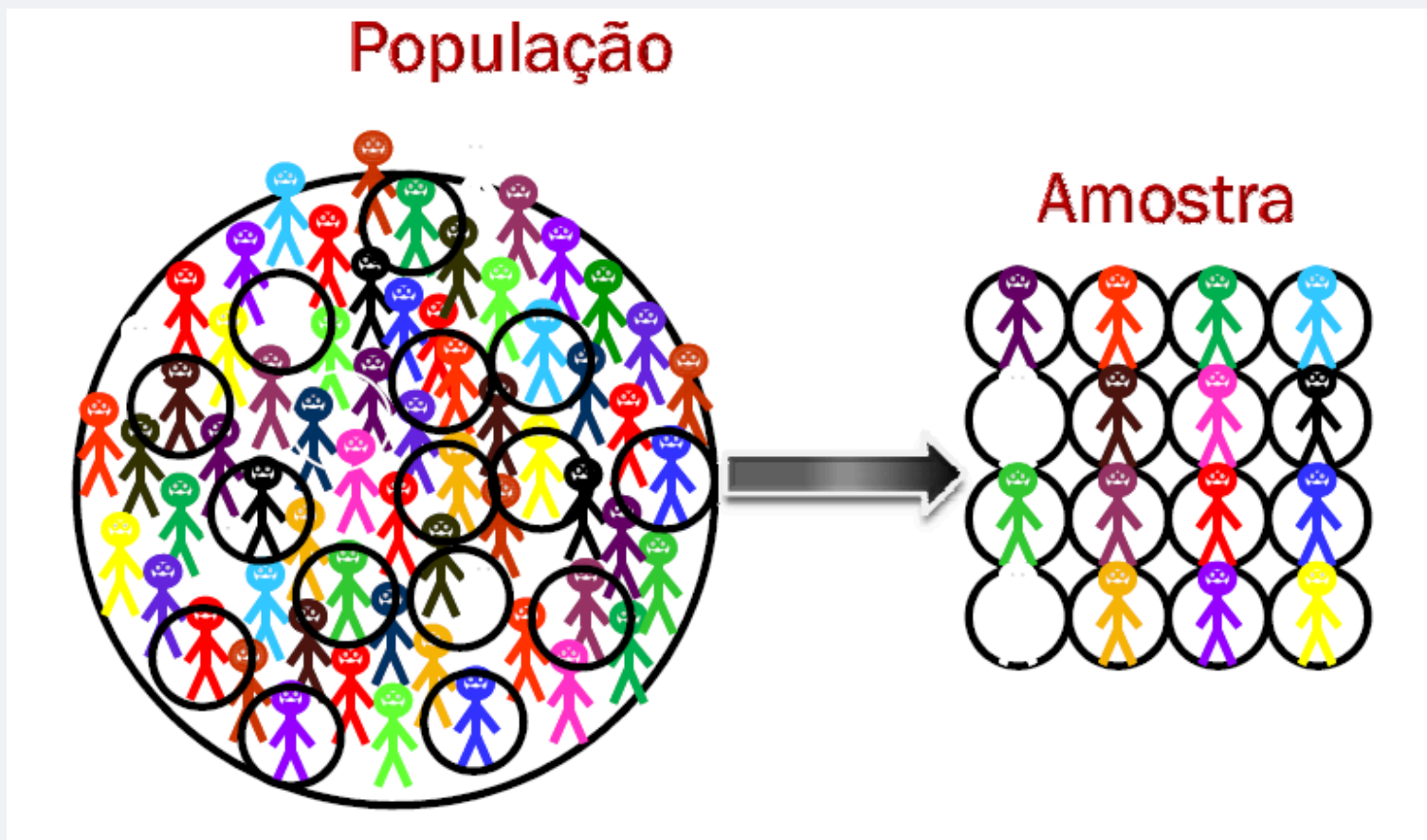
Exemplos Observacionais

1. Um estudo longitudinal acompanha pacientes com doença cardíaca para examinar a associação entre o uso de medicamentos anti-hipertensivos e a progressão da doença.
2. Um estudo de caso-controle investiga fatores de risco para resistência a antibióticos em pacientes de um hospital. Compara pacientes que desenvolveram resistência (casos) com aqueles que não desenvolveram (controles).

Exemplos Experimentais

1. Pesquisadores desenvolvem um medicamento para tratar enxaquecas e realizam um ensaio clínico randomizado. Um grupo recebe o medicamento experimental, e outro, placebo. Comparação é feita quanto à frequência e intensidade das enxaquecas.
2. Para investigar o impacto de um suplemento dietético na concentração de estudantes universitários, pesquisadores dividem voluntários em dois grupos. Um grupo recebe o suplemento, o outro placebo. O desempenho dos estudantes é avaliado em testes de concentração.

No meu TCC/Artigo: para obter essa tal de amostra basta escolher um pequeno grupo qualquer da população?



OK, parece legal. Mas, como a estatística se conecta com meu curso?



1.Bioestatística e Pesquisas Clínicas

- Estatística é usada para interpretar dados de pesquisas clínicas, ajudando a determinar a eficácia e segurança de novos medicamentos, incluindo testes clínicos de fase I, II, III e IV.

2.Controle de Qualidade em Farmácias e Indústrias Farmacêuticas

- Técnicas estatísticas como controle estatístico de processos (CEP) são usadas para monitorar a qualidade dos medicamentos durante a produção, garantindo que estejam dentro das especificações.

3.Farmacocinética e Farmacodinâmica

- Estatística auxilia no ajuste de modelos matemáticos para descrever a absorção, distribuição, metabolismo e excreção de fármacos, assim como sua interação com o organismo.

4. Análise de Estudos de Bioequivalência

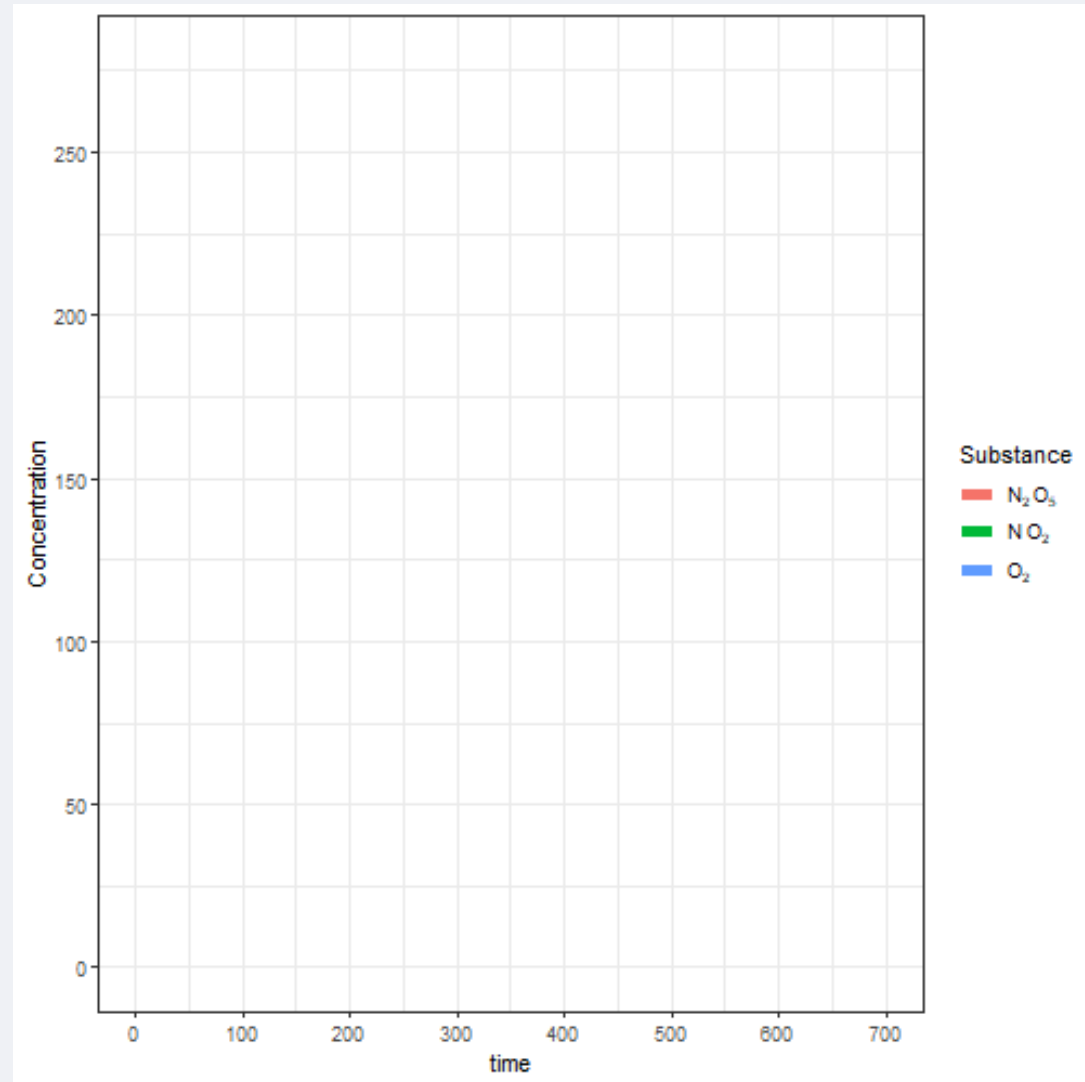
- Métodos estatísticos são aplicados para comparar a biodisponibilidade de medicamentos genéricos com os medicamentos de referência.

5. Ensaio de Estabilidade de Medicamentos

- Estatísticas ajudam a modelar e prever o prazo de validade de medicamentos através de estudos de estabilidade, avaliando a degradação dos compostos ao longo do tempo.

Quer mais exemplos?





- **N₂O₅** é o pentóxido de dinitrogênio, um composto químico formado por dois átomos de nitrogênio (N) e cinco átomos de oxigênio (O). Ele é um anidrido ácido do ácido nítrico (HNO₃), o que significa que, em presença de água, N₂O₅ se hidrata e forma ácido nítrico de acordo com a seguinte reação:
$$\text{N}_2\text{O}_5 + \text{H}_2\text{O} \rightarrow 2\text{HNO}_3.$$
- O **NO₂** é o dióxido de nitrogênio, um gás de coloração marrom-avermelhada com um odor característico e pungente. Ele é um óxido de nitrogênio e faz parte da família dos compostos conhecidos como **NO_x** (óxidos de nitrogênio), que são poluentes atmosféricos importantes.
- O **O₂** é o **oxigênio molecular**, um gás diatômico composto por duas moléculas de oxigênio. É uma substância essencial para a maioria dos organismos na Terra, pois é fundamental no processo de **respiração celular**.

Aplicações na pesquisa em Farmacologia

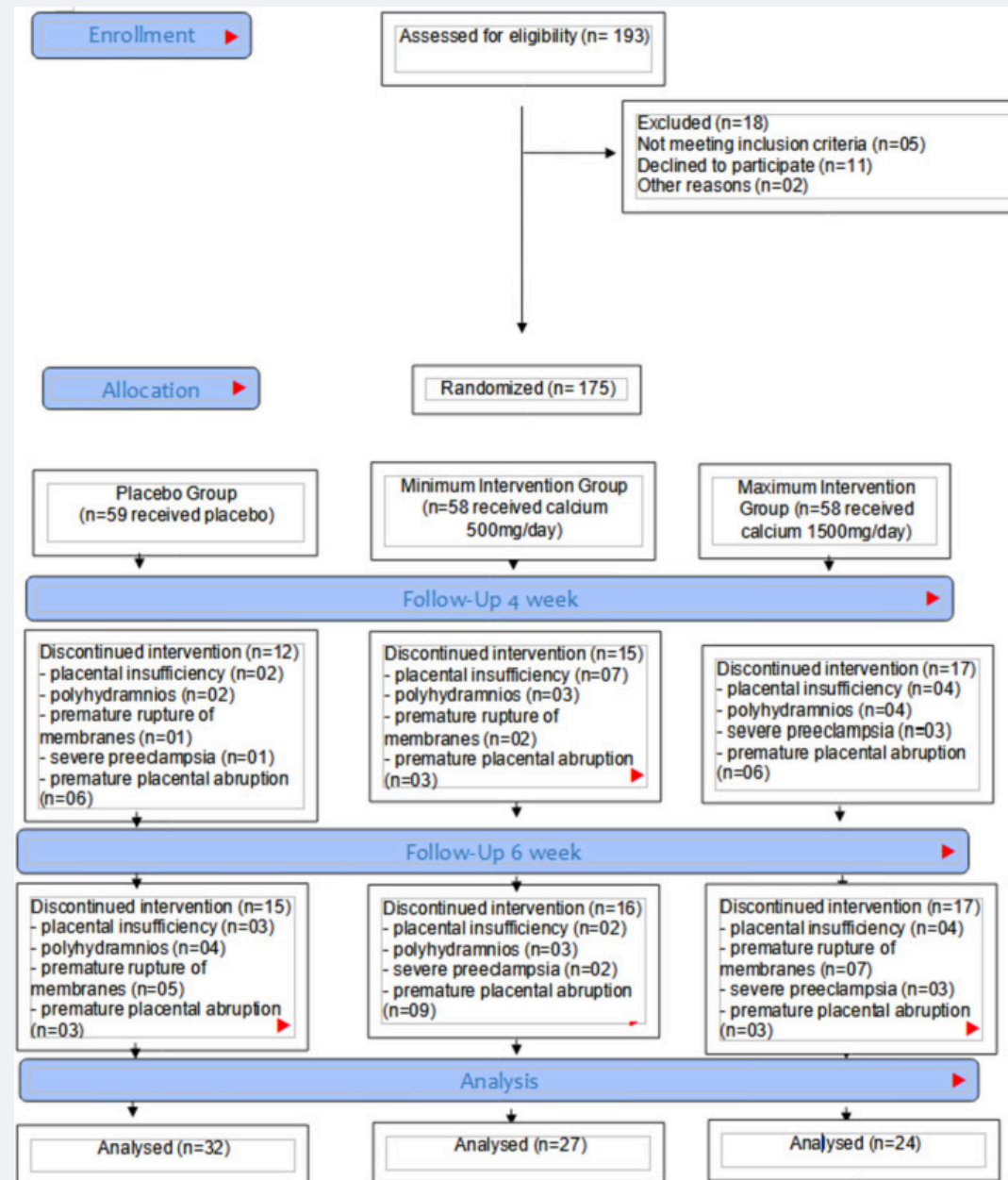
- Effects of calcium supplementation on changes in the IL2, IL4, IL6, IL10 axes and oxidative stress in pregnant women at risk for pre-eclampsia
- Calcium supplementation commencing before or early in pregnancy, or food fortification with calcium, for preventing hypertensive disorders of pregnancy (Review)
- Use of recombinant S1 protein with hFc for analysis of SARS-CoV-2 adsorption and evaluation of drugs that inhibit entry into VERO E6 cells

Sobre o primeiro artigo da lista

Efeitos da suplementação de cálcio nas alterações dos eixos IL2, IL4, IL6, IL10 e no estresse oxidativo em gestantes com risco de pré-eclâmpsia

- IL Interleucina
- Os eixos IL2,...,IL10 representam um conjunto de citocinas que regulam a resposta inflamatória e a atividade das células imunológicas no sistema imunológico.
- A dose de 500 mg/dia de cálcio melhorou o sistema purinérgico, protegendo os vasos, mas com menor efeito anti-inflamatório que doses maiores.

- O sistema purinérgico, composto por receptores P1 e P2 e enzimas como a NTPDase, regula inflamação e protege tecidos, sendo relevante em condições como hipertensão e pré-eclâmpsia.
- A dose de 1.500 mg/dia de cálcio reduziu marcadores inflamatórios e aumentou a defesa antioxidante, oferecendo maior proteção contra danos vasculares e oxidativos na pré-eclâmpsia.



- Inscrição: Dos 193 participantes avaliados, 18 foram excluídos por não atenderem aos critérios (n=5), recusarem participar (n=11) ou por outros motivos (n=2).
- Alocação: Dos 175 participantes elegíveis, 59 foram alocados ao grupo placebo, 58 ao grupo de intervenção mínima (500 mg de cálcio/dia) e 58 ao grupo de intervenção máxima (1.500 mg de cálcio/dia).
- Acompanhamento de 4 semanas: 12 participantes do grupo placebo, 15 do grupo de intervenção mínima e 17 do grupo de intervenção máxima descontinuaram devido a complicações como insuficiência placentária e pré-eclâmpsia.

- Acompanhamento de 6 semanas: 15 participantes do grupo placebo, 10 do grupo de intervenção mínima e 17 do grupo de intervenção máxima descontinuaram após 6 semanas.
- Análise: No final do estudo, 32 participantes do grupo placebo, 27 do grupo de intervenção mínima e 24 do grupo de intervenção máxima foram analisados.

Em resumo, a Estatística vai ser um super aliado na tua formação acadêmica



Pré-requisitos: Comparecer às aulas + Paciência + Estudo ativo + o principal:



Unidade 1: Análise Exploratória de Dados

Objetivos da Unidade

- Introduzir conceitos fundamentais de estatística descritiva.
- Explorar variáveis qualitativas e quantitativas.
- Apresentar metodologias de representação tabular e gráfica.
- Compreender medidas descritivas: tendência central, posição e dispersão.

1.1 Introdução à Análise Exploratória de Dados

A análise exploratória de dados (AED) é uma abordagem inicial usada para resumir as principais características de um conjunto de dados. Seu objetivo é:

- Entender as distribuições das variáveis.
- Identificar padrões.
- Detectar anomalias.
- Formular hipóteses iniciais para estudos mais aprofundados.

1.2 Variáveis Estatísticas

As variáveis estatísticas são os elementos fundamentais que descrevem as características de um conjunto de dados. Elas podem ser classificadas em dois grandes grupos:

- **Qualitativas:** variáveis categóricas que representam qualidades ou categorias.
 - Exemplo: Tipo de medicamento (analgésico, antibiótico), resposta a tratamento (positivo/negativo).
- **Quantitativas:** variáveis numéricas que expressam quantidade.
 - Exemplo: Dosagem de fármaco (mg), idade do paciente (anos), nível de colesterol (mg/dL).

1.2.1 Variáveis Qualitativas

As variáveis qualitativas (categóricas) podem ser subdivididas em dois tipos:

- **Nominais:** Não têm uma ordem intrínseca.
 - **Exemplo:** Tipo de medicamento (analgésico, antibiótico, anti-inflamatório).
- **Ordinais:** Têm uma ordem ou hierarquia natural.
 - **Exemplo:** Grau de dor relatado pelos pacientes (leve, moderado, intenso).

Representação Gráfica para Variáveis Qualitativas

- **Gráfico de Barras:** Frequência ou proporção de cada categoria.
 - **Exemplo:** Frequência de diferentes tipos de medicamentos vendidos em uma farmácia.
- **Gráfico de Pizza:** Proporção de cada categoria em relação ao total.
 - **Exemplo:** Proporção de pacientes respondendo positivamente a diferentes tratamentos.

Exemplo: Variáveis Qualitativas na Farmácia

Uma farmácia está conduzindo um estudo para verificar a eficácia de diferentes tipos de analgésicos. Os pacientes são categorizados de acordo com a resposta ao tratamento.

- Variável qualitativa nominal: **Tipo de analgésico** (paracetamol, ibuprofeno, dipirona).
- Variável qualitativa ordinal: **Nível de eficácia** (ineficaz, eficaz, muito eficaz).

A farmácia pode usar um gráfico de barras para representar a quantidade de pacientes que usaram cada tipo de analgésico, e um gráfico de pizza para mostrar a proporção de pacientes que relataram eficácia.

1.2.2 Variáveis Quantitativas

As variáveis quantitativas podem ser subdivididas em:

- **Quantitativas Discretas:** Assumem valores inteiros e contáveis.
 - **Exemplo:** Número de comprimidos tomados por dia.
- **Quantitativas Contínuas:** Podem assumir qualquer valor em um intervalo contínuo.
 - **Exemplo:** Dosagem de fármaco administrado (mg), concentração de substância no sangue (mg/dL).

Representação Gráfica para Variáveis Quantitativas

- **Histograma:** Distribuição dos valores numéricos em intervalos.
- **Boxplot:** Visualização da mediana, quartis e possíveis outliers.

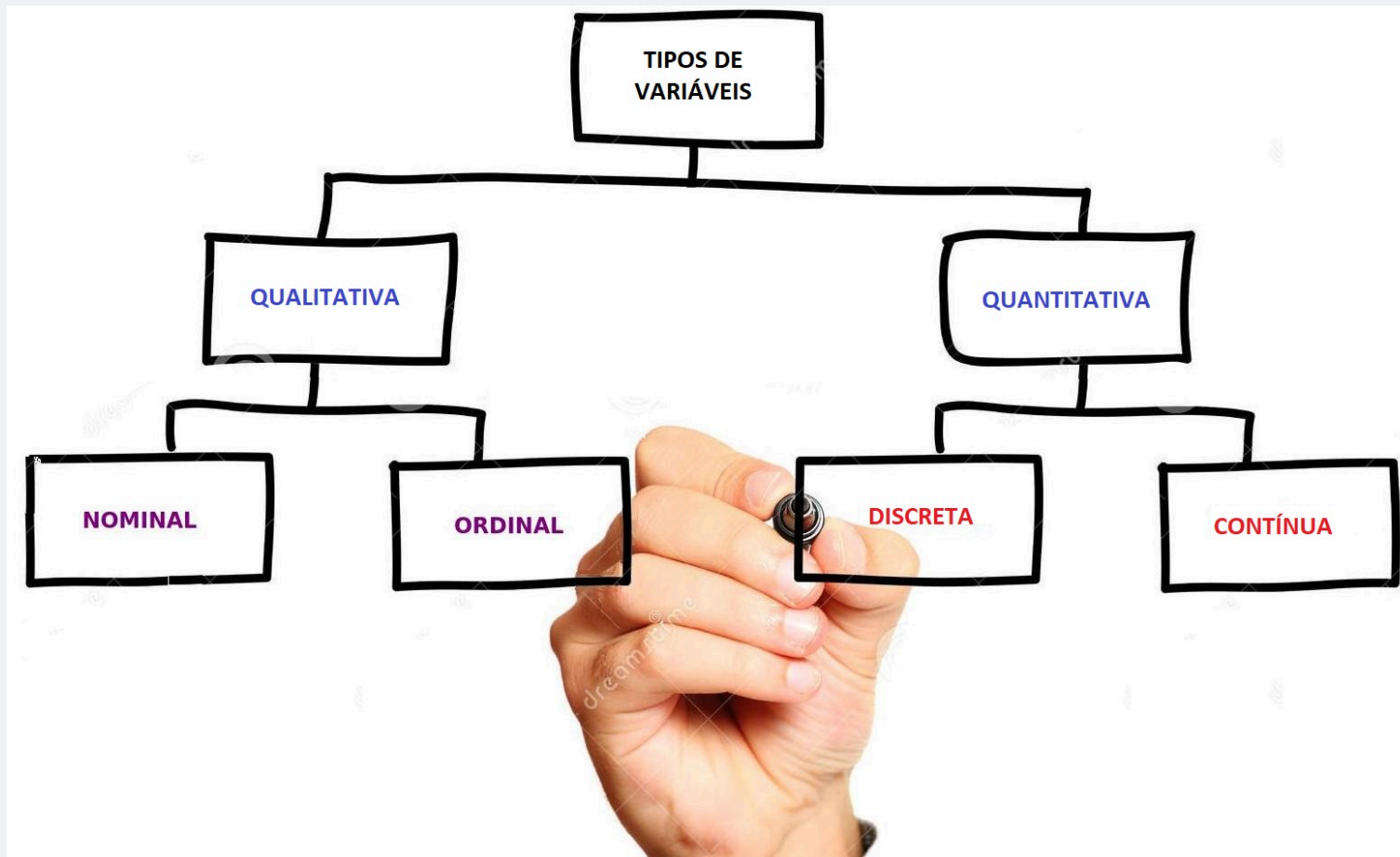
Exemplo: Variáveis Quantitativas na Farmácia

Em um estudo clínico, a dosagem de um medicamento é administrada a diferentes pacientes e medida em miligramas (mg).

- Variável quantitativa contínua: **Dosagem administrada** (mg).
- Variável quantitativa discreta: **Número de doses** tomadas por cada paciente em um dia.

Para visualizar a distribuição da dosagem administrada, podemos usar um **histograma**, e para identificar possíveis outliers, um **boxplot**.

Resume 🙌



Dados Fictícios para o Exemplo

Imagine que estamos analisando os dados de 15 pacientes que receberam um tratamento para hipertensão. Coletamos as seguintes informações:

1. **Idade dos pacientes** (variável quantitativa contínua, em anos):

- 45, 50, 62, 39, 47, 55, 60, 48, 42, 53, 57, 41, 65, 44, 59

2. **Tipo de medicamento** administrado (variável qualitativa nominal):

- A, B, A, C, B, A, C, A, B, C, B, A, C, B, A

1. **Resposta ao tratamento** (variável qualitativa ordinal):

- Eficaz, Eficaz, Ineficaz, Moderado, Moderado, Eficaz, Eficaz, Ineficaz, Moderado, Eficaz, Eficaz, Moderado, Ineficaz, Moderado, Eficaz

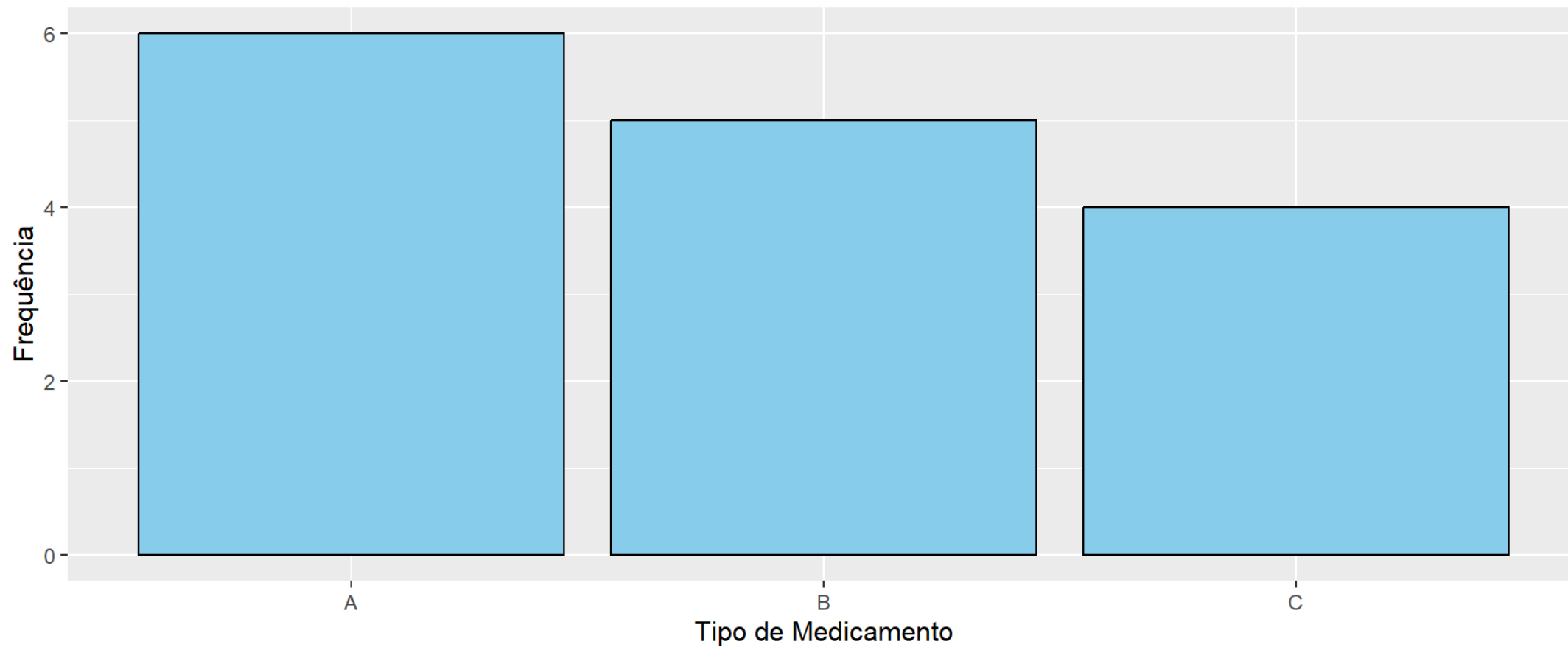
Com base nesses dados, podemos criar diferentes tipos de tabelas e gráficos para representar as variáveis qualitativas e quantitativas.

Resposta ao Tratamento	Frequência
Ineficaz	3
Moderado	5
Eficaz	7

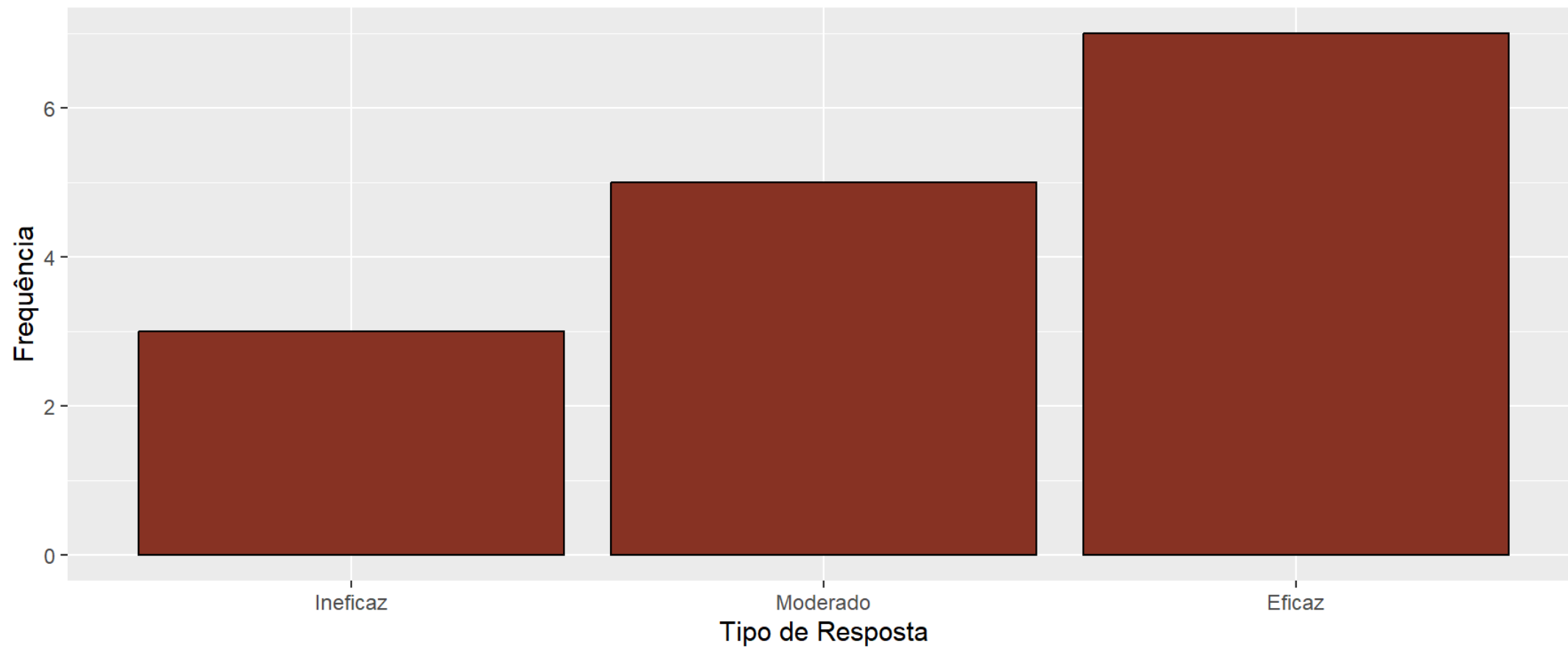
Tipo de Medicamento	Frequência
A	6
B	5
C	4

Intervalo (anos)	Frequência
39-44	4
45-50	4
51-56	3
57-62	3
63-68	1

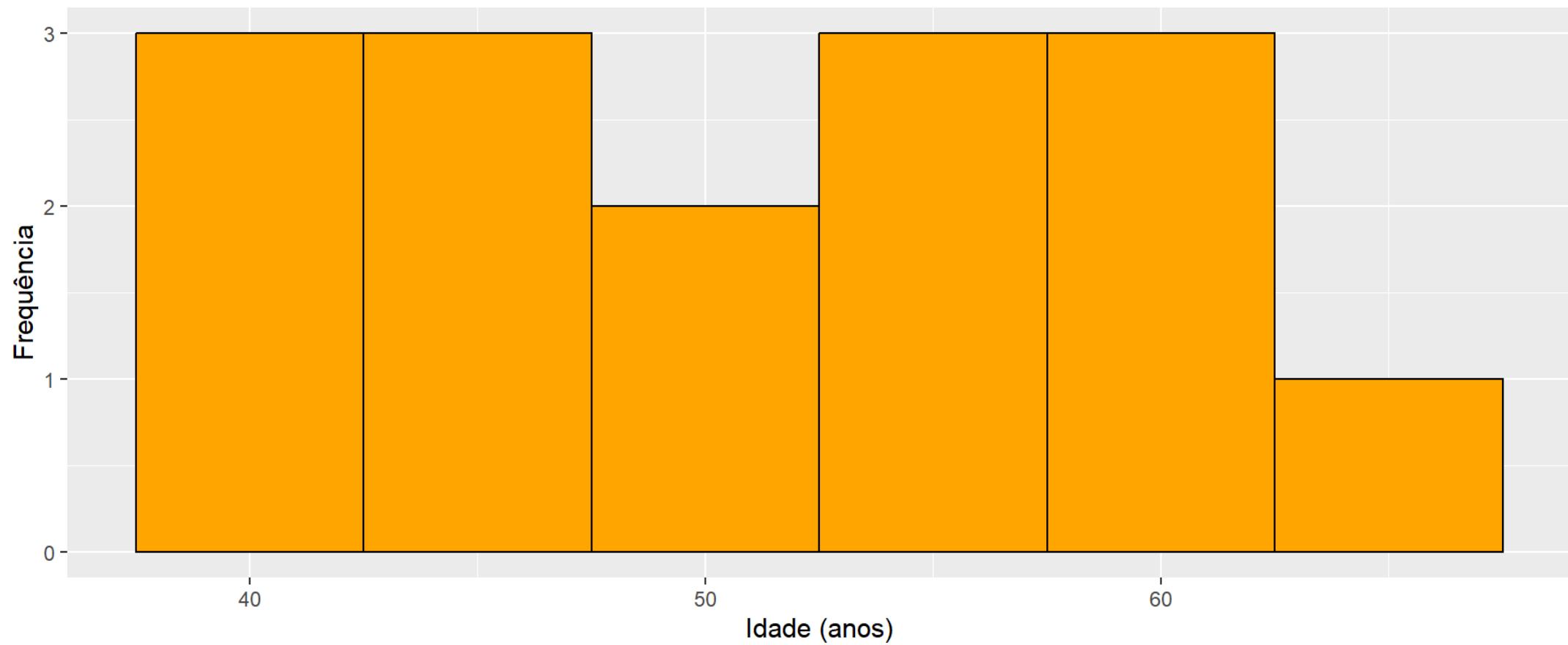
Frequência de Tipo de Medicamento



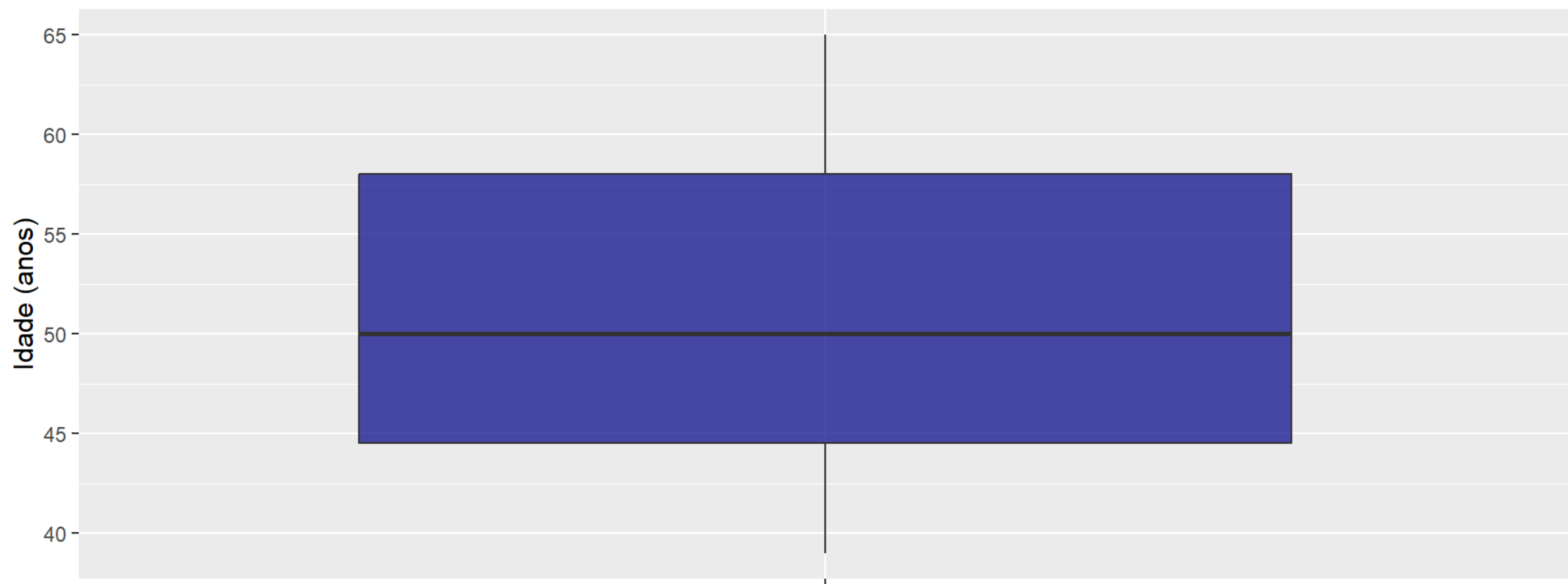
Frequência de Tipo de Resposta



Distribuição de Idade dos Pacientes



Boxplot da Idade dos Pacientes



1.3 Estatísticas Descritivas

As estatísticas descritivas resumem e descrevem características importantes dos dados. Vamos explorar três categorias principais:

1. **Medidas de Tendência Central:** localizam o centro dos dados.
2. **Medidas de Posição:** descrevem a localização de um dado em relação ao conjunto.
3. **Medidas de Dispersão:** medem a variabilidade dos dados.

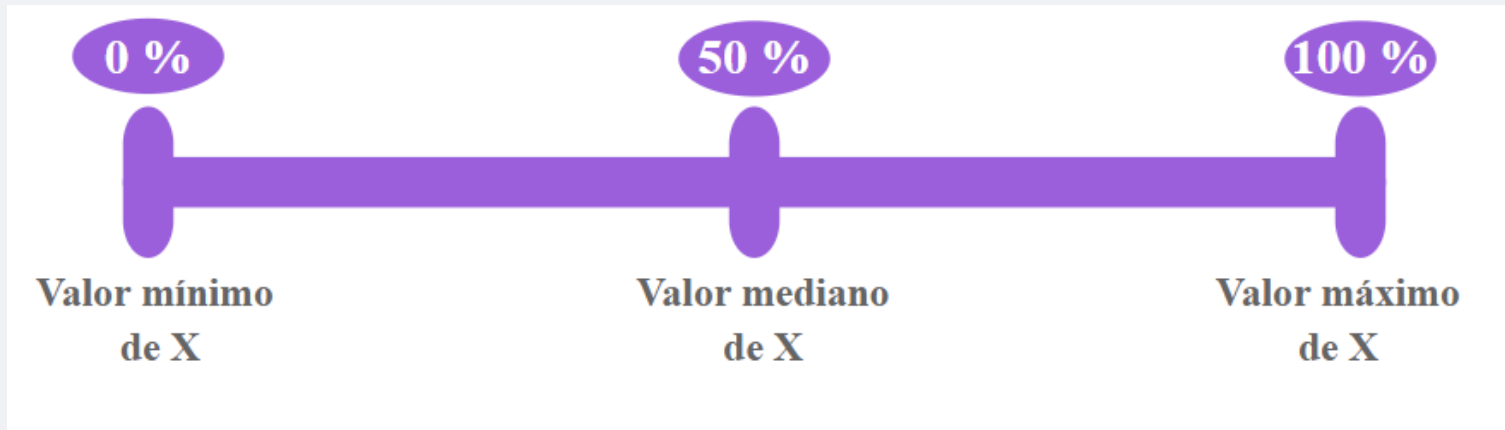
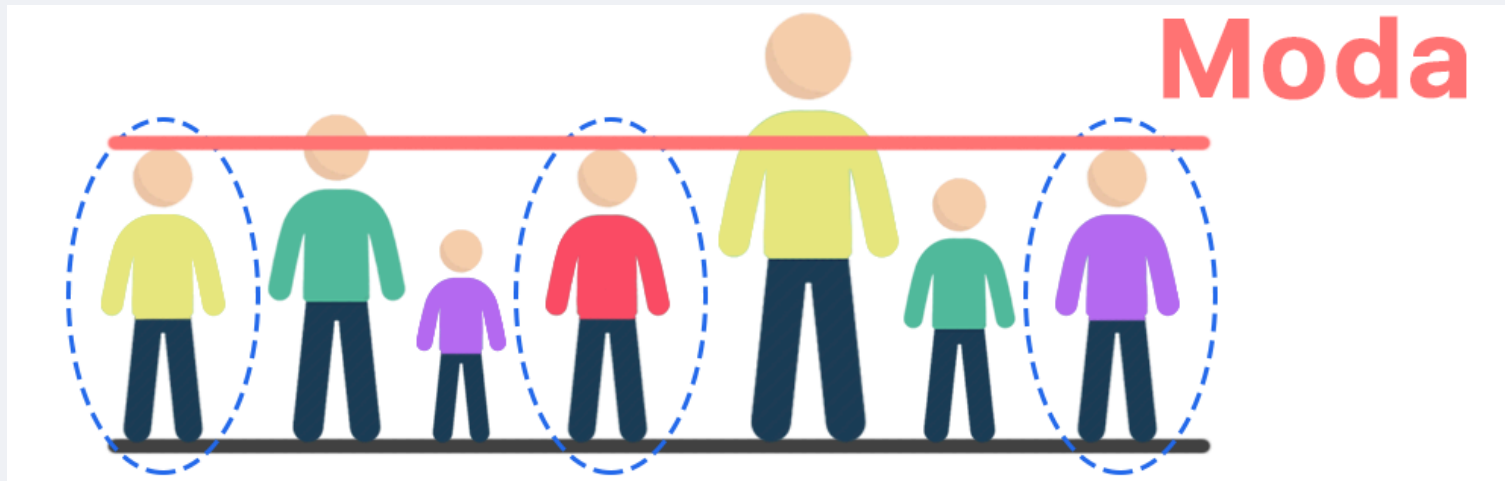
1.3.1 Medidas de Tendência Central

Essas medidas indicam onde a maioria dos dados está concentrada. As principais são:

- **Média (\bar{x})**: A soma dos valores dividida pelo número de observações.

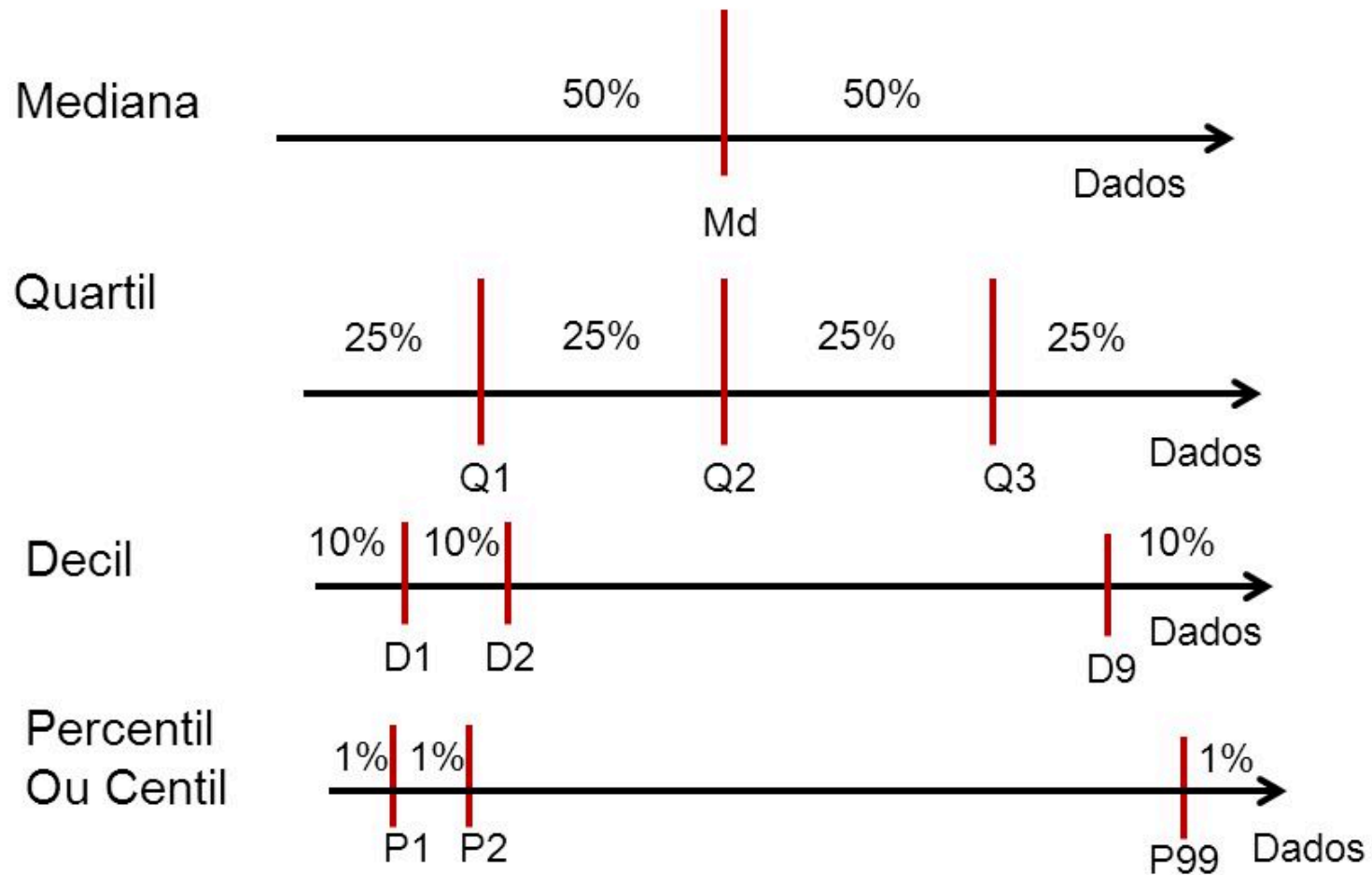
$$\bar{x} = \frac{\sum x_i}{n}$$

- **Mediana**: O valor central quando os dados estão ordenados.
- **Moda**: O valor que ocorre com maior frequência.



1.3.2 Medidas de Posição

- **Quartis:** Dividem os dados em quatro partes iguais.
- **Percentis:** Dividem os dados em 100 partes, facilitando comparações.



1.3.3 Medidas de Dispersão

Medem a dispersão ou variabilidade dos dados.

- **Variância:** A média dos quadrados das diferenças em relação à média.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

- **Desvio Padrão:** A raiz quadrada da variância. $s = \sqrt{s^2}$
- **Amplitude (Range):** A diferença entre o valor máximo e o valor mínimo.
- **Coeficiente de Variação (CV)** é dado pela fórmula: $CV = \frac{s}{\bar{x}} \times 100$

Exercícios Propostos (Dificuldade baixa)

Estamos interessados em estudar a idade em anos dos alunos da turma de Estatística para a Farmácia da UFSM. Temos a seguinte amostra das idades: 60, 19, 25, 30, 22, 23, 20, 18, 21. Calcule as medidas de tendência central e dispersão vistas em sala.

Exercícios Propostos (Dificuldade baixa)

Em um estudo sobre o efeito de um novo medicamento, coletaram-se os seguintes dados:

- a. **Idade dos pacientes** (em anos): 25, 34, 45, 52, 36, 41, 28, 33, 38, 44.
- b. **Tipo de resposta ao tratamento**: Positivo, Negativo, Positivo, Negativo, Positivo, Positivo, Negativo, Positivo, Negativo, Positivo.

Perguntas:

- Classifique as variáveis acima (idade e tipo de resposta) como qualitativas ou quantitativas.
- Para a variável quantitativa, calcule a média e a mediana.
- Para a variável qualitativa, construa uma tabela de frequência e represente os resultados em um gráfico de barras.

Exercícios Propostos (Dificuldade moderada)

Para cada amostra abaixo, calcule as medidas de tendência central e dispersão vistas em sala. Compare os resultados entre as amostras.

Amostra 1: 2, 2, 2, 7, 8, 9, 10, 3

Amostra 2: 3, 3, 3, 7, 7, 7, 8, 9

Amostra 3: 2, 4, 4, 1, 3, 8, 5, 7

Exercícios Propostos (Dificuldade moderada)

Considere os dados abaixo, relativos ao número de pessoas vacinadas em determinados dias do mês de abril, em um posto médico de Santa Maria:

14, 12, 11, 13, 14, 13, 12, 14, 13, 14, 11, 12, 12, 14, 10, 13, 15, 11, 15, 13, 16, 16, 14, 14.

Com base nos dados acima, calcule o que se pede a seguir.

- Construa uma tabela de frequência para os dados.
- Calcule a média, moda e mediana.
- Calcule o desvio padrão.

Exercícios Propostos (Dificuldade alta)

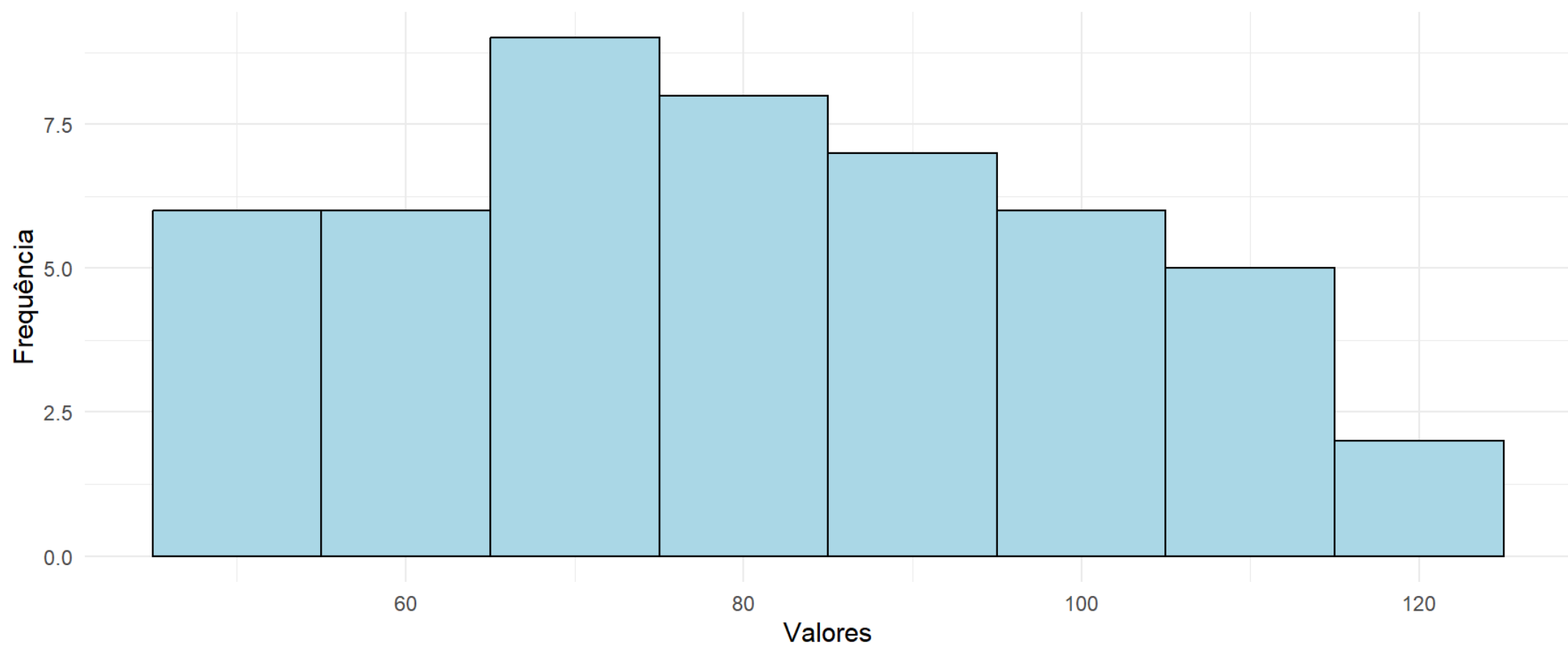
Uma farmácia coletou dados sobre a dosagem de um medicamento administrado a 49 pacientes, com as seguintes dosagens em mg: 50, 60, 70, 80, 90, 100, 60, 70, 80, 90, 50, 60, 70, 80, 100, 110, 50, 60, 70, 80, 90, 100, 110, 50, 70, 80, 90, 100, 110, 120, 50, 60, 70, 80, 90, 100, 60, 70, 80, 90, 50, 100, 110, 70, 80, 110, 120, 70, 90

- Calcule a média, a mediana e a moda das dosagens.
- Determine o desvio padrão dessas dosagens.
- Construa um histograma para visualizar a distribuição.

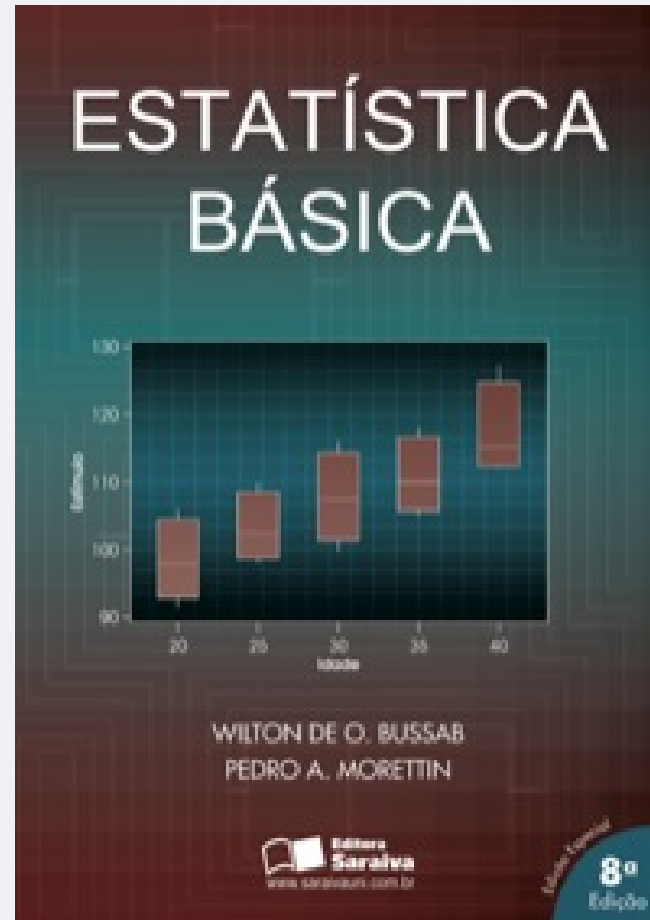
Dica para o tópico 3:

- Calcule a amplitude (AT)
- Calcule o número de classes com (k)
 - A **Regra de Sturges** propõe uma fórmula simples baseada no tamanho da amostra (n): $k = 1 + 3.322 \log_{10}(49) \approx 7$
 - A **Regra da Raiz Quadrada** é ainda mais simples: $k = \sqrt{49} = 7$
- Calcule o tamanho de cada intervalo (AI): $AI = \frac{AT}{k}$

Histograma com Comprimento de Intervalo pela Regra de Sturges

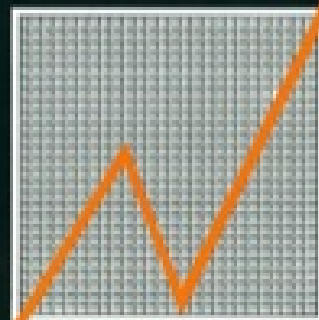


Referências



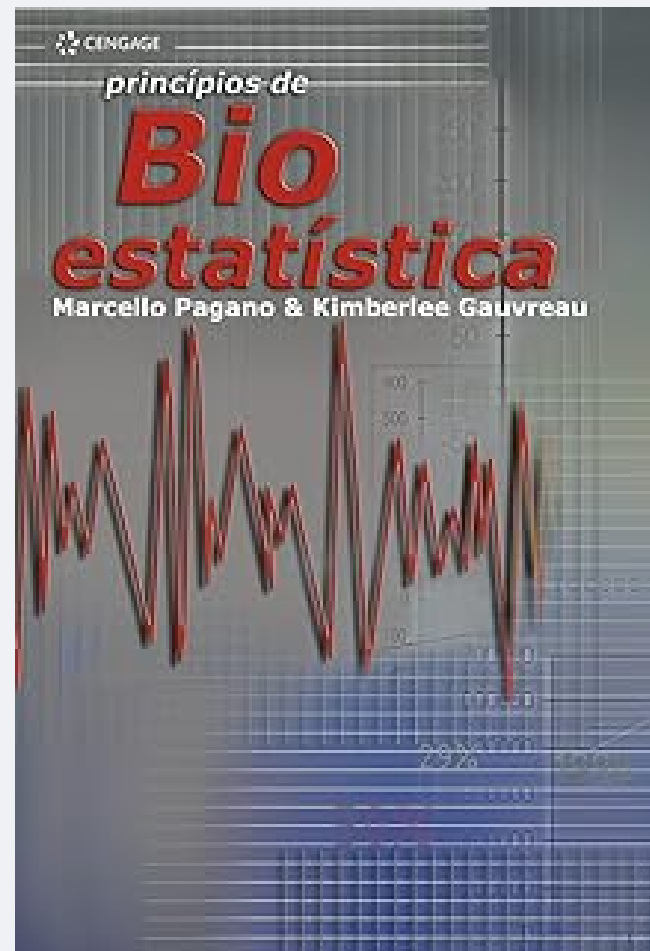
JAIRO SIMON DA FONSECA
GILBERTO DE ANDRADE MARTINS

CURSO DE ESTATÍSTICA



6ª edição

atlas



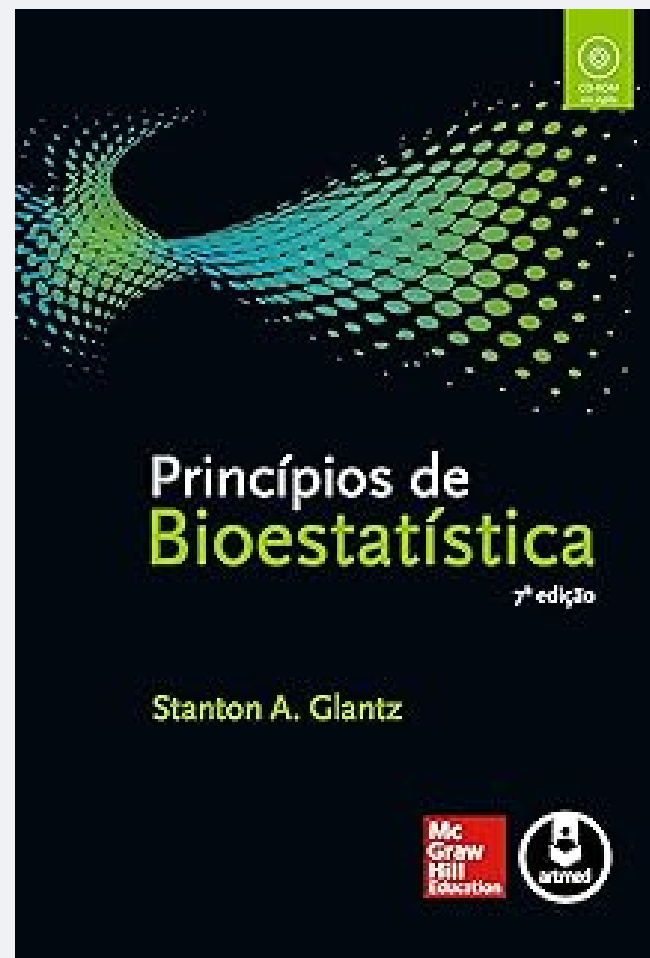
SONIA VIEIRA

INTRODUÇÃO À
Bioestatística



4ª EDIÇÃO







Referências complementares

- **Conceitos e análises estatísticas com R e JASP**
- **Regression Modelling for Biostatistics**
- **Amostragem: Teoria e prática usando o R**
- **R para Cientistas Sociais**

Não deixe de entrar em contato comigo para tirar suas dúvidas:
thiagoan.andrade@gmail.com

Obrigado!

Thanks!