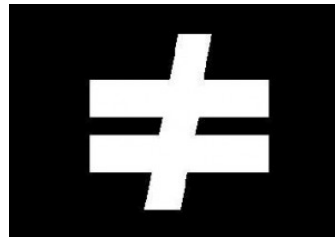


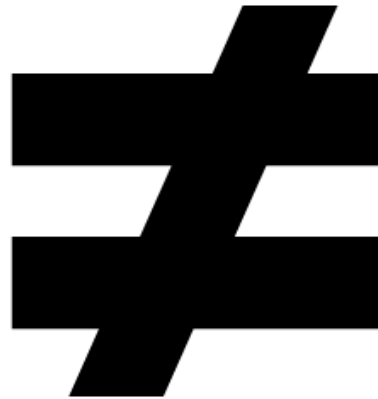
ESTATÍSTICAS E PARÂMETROS

POPULAÇÃO



AMOSTRA

**VALORES OBTIDOS
DA POPULAÇÃO**



**VALORES OBTIDOS
DA AMOSTRA**

ESTATÍSTICAS E PARÂMETROS

Parâmetro

É uma característica numérica estabelecida para toda uma população (valor singular que caracteriza a população).

Exemplo: A média da altura de alunos de uma certa turma é 1,72 cm.

Estimativa

É uma característica numérica estabelecida para uma amostra (valor aproximado do parâmetro, sendo obtido a partir de medidas na amostra).

ESTATÍSTICAS E PARÂMETROS

Quando você ouve no noticiário que, de acordo com a pesquisa de determinado instituto, 44% dos brasileiros aprovam determinada atitude do Presidente da República, você foi apresentado a uma **ESTATÍSTICA**

As pessoas que compuseram a amostra (provavelmente 1.500 ou 2.000) pensam da atitude em questão.

É um indicador ou uma estimativa do parâmetro correspondente a porcentagem da população brasileira que aprovou a atitude.

ESTATÍSTICAS E PARÂMETROS

COM QUANTAS UNIDADES SE COMPÕE UMA AMOSTRA?

Do ponto de vista do estatístico, as amostras devem ser GRANDES para dar maior confiança às conclusões obtidas.



A "qualidade" de uma estimativa depende, em muito, do número de unidades que compõe a amostra (tamanho da amostra).

Como se determina o tamanho da amostra?

Na prática

Determinado mais por considerações reais ou imaginárias a respeito do custo de cada unidade amostrada do que por técnicas estatísticas (orçamento curto X pesquisa ambiciosa)



na sua área

mas verifique também o que

fazer

veja o que se faz

seu orçamento permite

Como se determina o tamanho da amostra?

Estatisticamente

Determinado por alguns critérios (fórmulas de cálculo)

Para essa disciplina será
apresentada apenas uma
equação que dará ideia do
problema



Como se determina o tamanho da amostra?

Imagine que um antropólogo está estudando os habitantes de uma ilha isolada e que, entre outras coisas, quer determinar a porcentagem de pessoas dessa ilha com sangue tipo O.

Quantas pessoas (tamanho da amostra) devem ser examinadas?



Como se determina o tamanho da amostra?

1ª QUESTÃO

qual é a margem de erro que o antropólogo admite em seus resultados?

- ❖ Vamos imaginar que ele diz ficar satisfeito com uma margem de erro de $\pm 5\%$, isto é, se 43% das pessoas da amostra tiverem sangue tipo O.

A verdadeira porcentagem de pessoas com sangue tipo O na ilha deverá estar entre 38 e 48, ou seja, no intervalo 43% ($\pm 5\%$).

Como se determina o tamanho da amostra?

2ª QUESTÃO

como estará trabalhando com uma só amostra, existe a chance de ele, por azar, tomar uma amostra pouco representativa. O antropólogo então concorda em admitir a probabilidade de uma amostra errada em cada 20

Isto significa que ele terá probabilidade $(19/20)=0,95$ de obter a verdadeira porcentagem de sangue tipo O dentro do intervalo calculado. Temos então o nível de confiança: 95%.

margem de erro e o termo adicionado e subtraído do estimador para formar um intervalo de confiança

$$Z = 1,65 = 90\%$$

$$Z = 1,96 = 95\%$$

$$Z = 2,0 = 95.5\%$$

$$Z = 2,57 = 99\%$$

Como se determina o tamanho da amostra?

3ª QUESTÃO

qual o valor que o antropólogo espera para a porcentagem de pessoas com sangue tipo O na ilha.

- ❖ Ele diz que, com base no que sabe de outras populações, é razoável esperar que essa porcentagem esteja entre 30% e 60%.

Admitiremos, por simplicidade, que essa porcentagem seja 50%

Como se determina o tamanho da amostra?

$$n = \frac{z^2 p(100 - p)}{d^2}$$

Z (valor dado em tabela e associado ao nível de confiança) = 1,96~2

p (porcentagem de pessoas com sangue tipo O na ilha)

d (margem de erro em porcentagem)

Como se determina o tamanho da amostra?

$$n = \frac{2^2 50(100 - 50)}{5^2}$$

$$n = \frac{4 \times 50 \times 50}{25} = 400$$

Então o tamanho da amostra deve ser de 400 pessoas

A QUESTÃO DA REPRESENTATIVIDADE

- ❑ A amostra só traz informação sobre a população da qual foi retirada;
- ❑ A amostra deve ter o tamanho usual da área em que a pesquisa se enquadra (amostras pequenas x amostras grandes).

Amostras representativas ou não-representativas

Uma amostra é dita representativa se as suas características se assemelham o mais possível às da população alvo.

Uma amostra é representativa se as unidades que a constituem foram escolhidas por um processo tal que todos os membros da população tenham a mesma probabilidade de fazer parte da amostra.

Variável é uma condição ou característica das unidades da população; pode assumir valores diferentes em diferentes unidades

Dados são os valores da variável em estudo, obtidos por meio de uma amostra.

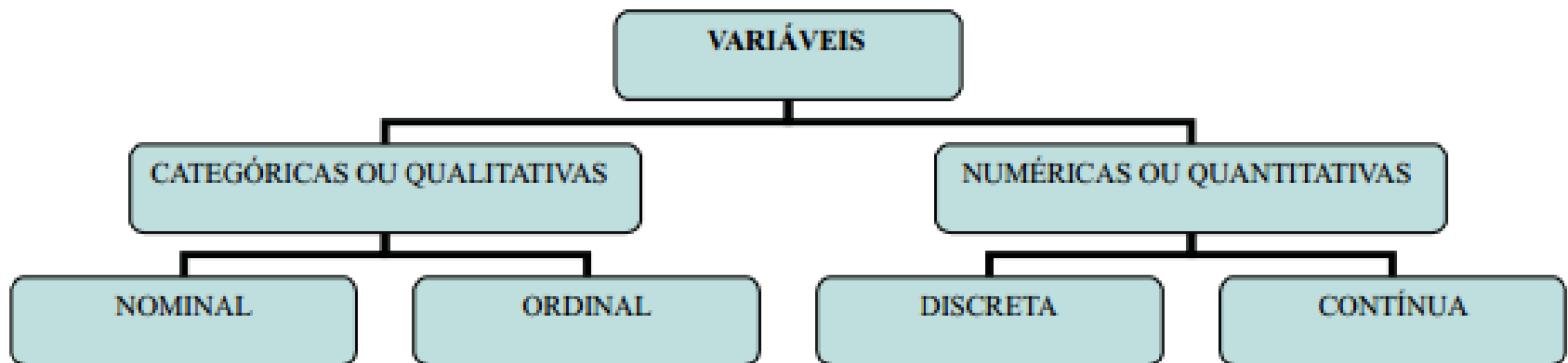
O dono de uma academia de ginástica quer saber a opinião de seus clientes sobre a qualidade dos serviços que presta. O que é variável e o que são dados nesse problema?

Solução

A *variável* de interesse é a opinião dos clientes. Os *dados* serão obtidos somente quando o dono da academia começar a pedir aos clientes que dêem uma nota a cada serviço. Então, se for pedido que o cliente dê uma nota de zero e 5 a cada serviço que utiliza — os dados coletados poderão ser, por exemplo, 4, 3, 2, 4, 1 etc., por serviço.

DADOS E VARIÁVEIS

As **variáveis** podem ter valores numéricos e não numéricos e são classificadas como:



DADOS E VARIÁVEIS

1. Variáveis qualitativas (ou categóricas)

são as características que não possuem valores quantitativos, mas, ao contrário, são definidas por várias categorias, ou seja, representam uma classificação dos indivíduos.

Variáveis nominais: quando o seu significado só se entende em função do nome e do número ou código que se lhe atribua, não nos dá nenhuma informação

Exemplo:
Sexo (masculino ou feminino)
Raça (branco, negro, pardo)

Variáveis ordinais: Constituem valores relativos, atribuídos para denotar ordem

Exemplo:
escolaridade (fundamental, médio e superior)
estágio da doença (inicial, intermediário, terminal)

DADOS E VARIÁVEIS

2. Variáveis quantitativas (numéricas)

são as características que podem ser medidas em uma escala quantitativa, ou seja, apresentam valores numéricos.

Variáveis discretas: são aquelas que só podem assumir valores inteiros dentro de um intervalo de interesse. Os dados discretos são resultados da contagem do número de itens referente à variável

Exemplo:
número de filhos
quantidade de consultas médicas

Variáveis contínuas: são aquelas que podem assumir qualquer valor dentro de um intervalo de interesse. Em geral, estão associadas a medidas que tenham unidade

Exemplo:
peso (balança)
altura (régua)
distância percorrida

APURAÇÃO DE DADOS

Dados são registrados em fichas, em cadernos, em computador.

Se a variável for **qualitativa**, a apuração se resume a simples contagem

Para estudar a razão de sexos¹ dos recém-nascidos em uma maternidade e seus pesos ao nascer, um pesquisador obteve uma amostra sistemática de 1.000 prontuários de recém-nascidos e escreveu numa folha de papel:

Masculino

Feminino

Depois examinou todos os prontuários e fez, então, um traço na linha que indicava cada sexo, toda vez que o prontuário registrava que o recém-nascido era desse sexo. Cada quadrado, cortado pela diagonal, representa cinco recém-nascidos. O total é dado pelo número de traços em cada linha.

Masculino ☒ ☒ ... ☒ ☐ = 509

Feminino ☒ ☒ ... ☐ | = 491

APURAÇÃO DE DADOS

Quando a variável for **quantitativa** é preciso anotar cada valor observado

Para apurar dados de peso ao nascer², o pesquisador deve anotar o número do prontuário e o peso ao nascer, numa folha de papel. O número do prontuário, escrito ao lado do peso ao nascer, facilita a posterior verificação da apuração.

<i>Nº do prontuário</i>	<i>Peso ao nascer</i>
10 525	3,250
10 526	2,010
.	.
.	.
.	.
10 624	2,208

A apuração de peso ao nascer pode ser feita por sexo se o interesse é comparar pesos ao nascer de meninos e meninas

COMPONENTES DAS TABELAS

- ❑ Os dados devem ser apresentados em tabelas construídas de acordo com as normas técnicas ditadas pela Fundação Instituto Brasileiro de Geografia e Estatística (IBGE) e NBR 14724:11 da ABNT.;
- ❑ As tabelas devem ser colocadas perto do ponto do texto em que são mencionadas pela primeira vez;
- ❑ Devem ser inseridas na ordem em que aparecem no texto.

COMPONENTES DAS TABELAS

De acordo com essas normas, uma tabela deve ter título, corpo, cabeçalho e coluna indicadora.

O *título* explica o que a tabela contém. O *corpo* é formado pelos dados, em linhas e colunas. O *cabeçalho* especifica o conteúdo das colunas. A *coluna indicadora* especifica o conteúdo das linhas.

TABELA 2.1

População residente no Brasil, segundo o sexo, de acordo com o censo demográfico de 2000.

<i>Sexo</i>	<i>População residente</i>
Masculino	83.576.015
Feminino	86.223.155
Total	169.799.170

Fonte: IBGE (2003)⁹

Na Tabela 2.1, observe o título:

População residente no Brasil, segundo o sexo, de acordo com o censo demográfico de 2000.

O cabeçalho é constituído por:

<i>Sexo</i>	<i>População residente</i>
--------------------	-----------------------------------

COMPONENTES DAS TABELAS

Toda tabela deve ser delimitada por traços *horizontais*, mas não deve ser delimitada por traços *verticais*.

Os traços verticais podem ser feitos somente para separar as colunas.

O cabeçalho deve ser separado do corpo da tabela por um traço horizontal.

COMPONENTES DAS TABELAS

As tabelas podem conter *fonte* e *notas*.

Fonte é a entidade, ou pesquisador, ou pesquisadores que publicaram ou forneceram os dados;

Notas esclarecem aspectos relevantes do levantamento dos dados ou da apuração

COMPONENTES DAS TABELAS

TABELA 2.2

Número de internações hospitalares, de mulheres, pelo Sistema Único de Saúde (SUS). Brasil, 2005.

<i>Grupo de doenças</i>	<i>Número</i>
Gravidez, parto e puerpério	2.640.438
Doenças do aparelho respiratório	736.012
Doenças do aparelho circulatório	612.415
Doenças do aparelho geniturinário	507.295
Doenças infecciosas e parasitárias	480.165
Doenças do aparelho digestivo	452.894
Transtornos mentais e comportamentais	105.354
Neoplasias	355.570
Causas externas	233.787
Demais causas	801.123
Total	6.925.053

Fonte: Ministério da Saúde/SE/Datasus — Sistema de Informações Hospitalares do SUS — SIH/SUS.

Nota: Suprimidos os casos com idade ou local de residência ignorados.

APRESENTAÇÃO DE DADOS QUALITATIVOS

- ❑ Quando observamos dados qualitativos, classificamos cada unidade da amostra em uma dada categoria;
- ❑ Nosso conhecimento sobre os dados aumenta se contarmos quantas unidades caem em cada categoria;
- ❑ A ideia seguinte é resumir as informações na forma de uma tabela que mostre as contagens (frequências) em cada categoria.

TABELA DE DISTRIBUIÇÃO DE FREQUÊNCIAS

APRESENTAÇÃO DE DADOS QUALITATIVOS

Foram entrevistados 2.500 brasileiros, com 16 anos ou mais, para saber a opinião deles sobre determinado técnico de futebol. Veja o que eles responderam: 1.300 achavam que o técnico era bom, 450 achavam regular e 125 achavam ruim; 625 não tinham opinião ou não quiseram opinar. Como se organizam estes dados em uma tabela de distribuição de freqüências?

Solução

Na Tabela 2.3 estão as respostas dadas pelos entrevistados (primeira coluna) e as freqüências dessas respostas (segunda coluna). A soma das freqüências é 2.500 (número de entrevistados).

TABELA 2.3

Opinião dos brasileiros sobre determinado técnico de futebol.

<i>Respostas</i>	<i>Freqüência</i>
Bom	1.300
Regular	450
Ruim	125
Não sabe	625
Total	2.500

APRESENTAÇÃO DE DADOS QUALITATIVOS

Nas tabelas de distribuição de frequências, é usual fornecer a **proporção (frequência relativa)** de unidades que caem em cada categoria.

Para obter a frequência relativa de uma dada categoria:

$$\text{Frequência relativa} = \frac{\text{frequência}}{\text{tamanho da amostra}}$$

APRESENTAÇÃO DE DADOS QUALITATIVOS

Solução

Na Tabela 2.4 estão as respostas dadas pelos entrevistados (primeira coluna), as freqüências dessas respostas (segunda coluna) e as freqüências relativas (terceira coluna). Note que as freqüências relativas somam 1,00.

TABELA 2.4

Opinião dos brasileiros sobre determinado técnico de futebol.

<i>Respostas</i>	<i>Freqüência</i>	<i>Freqüência relativa</i>
Bom	1.300	$\frac{1300}{2500} = 0,52$
Regular	450	$\frac{450}{2500} = 0,18$
Ruim	125	$\frac{125}{2500} = 0,05$
Não sabe	625	$\frac{625}{2500} = 0,25$
Total	2.500	1,00

APRESENTAÇÃO DE DADOS QUALITATIVOS



As frequências relativas são, em geral, dadas em porcentagens. Para transformar uma frequência relativa em porcentagem, basta multiplicar por 100.

As frequências relativas dadas em porcentagens fornecem a informação mais relevante. Mas sempre convém exibir o total (tamanho da amostra), que é indicador da credibilidade da informação dada

APRESENTAÇÃO DE DADOS QUALITATIVOS

TABELA DE CONTINGÊNCIA

os elementos da amostra ou da população são classificados de acordo com duas variáveis qualitativas.

Os dados devem então ser apresentados em tabelas de contingência (tabelas de dupla entrada, cada entrada relativa a uma das variáveis)

APRESENTAÇÃO DE DADOS QUALITATIVOS

Foram feitos diagnósticos de depressão em 500 estudantes com idades entre 10 e 17 anos, metade de cada sexo. Foram identificados 98 casos de depressão, sendo 62 no sexo feminino. Apresente os dados em uma tabela.

Solução

Note que os dados estão classificados segundo duas variáveis: sexo e presença de depressão.

TABELA 2.5
Sexo e presença de depressão.

<i>Sexo</i>	<i>Depressão</i>		<i>Total</i>
	<i>Sim</i>	<i>Não</i>	
Masculino	36	214	250
Feminino	62	188	250
Total	98	402	500

APRESENTAÇÃO DE DADOS QUALITATIVOS

TABELA DE CONTINGÊNCIA

As tabelas de contingência podem apresentar **frequências relativas** em porcentagens, além das frequências

Para verificar se o risco de óbito neonatal é maior quando a gestante é diabética, foram obtidos os dados apresentados na Tabela 2.6. Discuta.

TABELA 2.6
Óbito neonatal e diabetes mellitus.

<i>Gestante</i>	<i>Óbito neonatal</i>		<i>Total</i>	<i>Percentual de óbitos</i>
	<i>Sim</i>	<i>Não</i>		
Diabética	3	21	24	12,5%
Não-diabética	21	830	851	2,5%
Total	24	851	875	

O risco de óbito neonatal, dado pelo percentual de óbitos, é maior quando a gestante é diabética.

APRESENTAÇÃO DE DADOS NÚMERICOS

Os dados numéricos são apresentados na ordem em que são coletados;

Geralmente são obtidos dados relativos a diversas variáveis em cada paciente;

Os pacientes são identificados nas pesquisas por números.

Para estudar o desempenho cardíaco de pacientes submetidos à diálise renal, foram obtidos valores de diversas variáveis de interesse da Cardiologia. Na Tabela 2.7 são apresentadas apenas algumas informações, para mostrar como se apresentam dados numéricos.

APRESENTAÇÃO DE DADOS NÚMERICOS

TABELA 2.7

Idade em anos completos, tempo de diálise em meses, altura em metros, peso em quilogramas, pressão sistólica e diastólica em milímetros de mercúrio de mulheres submetidas à diálise renal.

<i>Número da paciente</i>	<i>Idade</i>	<i>Tempo de diálise</i>	<i>Altura</i>	<i>Peso</i>	<i>Pressão sistólica</i>	<i>Pressão diastólica</i>
1	45	14	1,60	62,0	140	85
2	62	54	1,65	52,5	100	70
3	38	52	1,55	67,8	140	100
4	26	34	1,59	48,2	165	105
5	35	18	1,58	46,0	170	105
6	44	71	1,48	40,4	150	100
7	53	39	1,69	67,7	155	95
8	44	79	1,59	55,5	160	105
9	58	23	1,62	63,0	175	110
10	55	64	1,51	50,3	155	105
11	24	16	1,79	77,0	160	95
12	70	46	1,51	44,0	150	95
13	56	48	1,58	64,0	175	110

APRESENTAÇÃO DE DADOS NÚMERICOS

Dados numéricos também podem ser apresentados em tabelas de distribuição de frequências;

Se os dados são discretos, para organizar a tabela de distribuição de frequências:

- escreva os dados em ordem crescente;
- conte quantas vezes cada valor se repete;
- organize a tabela como já foi feito para dados qualitativos, colocando no lugar das categorias, os valores numéricos em ordem natural.

APRESENTAÇÃO DE DADOS NÚMERICOS

As faltas ao trabalho de 30 empregados de uma clínica em determinado semestre estão na Tabela 2.8. A partir dela, faça uma tabela de distribuição de frequências.

TABELA 2.8

Número de faltas dadas por 30 empregados de uma clínica no semestre.

1	3	1	1	0	1	0	1	1	0
2	2	0	0	0	1	2	1	2	0
0	1	6	4	3	3	1	2	4	0

Solução

TABELA 2.9

Distribuição do número de faltas de 30 empregados de uma clínica no semestre.

<i>Número de faltas</i>	<i>Frequência</i>	<i>Percentual</i>
0	9	30,0
1	10	33,3
2	5	16,7
3	3	10,0
4	2	6,7
5	0	0,0
6	1	3,3
Total	30	100,0

APRESENTAÇÃO DE DADOS NÚMERICOS

Tabelas com grande número de dados não oferecem ao leitor visão rápida e global do fenômeno;

Por esta razão, **dados contínuos** (desde que em grande número) são apresentados em tabelas de distribuição de frequências.

Exemplo 2.10: Uma tabela com dados contínuos.

TABELA 2.10

Peso ao nascer de nascidos vivos, em quilogramas.

2,522	3,200	1,900	4,100	4,600	3,400
2,720	3,720	3,600	2,400	1,720	3,400
3,125	2,800	3,200	2,700	2,750	1,570
2,250	2,900	3,300	2,450	4,200	3,800
3,220	2,950	2,900	3,400	2,100	2,700
3,000	2,480	2,500	2,400	4,450	2,900
3,725	3,800	3,600	3,120	2,900	3,700
2,890	2,500	2,500	3,400	2,920	2,120
3,110	3,550	2,300	3,200	2,720	3,150
3,520	3,000	2,950	2,700	2,900	2,400
3,100	4,100	3,000	3,150	2,000	3,450
3,200	3,200	3,750	2,800	2,720	3,120
2,780	3,450	3,150	2,700	2,480	2,120
3,155	3,100	3,200	3,300	3,900	2,450
2,150	3,150	2,500	3,200	2,500	2,700
3,300	2,800	2,900	3,200	2,480	-
3,250	2,900	3,200	2,800	2,450	-

APRESENTAÇÃO DE DADOS NÚMERICOS

Para construir uma tabela de distribuição de frequências com dados contínuos:

- Ache o *valor máximo* e o *valor mínimo* do conjunto de dados;
- Calcule a *amplitude*, que é a diferença entre o valor máximo e o valor mínimo;
- Divida a amplitude dos dados pelo número de faixas que pretende organizar. Essas faixas recebem, tecnicamente, o nome de *classes*;
- O resultado da divisão é o *intervalo de classe*. É sempre melhor arredondar esse número para um valor mais alto, o que facilita o trabalho;
- Organize as *classes*, de maneira que a primeira contenha o menor valor observado.

APRESENTAÇÃO DE DADOS NÚMERICOS

Exemplo 2.10: Uma tabela com dados contínuos.

TABELA 2.10
Peso ao nascer de nascidos vivos

2,522	3,200	1,900			
2,720	3,720	3,600	2,400		3,400
3,125	2,800	3,200	2,700	2,750	1,570
2,250	2,900	3,300	2,450	4,200	3,800
3,220	2,950	2,900	3,400	2,100	2,700
3,000	2,480	2,500	2,400	4,450	2,900
3,725	3,800	3,600	3,120	2,900	3,700
2,890	2,500	2,500	3,400	2,920	2,120
3,110	3,550	2,300	3,200	2,720	3,150
3,520	3,000	2,950	2,700	2,900	2,400
3,100	4,100	3,000	3,150	2,000	3,450
3,200	3,200	3,750	2,800	2,720	3,120
2,780	3,450	3,150	2,700	2,480	2,120
3,155	3,100	3,200	3,300	3,900	2,450
2,150	3,150	2,500	3,200	2,500	2,700
3,300	2,800	2,900	3,200	2,480	-
3,250	2,900	3,200	2,800	2,450	-

Estabelecer
de 5 a 20
classes

$$k = \sqrt{n}$$

menor valor é 1,570 kg e o maior valor é 4,500 kg

■ A amplitude dos dados é:
 $4,500 - 1,570 = 3,030$

■ Vamos definir 7 classes. Então:
 $3,030 : 7 = 0,433 \sim 0,5$

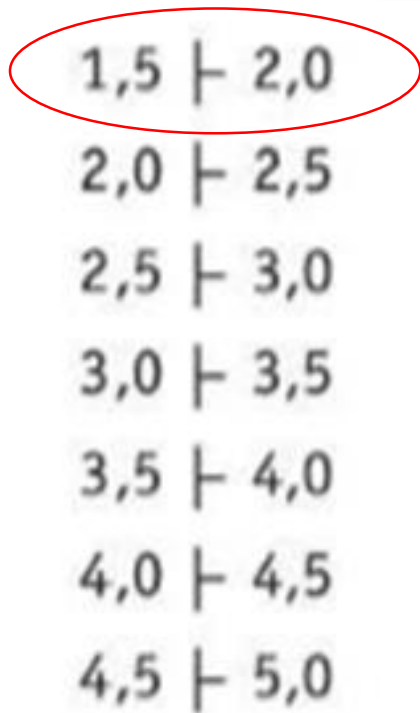
■ Construa a primeira classe, que será de 1,5 kg a 2,0 kg (esta classe contém o menor valor); depois, construa a segunda classe, que será de 2,0 kg a 2,5 kg, e assim por diante:

1,5 | 2,0
2,0 | 2,5
2,5 | 3,0
3,0 | 3,5
3,5 | 4,0
4,0 | 4,5
4,5 | 5,0

APRESENTAÇÃO DE DADOS NÚMERICOS

OBSERVAÇÃO

Denominam-se *extremos de classe* os limites dos intervalos de classe.



1,5 2,0
2,0 2,5
2,5 3,0
3,0 3,5
3,5 4,0
4,0 4,5
4,5 5,0

O intervalo é fechado à esquerda, isto é, pertencem à classe os valores iguais ao extremo inferior da classe.

Também significa que o intervalo é aberto à direita, isto é, não pertencem à classe os valores iguais ao extremo superior.

APRESENTAÇÃO DE DADOS NÚMERICOS

Exemplo 2.11: Uma tabela de distribuição de freqüências para dados contínuos.

Para dar idéia geral sobre peso ao nascer de nascidos vivos, o pesquisador quer apresentar não os pesos observados — mas o número de nascidos vivos por faixas de peso. A Tabela 2.11 apresenta a distribuição de freqüências.

TABELA 2.11
Distribuição de freqüências para peso ao nascer de nascidos vivos, em quilogramas.

<i>Classe</i>	<i>Freqüência</i>
1,5 2,0	3
2,0 2,5	16
2,5 3,0	31
3,0 3,5	34
3,5 4,0	11
4,0 4,5	4
4,5 5,0	1

APRESENTAÇÃO DE DADOS NÚMERICOS

Numa distribuição de frequências, o *extremo inferior* da primeira classe, o *extremo superior* da última classe ou ambos podem não estar definidos.

Exemplo 2.13: Uma tabela de distribuição de frequências para dados contínuos com classes de tamanhos diferentes e extremo superior não definido.

Para dar uma idéia geral sobre pressão sanguínea sistólica de mulheres com 30 anos de idade, o pesquisador apresentou não os valores observados — mas o número de mulheres por faixas de pressão. Veja a Tabela 2.12, que também é um exemplo em que o extremo superior da última classe não está definido.

APRESENTAÇÃO DE DADOS NÚMERICOS

TABELA 2.12

Distribuição de freqüências para a pressão sanguínea sistólica, em milímetros de mercúrio, de mulheres com 30 anos de idade.

<i>Classe</i>	<i>Freqüência</i>
90 100	6
100 105	11
105 110	12
110 115	17
115 120	18
120 125	11
125 130	9
130 135	6
135 140	4
140 150	4
150 160	1
160 e mais	1

As tabelas de distribuição de freqüências mostram a distribuição da variável, mas perdem em exatidão