

Capítulo

5

Distribuição de frequência

Objetivos

- Apresentar a forma de sintetizar dados qualitativos quantitativos.
- Fornecer o passo a passo da construção das tabelas de distribuição de frequência simples e para dados agrupados.
- Mostrar representações gráficas: histograma, polígono de frequência e ogiva.

1. Sintetizando dados qualitativos

Os dados qualitativos são o resultado da análise de variáveis qualitativas. A análise estatística desse tipo de dados resume-se à contagem do número de indivíduos em cada categoria e ao cálculo das respectivas porcentagens.

Quadro 1

Coca-cola	Sprite	Coca-cola	Pepsi-cola
Coca-cola zero	Coca-cola	Coca-cola	Coca-cola
Pepsi-cola	Sprite	Pepsi-cola	Pepsi-cola
Coca-cola	Pepsi-cola	Coca-cola	Sprite
Coca-cola zero	Coca-cola	Sprite	Coca-cola zero
Pepsi-cola	Pepsi-cola	Sprite	Coca-cola zero

1.1 Distribuição de frequência para dados qualitativos

Uma distribuição de frequência agrupa os dados por classes de ocorrência, resumindo a análise de conjunto de dados grandes. Vejamos um exemplo. Adotemos o conjunto de dados do quadro anterior que representa o consumo de refrigerantes de um grupo de amigos.

Para construirmos a tabela de frequências, organizamos os dados em 3 colunas: coluna das categorias ou classes – onde se indicam todas as categorias da variável em estudo; coluna das frequências absolutas – onde se registra o total de elementos da amostra que pertencem a cada categoria – e coluna das frequências relativas (ou porcentagens) – onde se coloca, para cada categoria, o valor que se obtém dividindo a respectiva frequência absoluta pela dimensão da amostra.

Tabela 1

DISTRIBUIÇÃO DE FREQUÊNCIA DAS COMPRAS DE REFRIGERANTES		
Refrigerantes	Frequência Absoluta	Frequência Relativa
Coca-cola	8	0,3333
Coca-cola zero	4	0,1667
Pepsi-cola	7	0,2917
Sprite	5	0,2083
Total	24	1

Dados Hipotéticos

¹⁰ Em alguns casos, as frequências relativas são dízimas infinitas obrigando, por isso, a arredondamentos. Estes podem ser feitos desde que o total seja igual a 1.

Ao organizarmos os dados de uma amostra numa tabela de frequências, é bom verificarmos se as frequências estão bem calculadas. Para isso é suficiente somá-las para todas as classes e verificar que a soma das frequências absolutas é igual à dimensão da amostra e que a soma das frequências relativas¹⁰ é igual a 1.

Saiba Mais



Arredondamento de Dados

Regras: Portaria 36 de 06/07/1965 - INPM - Instituto Nacional de Pesos e Medidas.

- Se o primeiro algarismo após aquele que formos arredondar for de 0 a 4, conservamos o algarismo a ser arredondado e desprezamos os seguintes. Ex.: 7,34856 (para décimos) 7,3
- Se o primeiro algarismo após aquele que formos arredondar for de 6 a 9, acrescenta-se uma unidade no algarismo a ser arredondado e desprezamos os seguintes. Ex.: 1,2734 (para décimos) 1,3
- Se o primeiro algarismo após aquele que formos arredondar for 5, seguido apenas de zeros, conservamos o algarismo se ele for par ou aumentamos uma unidade se ele for ímpar, desprezando os seguintes. Ex.: 6,2500 (para décimos) 6,2 12,350 (para décimos) 12,4
- Se o 5 for seguido de outros algarismos dos quais, pelo menos um é diferente de zero, aumentamos uma unidade no algarismo e desprezamos os seguintes. Ex.: 8,2502 (para décimos) 8,3 8,4503 (para décimos) 8,5

1.2 Representação gráfica para dados qualitativos

O gráfico estatístico¹¹ é uma forma de apresentação/visualização dos dados estatísticos, que permite que a pessoa que está analisando esses dados tenha uma impressão mais rápida e mais viva do fenômeno em estudo, uma vez que os gráficos falam mais à compreensão do que as séries.

A grande vantagem dos gráficos em relação às tabelas de frequências está na rapidez da leitura, em que **não só há uma percepção imediata de qual a categoria de maior** frequência, como também se evidencia com mais clareza a ordem de grandeza de cada categoria relativamente às restantes.

Vejamos um exemplo de gráfico em colunas usando como base a nossa tabela de distribuição de frequência das compras de refrigerantes

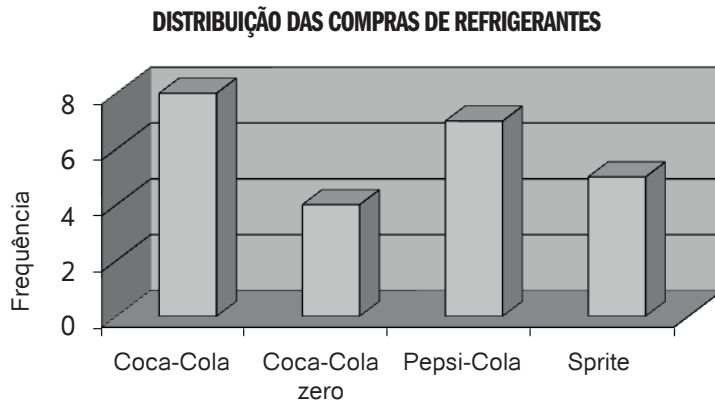


Gráfico 1

¹¹ Requisitos fundamentais para a construção de um gráfico:

Simplicidade: o gráfico deve ser destituído de detalhes de importância secundária, evitando-se, também, traços desnecessários que possam levar a pessoa que está analisando a uma interpretação equivocada dos dados.

Clareza: o gráfico deve possibilitar uma correta interpretação dos valores representativos do caso em estudo.

Veracidade: o gráfico deve ser a verdadeira expressão do fenômeno em estudo.

2. Sintetizando dados quantitativos

O dado quantitativo faz referência aos números no sentido de quantidade e podem se dividir em duas categorias.

- **Quantitativo discreto:** aquele que pode assumir apenas valores pertencentes a um conjunto enumerável.

Ex: população: habitações de uma cidade (variável: número de cômodos); população: casais em uma cidade (variável: número de filhos)

- **Quantitativo contínuo:** pode assumir qualquer valor em certo intervalo de variação.

Exemplo: população: pessoas residentes em uma cidade (variável: idade); estudantes: alunos de uma universidade (variável: estatura dos alunos)

2.1 Distribuição de frequência

Vamos imaginar uma pesquisa referente às estaturas de quarenta alunos que compõem uma amostra dos alunos de uma universidade, o que resultou na tabela de valores a seguir:

Tabela 2

ESTATURA DOS ALUNOS DE UMA UNIVERSIDADE									
166	160	161	150	162	160	165	167	164	160
162	168	161	163	156	173	160	155	164	168
155	152	163	160	155	155	169	151	170	164
154	161	156	172	153	157	156	158	158	161

Fonte: Dados Hipotéticos

A tabela anterior é denominada de "tabela primitiva" por considerar que os elementos ainda não foram numericamente organizados.

Visualizando a tabela anterior percebemos que fica difícil saber em torno de que valor tende a se concentrar as estaturas dos alunos, qual a menor ou qual a maior estatura ou, ainda, quantos alunos se acham abaixo ou acima de uma determinada estatura. Logo, é importante organizar os dados, e a maneira mais simples de fazer isso é através de uma certa ordenação (crescente ou decrescente). A tabela obtida através da ordenação dos dados recebe o nome de rol.

Tabela 3

ESTATURA DOS ALUNOS DE UMA UNIVERSIDADE									
150	154	155	157	160	161	162	164	166	169
151	155	156	158	160	161	162	164	167	170
152	155	156	158	160	161	163	164	168	172
153	155	156	160	160	161	163	165	168	173

Dados Hipotéticos

Depois de feita esta ordenação, ficou fácil visualizar que a menor estatura é 150 cm, que a amplitude de variação foi de $173 - 150 = 23$ cm e, ainda, a ordem que um valor particular da variável ocupa no conjunto. Com um exame mais minucioso, vemos que há uma concentração das estaturas em algum valor entre 160 cm e 165 cm e, mais ainda, que há poucos valores abaixo de 155 cm e acima de 170 cm.

Construindo a tabela de frequência absoluta teremos:

Tabela 4

Distribuição de frequência da estatura de 40 alunos de uma universidade	
Estatura (cm)	Frequência
150	1
151	1
152	1
153	1
154	1
155	4
156	3
157	1
158	2
160	5
161	4
162	2
163	2
164	3
165	1
166	1
167	1
168	2
169	1
170	1
172	1
173	1
Total	40

Dados Hipotéticos

Observando a tabela anterior, percebemos que a sua construção não é adequada quando o número de valores da variável (n) é de tamanho razoável. Então, é importante fazermos o agrupamento dos valores em vários intervalos que, em Estatística, chamamos de classes.

A tabela a seguir foi construída considerando a frequência de uma classe o número de valores da variável pertencente à classe. Essa tabela é denominada "distribuição de frequência com intervalos de classe".

¹² $154 \vdash 158$ (é um intervalo fechado à esquerda e aberto à direita, tal que: $154 \leq x < 158$), em vez de dizermos que a estatura de 1 aluno é de 154 cm; de 4 alunos, 155 cm; de 3 alunos, 156 cm; e de 1 aluno, 157 cm, dizemos que 9 alunos têm estaturas entre 154, inclusive, e 158 cm.

Tabela 5

DISTRIBUIÇÃO DE FREQUÊNCIA DA ESTATURA DE 40 ALUNOS DE UMA UNIVERSIDADE	
Classes	Frequência
150 \vdash 154	4
154 \vdash 158 ¹²	9
158 \vdash 162	11
162 \vdash 166	8
166 \vdash 170	5
170 \vdash 174	3

Dados Hipotéticos

É importante salientar que, ao agruparmos os valores da variável em classes, ganhamos em simplicidade para perdermos os detalhes. Perceba que, na Tabela 3, é fácil verificar que quatro alunos têm 161 cm de altura e que na Tabela 4, não podemos ver se algum aluno tem a estatura de 159 cm. No entanto, sabemos com segurança que onze alunos têm estatura compreendida entre 158 e 162 cm. Porém a nova tabela realça o que há de essencial nos dados e, também, torna possível o uso de técnicas analíticas para sua total descrição, o que vem ao encontro da finalidade da estatística, que é especificar e analisar o conjunto de valores, desinteressando-se por casos isolados.

¹³ Na distribuição de frequência com classe, o h_i será igual em todas as classes. A amplitude total da distribuição jamais coincide com a amplitude amostral, AT é sempre maior que AA.

2.2 Elementos de uma distribuição de frequência¹³

- **Classe:** são os intervalos de variação da variável; é simbolizada por i e o número total de classes, simbolizada por k . Ex: na tabela anterior, $k = 6$ e $158 \vdash 162$ é a 3ª classe, onde $i = 3$.
- **Limites de classe:** são os extremos de cada classe. O menor número é o limite inferior de classe (l_i) e o maior número, limite superior de classe (L_i). Ex: em $158 \vdash 162$, $l_3 = 158$ e $L_3 = 162$.
- **Amplitude do intervalo de classe:** é obtida através da diferença entre o limite superior e inferior da classe e é simbolizada por $h_i = L_i - l_i$. Ex: na tabela anterior, $h_i = 162 - 158 = 4$.
- **Amplitude total da distribuição:** é a diferença entre o limite superior da última classe e o limite inferior da primeira classe. $AT = L(\max) - l(\min)$. No nosso exemplo, $AT = 174 - 150 = 24$.
- **Amplitude total da amostra (ROL):** é a diferença entre o valor máximo e o valor mínimo da amostra (ROL), onde $AA = X_{\max} - X_{\min}$. Em nosso exemplo, $AA = 173 - 150 = 23$.

- **Ponto médio de classe:** é o ponto que divide o intervalo da classe em duas partes iguais. No nosso exemplo, em $158 \vdash 162$, $x_i = (l_i + L_i)/2 \Rightarrow x_2 = (158 + 162)/2 = 160$

2.3 Passos para construção das categorias em classes

1. Organize os dados brutos em um ROL.
2. Calcule a amplitude amostral AA (no nosso exemplo, $AA = 173 - 150 = 23$)
3. Calcule o número de classes através da “Regra de Sturges¹⁴” (no nosso exemplo: $v = 40$ dados, então, a princípio, a regra sugere a adoção de 6 classes)
4. Decidido o número de classes, calcule a amplitude do intervalo de classe dividindo a amplitude total da amostra pelo número de classes $h \approx AA / k$ (no nosso exemplo $h \approx 23/6 = 3,8 \approx 4$)
5. Temos então o menor número da amostra, o número de classes e a amplitude do intervalo; com isso, é possível montar a tabela, com o cuidado para não aparecer em classes com frequência = 0 (zero). No nosso exemplo, o menor número da amostra = $150 + 4 = 154$, logo a primeira classe será representada por $150 \vdash 154$. As classes seguintes respeitarão o mesmo procedimento.

2.4 Tipos de frequências

- **Frequência simples ou absoluta (f_i):** são os valores que realmente representam o número de dados de cada classe. A soma das frequências simples é igual ao número total dos dados da distribuição.

$$f_i = n$$

- **Frequência relativa (fr_i):** são os valores das razões entre as frequências absolutas de cada classe e a frequência total da distribuição. A soma das frequências relativas é igual a 1 (100 %).

$$fr_i = f_i / f$$

- **Frequência acumulada (F_i):** é o total das frequências de todos os valores inferiores ao limite superior do intervalo de uma dada classe.

$$F_k = f_1 + f_2 + \dots + f_k \quad \text{ou} \quad F_k = \sum_{i=1}^k f_i \quad (i = 1, 2, \dots, k)$$

- **Frequência acumulada relativa (Fr_i):** é a frequência acumulada da classe dividida pela frequência total da distribuição.

$$Fr_i = F_i / f$$

¹⁴ A Regra de Sturges nos dá o número de classes em função do tamanho da amostra

$$k \approx 1 + 3,3 \cdot \log n$$

Onde:

k é o número de classe;

n é o tamanho da amostra.

No cálculo da amplitude do intervalo de classe, quando o resultado não é exato, devemos arredondá-lo para mais.

Podemos agora construir a tabela a seguir com as frequências estudadas considerando o nosso exemplo.

Tabela 11

DISTRIBUIÇÃO DE FREQUÊNCIA DA ESTATURA DE 40 ALUNOS DE UMA UNIVERSIDADE						
i	Estaturas (cm)	f_i	x_i	fr_i	F_i	Fr_i
1	150 – 154	4	152	0,100	4	0,100
2	154 – 158	9	156	0,225	13	0,325
3	158 – 162	11	160	0,275	24	0,600
4	162 – 166	8	164	0,200	32	0,800
5	166 – 170	5	168	0,125	37	0,925
6	170 – 174	3	172	0,075	40	1,000
		$\Sigma = 40$		$\Sigma = 1,000$		

Dados Hipotéticos

De acordo com o estudo dessas frequências, podemos responder às questões a seguir:

1. Quantos alunos têm estatura entre 158 cm, inclusive, e 156 cm?

Resposta: 11 – esses são os valores que formam a terceira classe ($f_3 = 11$)

2. Qual a porcentagem de alunos cujas estaturas são inferiores a 154 cm?

Resposta: 10% – esses valores são os que formam a primeira classe ($fr_1 = 0,100$). Obtemos a resposta multiplicando a frequência relativa por 100: $0,100 \times 100 = 10$

3. Quantos alunos têm estatura abaixo de 162?

Resposta: 24 – as estaturas abaixo de 162 são aquelas que formam as classes de ordem 1, 2 e 3. Assim, o número de alunos é dado por $F_3 = (\sum_{i=1}^3 f_i) = f_1 + f_2 + f_3 \Rightarrow F_3 = 24$

4. Quantos alunos têm estatura maior ou igual a 158 cm?

Resposta: 27 – as estaturas maior ou igual a 158 são aquelas que formam as classes de ordem 3, 4, 5 e 6. Assim, o número de alunos é dado por $(\sum_{i=3}^6 f_i) = f_3 + f_4 + f_5 + f_6 = 11 + 8 + 5 + 3 = 27$

2.5 Distribuição de frequências e sua representação gráfica para variáveis quantitativas contínuas

As variáveis quantitativas contínuas diferem um pouco das discretas na sua forma de representação gráfica. A diferença mais importante é que as frequências são associadas a intervalos de valores (classes de frequências) e não a valores individuais da variável em estudo. Graficamente, podemos representar pelo histograma, pelo polígono de frequência e pelo polígono de frequência acumulada.

a) Histograma

O histograma é um gráfico formado por um conjunto de colunas retangulares. No eixo das abscissas, marcamos as classes, cujas amplitudes correspondem às bases dos retângulos. No eixo das ordenadas, marcamos as frequências absolutas, que correspondem às alturas dos retângulos. Os pontos médios das bases dos retângulos coincidem com os pontos médios dos intervalos de classes.

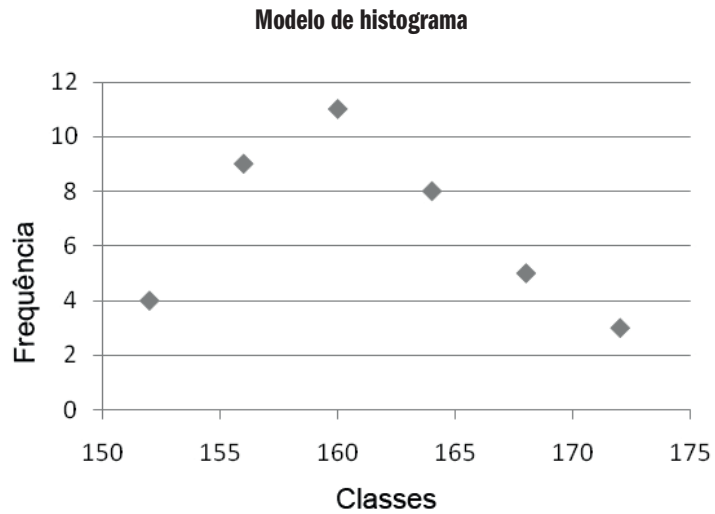
Roteiro para construção do histograma:

1. Obtenha a tabela de frequência a partir dos dados, agrupando-os em classes.
2. Desenhe dois eixos ortogonais de tamanho médio.
3. Divida o eixo horizontal em tantas partes quanto for o número de classes mais dois (considere uma classe à esquerda da primeira classe e outra à direita da última classe para deixar espaço suficiente para traçar o polígono de frequência, que veremos mais adiante) e marque os números correspondentes aos limites inferior e superior de cada classe.
4. Identifique a maior frequência da classe na tabela de frequência; escolha um número adequado, maior ou igual àquela frequência; marque esse número na extremidade do eixo vertical; divida o eixo vertical em algumas partes e marque os números correspondentes.
5. Para cada classe, desenhe um retângulo com largura igual à amplitude da classe com altura igual à frequência da classe.

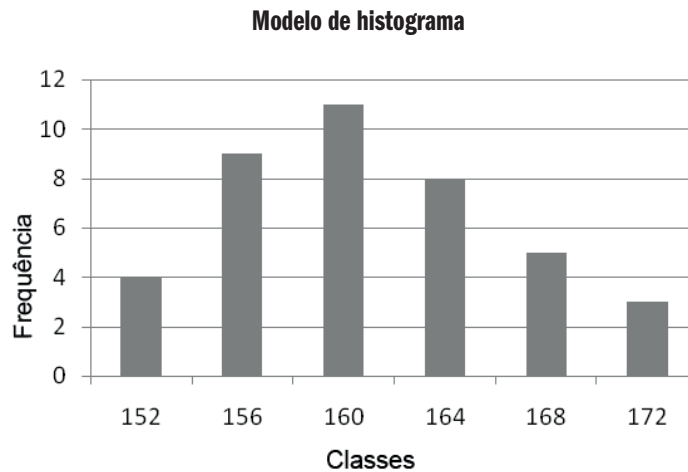
Para construir o histograma com o software Excel, depois de feita a tabela de distribuição de frequência em classes, adicione uma nova coluna com os pontos médios dos intervalos de classe, que serão as abscissas de um gráfico de dispersão, com as frequências correspondendo às alturas.

Classes	Frequência	Ponto médio
150 154	4	152
154 158	9	156
158 162	11	160
162 166	8	164
166 170	5	168
170 174	3	172

Veja o exemplo a seguir.

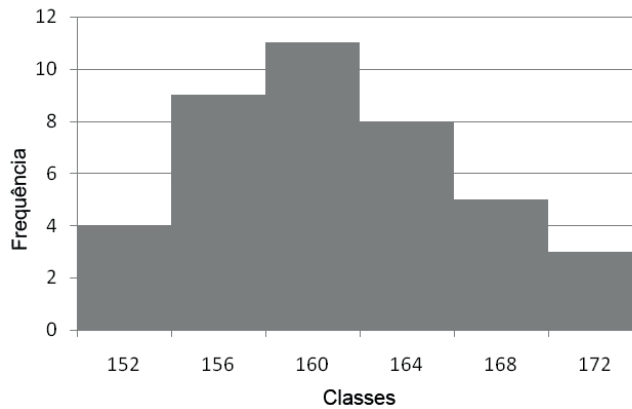


Com esse gráfico construído, clique com o botão direito do *mouse* sobre ele, selecione a opção ALTERAR TIPO DE GRÁFICO escolhendo colunas.



Sobre uma das colunas, com o botão direito do *mouse*, escolha a opção FORMATAR SEQUÊNCIA DE DADOS, em seguida ESPAÇAMENTO e zere-o.

Veja mais um modelo de histograma a seguir.

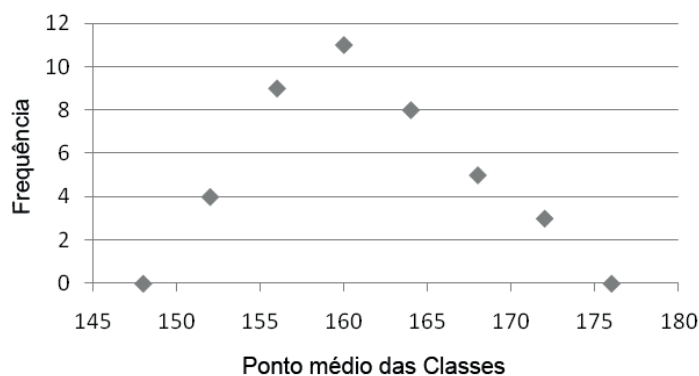
Modelo de histograma

b) Polígono de frequência

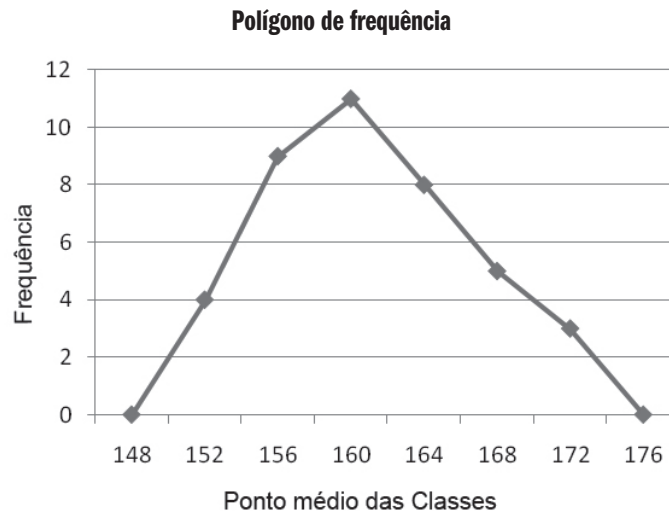
É a representação gráfica de uma distribuição por meio de um polígono. É obtido através da ligação dos pontos médios dos intervalos de classe em um histograma de frequências.

Para construir o polígono de frequência com o Excel, inclua uma classe anterior à primeira e uma posterior à segunda, ambas com frequência zero.

Classes	Frequência	Ponto médio
146 150	0	148
150 154	4	152
154 158	9	156
158 162	11	160
162 166	8	164
166 170	5	168
170 174	3	172
174 178	0	176

Polígono de frequência

Uma vez construído esse gráfico, clique com o botão direito do *mouse* sobre o gráfico selecionando a opção TIPO DE GRÁFICO escolhendo GRÁFICOS EM LINHA.



c) Ogiva

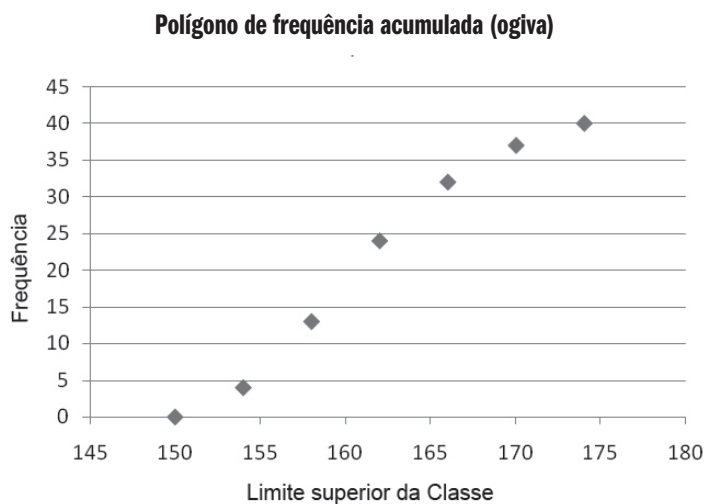
É um polígono de frequências acumuladas no qual elas são localizadas sobre perpendiculares levantadas nos limites inferiores ou superiores das classes, dependendo se a ogiva representar as frequências acumuladas abaixo ou acima, respectivamente.

Para construir o polígono de frequência acumulada (ogiva) com o Excel, inclua uma classe anterior à primeira com frequência nula. O gráfico será construído com os limites superiores de cada classe e as respectivas frequências acumuladas.

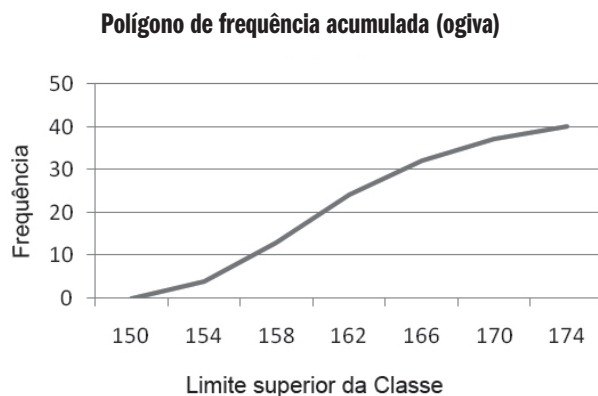
Classes	Frequência	Frequência acumulada
146 150	0	0
150 154	4	4
154 158	9	13
158 162	11	24
162 166	8	32
166 170	5	37
170 174	3	40

Com o Excel aberto, opção ASSISTENTE DE GRÁFICO, escolha tipo de gráfico (Dispersão), complete com o título do Gráfico, e com os valores dos eixos X e Y.

Veja o exemplo a seguir.



Uma vez construído esse gráfico, clique com o botão direito do *mouse* sobre o gráfico selecionando a opção tipo de gráfico escolhendo GRÁFICOS EM LINHA.



2.6 Distribuição de frequência sem intervalo de classes

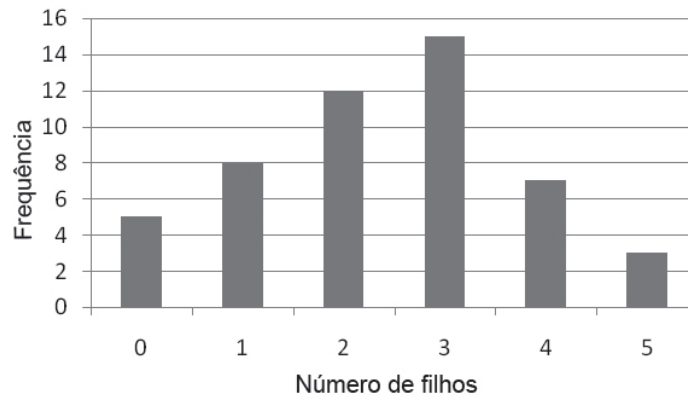
É a simples condensação dos dados conforme as repetições de seus valores. Cada valor é tomado como um intervalo de classe tratando-se de variável discreta de variação relativamente pequena.

Veja o exemplo a seguir.

Considere x a variável “número de filhos de 50 famílias”

i	x_i	f_i
1	0	5
2	1	8
3	2	12
4	3	15
5	4	7
6	5	3
		$\Sigma = 50$

A representação gráfica de uma distribuição de frequência de uma variável quantitativa discreta é denominada **gráfico de frequências**. Utilizando a tabela de frequência anterior obtemos o gráfico a seguir.



Outra representação utilizada é a do gráfico das frequências acumuladas e das frequências relativas acumuladas. Tomando-se os dados do nosso exemplo, podemos calcular as frequências, frequências acumuladas e frequências relativas acumuladas dos diversos valores. Esse cálculo está ilustrado na tabela a seguir.

i	x_i	f_i	fr_i	F_i	Fr_i
1	0	5	0,10	5	0,10
2	1	8	0,16	13	0,26
3	2	12	0,24	25	0,50
4	3	15	0,30	40	0,80
5	4	7	0,14	47	0,94
6	5	3	0,06	50	1,00
		$\Sigma = 50$	$\Sigma = 1,00$		

Veja a representação nos gráficos a seguir.

Gráfico de frequências relativas

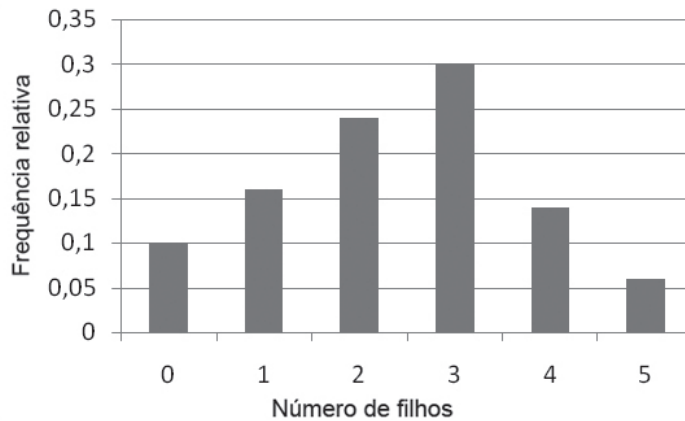
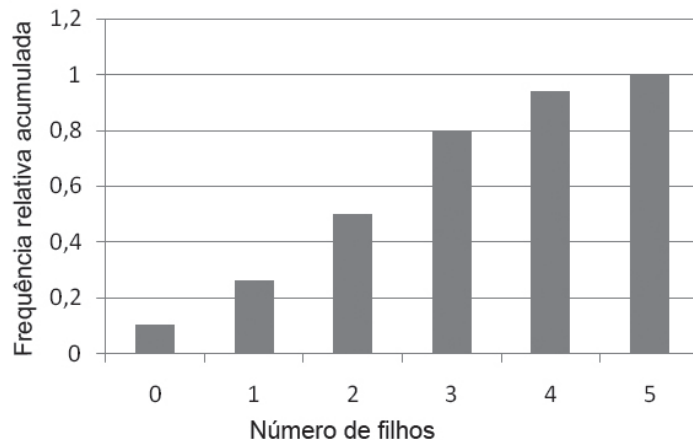


Gráfico de frequências relativas acumuladas



Atividades de avaliação



- Conhecidas as notas dos alunos em uma disciplina, de acordo com os dados a seguir, obtenha a distribuição de frequência, considerando 30 para limite inferior da primeira classe e, 10 para intervalo de classe.

84 68 33 52 47 73 68 61 73 77
 74 71 81 91 65 55 57 35 85 88
 59 80 41 50 53 65 76 85 73 60
 67 41 78 56 94 35 45 55 64 74
 65 94 66 48 39 69 89 86 42 54

2. Complete a tabela a seguir.

i	Classes	f_i	x_i	fr_i	F_i	Fr_i
1	0 – 8	4				
2	08 – 16	10				
3	16 – 24	14				
4	24 – 32	9				
5	32 – 40	3				
		$\Sigma = 40$				

3. Considere a seguinte tabela

i	Classes	fi
1	2,75 – 2,80	2
2	2,80 – 2,85	3
3	2,85 – 2,90	10
4	2,90 – 2,95	11
5	2,95 – 3,00	24
6	3,00 – 3,05	14
7	3,05 – 3,10	9
8	3,10 – 3,15	8
9	3,15 – 3,20	6
10	3,20 – 3,25	3
	Total	90

Identifique os seguintes elementos da tabela.

- Frequência simples absoluta da quinta classe.
- Frequência total.
- Limite inferior da sexta classe.
- Limite superior da quarta classe.
- Amplitude do intervalo de classe.
- Amplitude total.
- Ponto médio da terceira classe.

4. Construa o histograma da distribuição a seguir:

i	Áreas (em m ²)	Número de lotes
1	300 – 400	14
2	400 – 500	46
3	500 – 600	58
4	600 – 700	76

i	Áreas (em m ²)	Número de lotes
5	700 – 800	68
6	800 – 900	62
7	900 – 1000	48
	Total	372

Referências



- BARBETA, P. A., REIS, M. M., BORNIA, A. C. **Estatística para Cursos de Engenharia e Informática**, São Paulo: Atlas, 2004.
- BOLFARINE, Heleno & BUSSAB, Wilton O. **Elementos de Amostragem**. 1. ed. São Paulo: Edgard Blücher, 2005.
- BUSSAB, Wilton de OI., MORETTIN, Pedro A. **Estatística Básica**. São Paulo: Saraiva, 2002.
- COSTA, F. S. **Introdução Ilustrada à Estatística**. São Paulo: Harbra, 1998.
- CRESPO, A. **Estatística Fácil**. Rio de Janeiro: Saraiva, 1996.
- FONSECA, Jairo Simon da. **Introdução à Estatística**. Rio de Janeiro: LTC. 1993.
- JAIRO, Simon da Fonseca, MARTINS, Gilberto de Andrade. **Curso de Estatística**, São Paulo: Atlas, 1996.
- MARTINS G. A. **Estatística Geral e Aplicada**. São Paulo: Atlas, 2001.
- MEYER, Paul L. **Aplicações a Estatística**. Rio de Janeiro: LTC. 1996.
- MORETTIN, Pedro A. **Estatística Básica**. São Paulo: Saraiva. 2013.
- PEREIRA, Wilson., TANAKA, Oswaldo K. **Estatística – Conceitos Básicos**, São Paulo: Makron Books, 1990.
- SPIEGEL M. R. **Estatística. Coleção Schaum**. São Paulo: Editora Afiliada 1993.
- TRIOLA. M. F. **Introdução à Estatística e Probabilidade - Exercícios Resolvidos e Propostos**. Rio de Janeiro: LTC, 1999.
- TRIOLA, Mario F. **Introdução a Estatística**. Rio de Janeiro: LTC. 2005.
- VIEIRA, S. **Estatística Experimental**. São Paulo: Atlas, 1999.

