

ESTATÍSTICA DESCRITIVA

- Organização
- Descrição
- Quantificação de variabilidade
- Identificação de valores típicos e atípicos

- Elementos básicos:
 - Tabelas
 - Gráficos
 - Resumos numéricos

CONCEITOS BÁSICOS

- Variável
 - Quantificação ou categorização do fenômeno de interesse (na Medicina chamado de parâmetro)

Inquérito epidemiológico com as perguntas:

- Qual é a sua idade? Variável: Idade
- Qual o número de pessoas da sua família? Variável: Tamanho da família
- Qual é a renda total de sua família? Variável: Renda familiar
- Qual é o seu estado civil? Variável: Estado civil
- Você tem emprego fixo? Variável: Emprego

Tipos de Variáveis

- Facilita o tratamento estatístico classificar variáveis em: Categóricas e Quantitativas
- **Variáveis Categóricas**
 - **Variáveis nominais:** Sexo, Tipo sanguíneo, Hábito de fumar
 - **Variáveis ordinais:** Grau de Escolaridade, Estadiamento, Níveis pressóricos

Tipos de Variáveis

- **Variáveis Quantitativas**
 - **Variáveis discretas:** Tamanho da família, Número de bactérias num volume de urina, Número de consultas durante um período, Número de batimentos cardíacos por minuto
 - **Variáveis contínuas:** Pressão sanguínea, peso, altura, Tempo de sobrevivência, Renda

V. Discreta x V. Categ. Ordinal

- A ordenação tem significado diferente:
 - **Estadiamento de câncer de mama (I, II, III e IV):** Não se pode dizer que IV é duas vezes pior do que II, ou que a diferença entre I e II é equivalente à entre III e IV.
 - **Número de crianças (0, 1, 2, 3, 4):** 4 crianças corresponde ao dobro de 2 crianças, e uma família com 4 crianças tem uma criança a mais do que uma família com 3, que por sua vez tem uma criança a mais do que uma família com 2 crianças.

DADOS BRUTOS

- Obtidos diretamente da pesquisa
- Ainda não sofreram processo de análise ou síntese
- Apresentados em tabelas mas omitidos das publicações por questão de espaço
- O conjunto de dados constitui uma ***amostra***. O tamanho da amostra é geralmente denotado por ***n***.

Exemplo: Teor de gordura fecal em crianças

- Utilidade diagnóstica do teor de gordura fecal inquestionável mas até 1984 não existia um padrão de referência para crianças brasileiras.
- Prof. Francisco Penna (titular de Pediatria da UFMG) examinou 43 crianças saudáveis

Tabela: Teor de gordura fecal (g/24 hs)

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

- Note a grande variação dos resultados!
- Como definir um padrão de referência?

Exemplo 3.2: Nível de colesterol

- 1948, cidade de Framingham selecionada para um estudo prospectivo
- **Objetivo:** verificar como hábitos de vida influenciam o risco de desenvolvimento doenças cardíacas.
- Necessidade de controle do nível de colesterol foram obtidos a partir deste estudo.

Tabela: taxa de colesterol (mg/dL) em 1952

278	182	247	227	277	194	196	276	244	192
118	219	255	201		209	219	228	209	209
171	213	233	226	209	200	200	363	209	200
179	167	192	277	317	146	217	292	217	255
212	233	250	243	150	209	184	199	250	479
175	194	221	233		184	217	150	167	265
242	180	255	170	209	161	196	165	234	179
248	184	291	185	242	276	243	229	242	250

- Observando a tabela o que apreendemos sobre o nível do colesterol à época do exame?
- Como saber o valor em torno do qual as medidas estão agrupadas, a forma da distribuição e a extensão da variabilidade?

Organização e apresentação de dados

- Para uma variável ou para o cruzamento de variáveis
 - Tabelas de frequências
 - Gráficos

Tabelas de frequências

- Sintetiza os dados através de uma **distribuição de frequências**
- Consiste na construção de uma tabela a partir dos dados brutos com a frequência com que cada observação ocorre.
- A partir das tabelas são construídos os gráficos.

Exemplo 3.2: Taxas de Colesterol (cont.)

- $n=78$, $\text{Min}=118$, $\text{Max}=479$
- Número de classes = $1 + \log(n, \text{base}=2) = 7,28 \approx 8$
- Tamanho de classe = $(479 - 118) / 8 = 45,125 \approx 50$

Tabela 3.10: Distribuição do nível de colesterol

Nível de Colesterol	Frequência absoluta		Frequência relativa	
	simples	acumulada	simples	acumulada
100 -150	2	2	0,03	0,03
150 -200	24	26	0,31	0,34
200 -250	35	61	0,45	0,79
250 -300	14	75	0,18	0,97
300 -350	1	76	0,01	0,98
350 -400	1	77	0,01	0,99
400 -450	0	77	0	0,99
450 -500	1	78	0,01	1
Total	78	-	1	-

Exemplo 3.5: Tentativas de suicídio

- Estudo retrospectivo (Fernandes et al., 1995)
- Tentativas de suicídio por intoxicação aguda registradas no Centro de Assistência Toxicológica do Hospital de Base de São Paulo.
- Período de 01/92 a 02/93
- 302 casos
- 27% do total de atendimentos no período
- 67% das tentativas de suicídio do sexo feminino

Tabela 3.3: Distribuição de profissões entre pacientes potencialmente suicidas



A small icon representing an OLE (Object Linking and Embedding) object, consisting of a square with the letters 'OLE' inside.

Tabela 3.4: Distribuição de tentativas de suicídio segundo faixa etária

A small icon with the letters 'OLE' inside a square frame, positioned in the center of a large, empty rectangular box. This likely indicates a missing or broken image or table content.

Etapas para construção de tabelas de frequências para dados agrupados

- 1. Encontrar o menor e o maior valores** (mínimo e máximo) do conjunto de dados
- 2. Escolher número de classes** (de igual amplitude), que englobem todos os dados sem superposição de intervalos.
- 3. Contar o número de elementos em cada classe** (este número é a **frequência absoluta**)
- 4. Calcular a frequência relativa** em cada classe

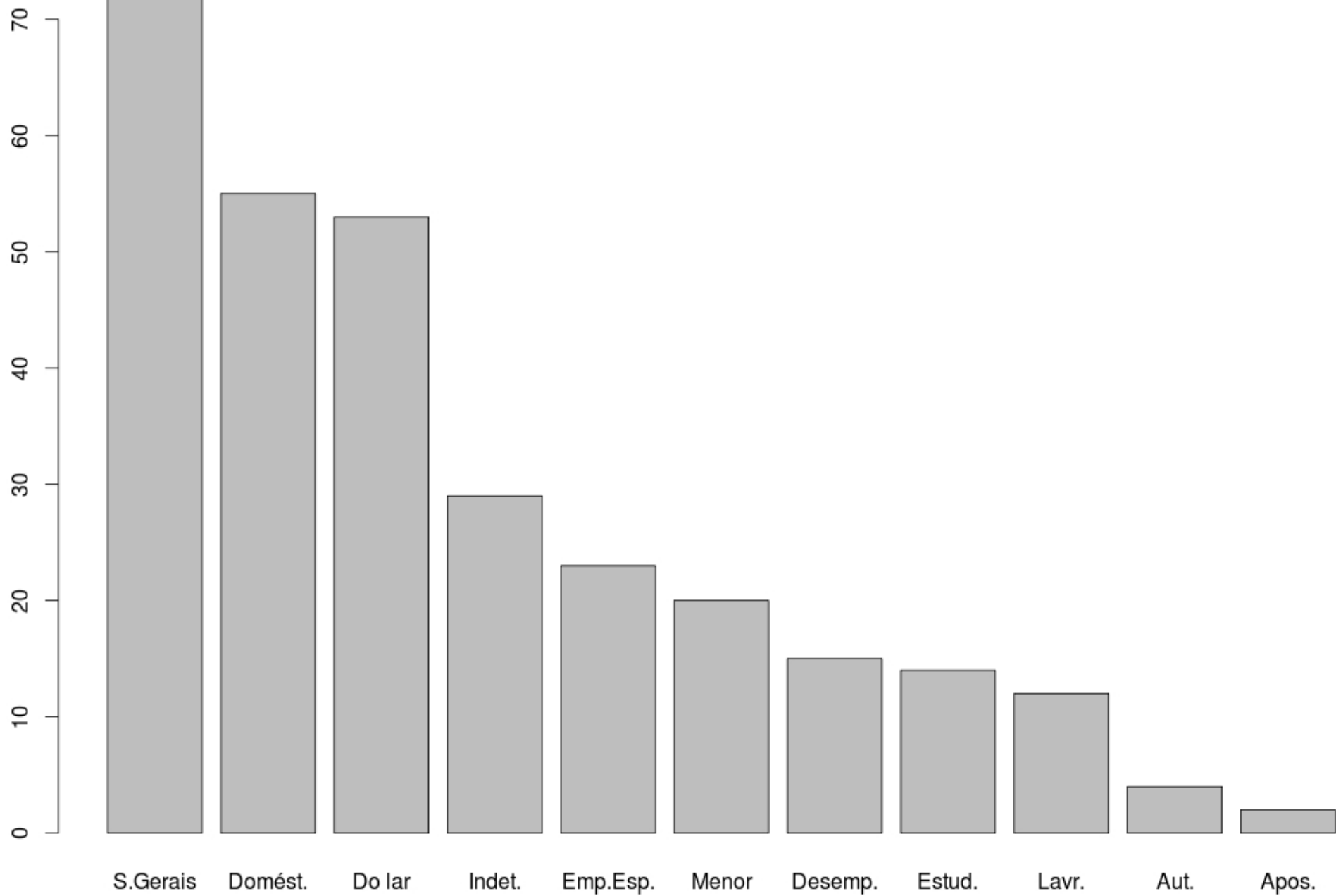
GRÁFICOS

- Diagrama de barras
- Histograma
- Ogiva
- Gráfico de linhas
- Diagrama de pontos
- Diagrama de dispersão

Representação gráfica para variáveis categóricas

- Diagrama de barras
- Exemplo 3.5: Distribuição de profissões entre pacientes potencialmente suicidas (cont.)

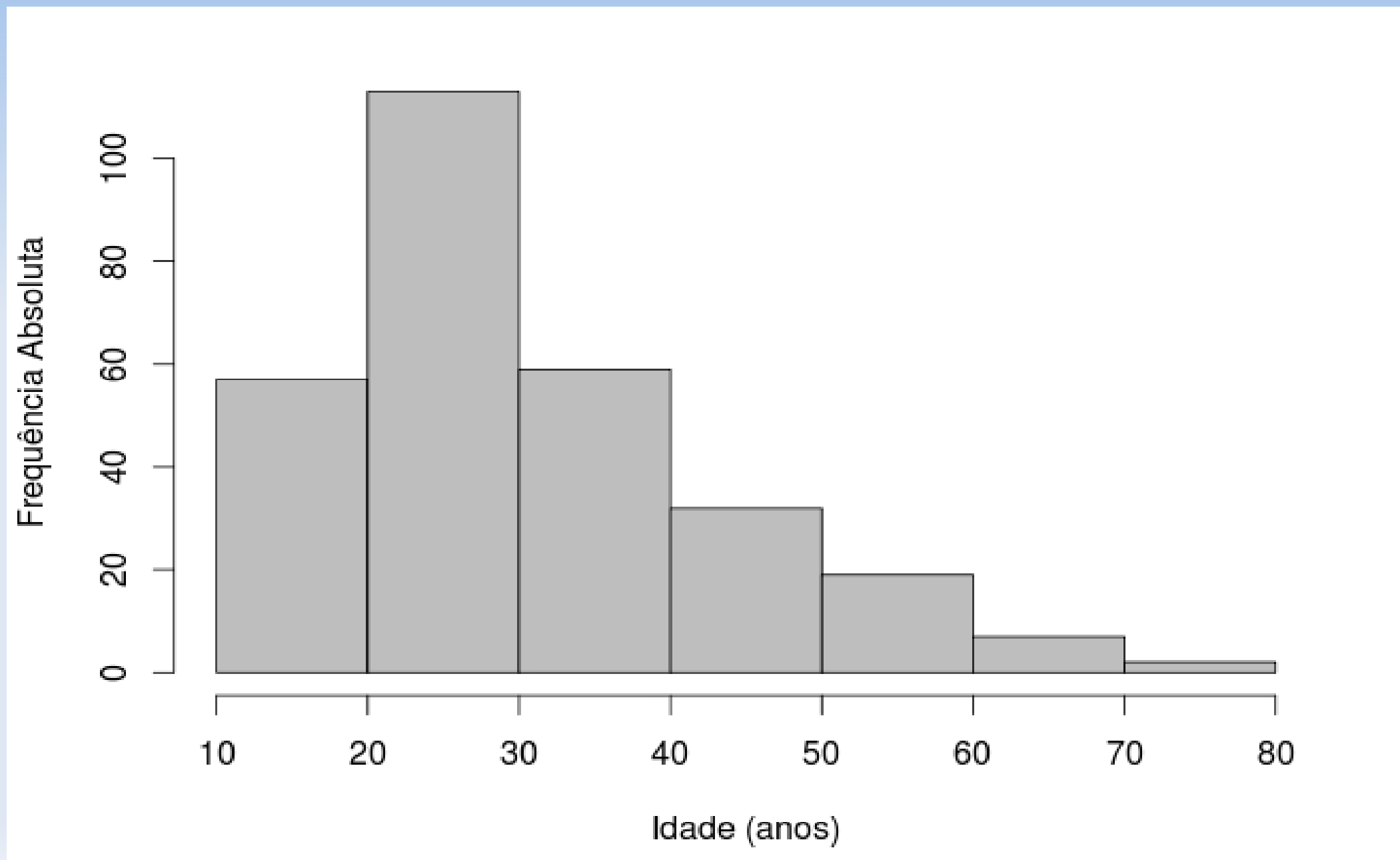
Frequência Absoluta



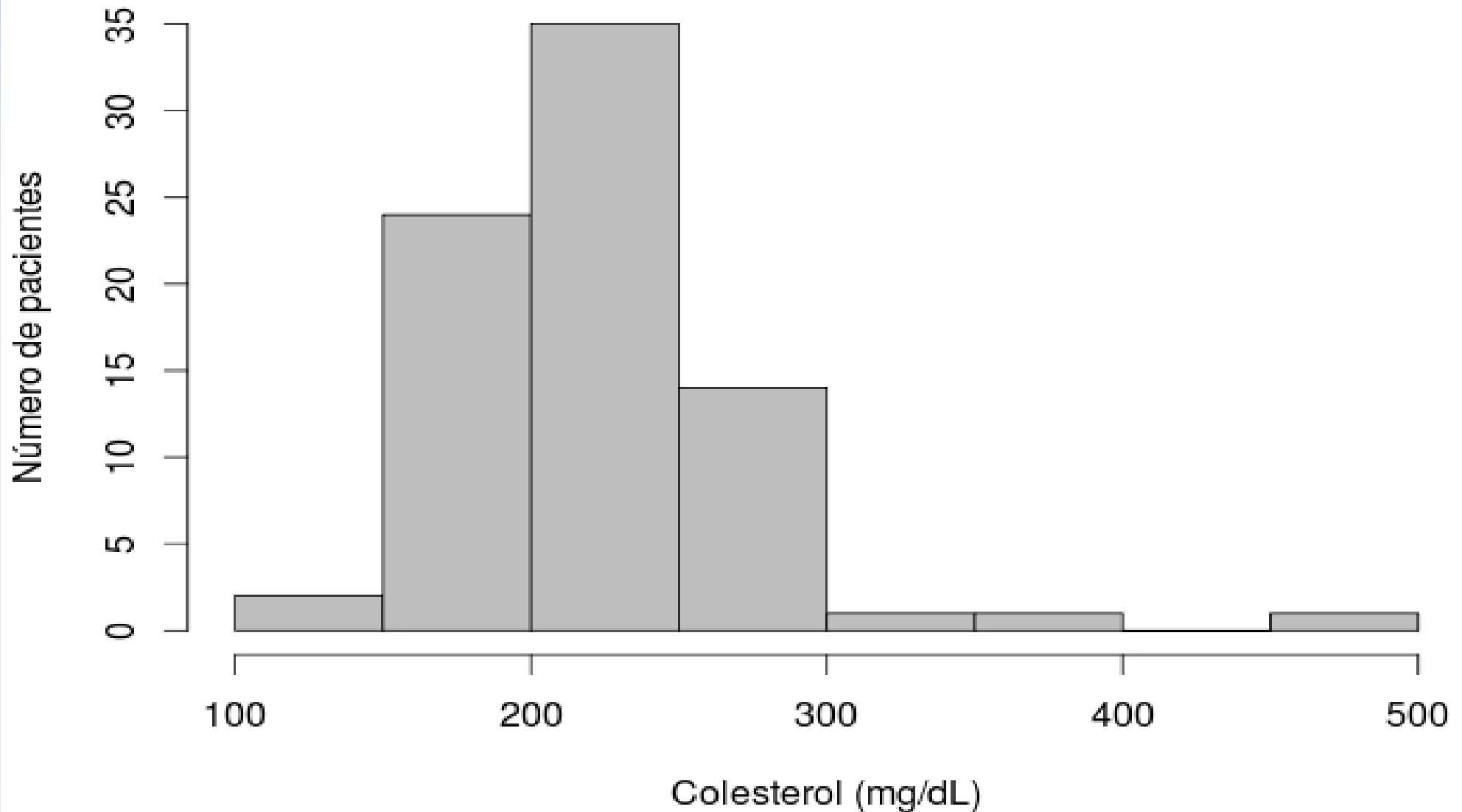
Representação gráfica de variáveis quantitativas

- Histograma
 - Serve para visualizar a forma da distribuição da variável estudada.

Exemplo 3.5: Distribuição de tentativas de suicídio segundo faixa etária



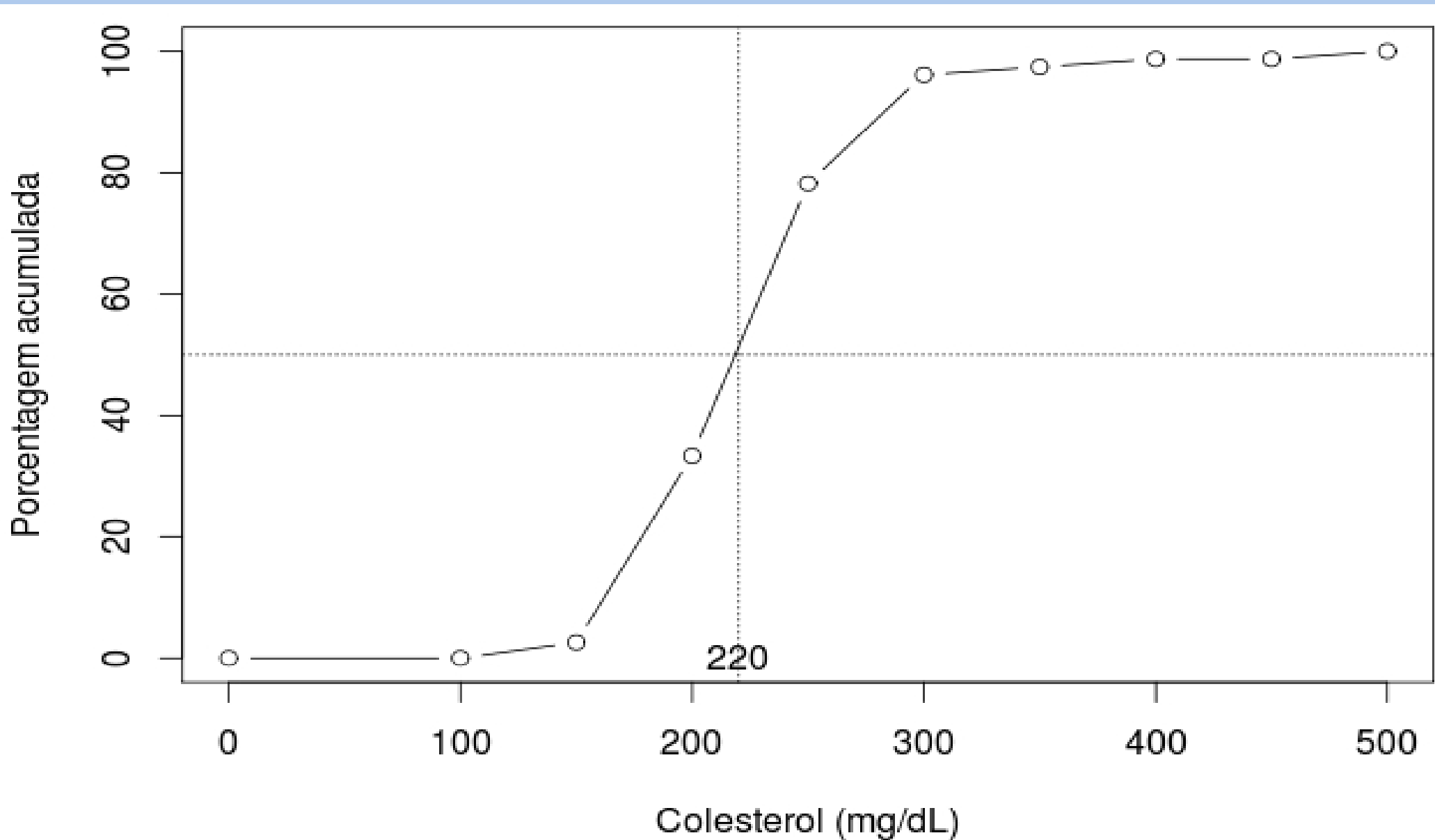
Exemplo 3.2: Distribuição do nível de colesterol



Ogiva

- Gráfico de percentuais acumulados
- Através da ogiva podemos **estimar percentis** da distribuição, isto é, o valor que é precedido por uma porcentagem pré-estabelecida.
- Exemplo: estimar o valor da variável abaixo do qual se tem 50% dos indivíduos.

Ogiva das taxas de colesterol



Representação gráfica de dados temporais

- Dados coletados ao longo do tempo são comuns em pesquisas médicas
- Diagrama de barras para períodos agrupados (ex: menos de 1 ano, 1 a 5 anos, 5 a 10 anos)
- **Gráfico de linhas** é o mais apropriado
 - Eixo horizontal: escala temporal
 - Eixo vertical: variável de interesse
- Permite constatar tendências e identificar eventos extremos

Coleta de dados

- Há várias maneiras de se coletar dados, dependendo do tipo de estudo
- **Experimento com cobaias:** medida direta das variáveis de interesse
- **Inquérito:** questionário é o instrumento de medida mais utilizado
- **Pesquisas clínicas:** formulário, prontuário do paciente, ou ficha preenchida na anamnese.
- Referências na Página 64 do livro texto.

Banco de dados

- Uma linha para cada indivíduo
- Uma coluna para cada variável observada
- Para variáveis categóricas:
 - Criar códigos para cada categoria
- Para variáveis contínuas:
 - Entrar com os dados originais e não os codificados para classes de interesse, pois pode haver mudança nas classes de interesse durante a análise
- Para dados omissos: usar código que facilmente identifique esse tipo de dado (Ex: 999 para pressão arterial)

Exemplo 3.20: Tentativas de suicídio (cont.)

- Dicionário das variáveis:
- Sexo: 0 para masculino e 1 para feminino
- Profissão: 1-Serviços Gerais, 2-Doméstica, 3-Do lar, 4-Indeterminado, 5-Emprego Especializado, 6-Menor, 7-Desempregado, 8-Estudante, 9-Lavrador, 10-Autônomo, 11-Aposentado
- Idade: anos

Indivíduo	Sexo	Profissão	Idade
1	0	1	25
2	1	2	48
...
302	1	8	13