

Aula 1 – Apresentação de dados – parte 1

Nesta aula, você aprenderá:

- os conceitos básicos de população e amostra de uma pesquisa estatística;
- a distinção entre variáveis qualitativas e variáveis quantitativas;
- a construir distribuições de frequências para variáveis qualitativas e quantitativas discretas;
- a construir gráficos de colunas e de setores para representar dados qualitativos e quantitativos discretos.

Pesquisa estatística - conceitos básicos

População e amostra

Estatística é a ciência da aprendizagem a partir dos dados. Em geral, fazemos levantamentos de dados para estudar e compreender características de uma população. Por exemplo, um grande banco, querendo lançar um novo produto, precisa conhecer o perfil socioeconômico dos seus clientes e, nesse caso, a população de interesse é formada pelos clientes de todas as agências do banco. A Federação das Indústrias do Estado do Rio de Janeiro – FIRJAN – mede o grau de confiança dos empresários industriais através de uma pesquisa junto às empresas industriais, sendo a população de interesse, aqui, o conjunto das empresas industriais do estado do Rio de Janeiro. Com esses dois exemplos apenas, já podemos ver que o conceito de *população de uma pesquisa estatística* é mais amplo, não se restringindo a seres humanos; ela é definida exatamente a partir dos objetivos da pesquisa. Mais precisamente, *população é o conjunto de elementos para os quais se deseja estudar determinada(s) característica(s)*.

Embora tenham populações bastante distintas, essas duas pesquisas têm em comum o fato de os resultados desejados serem obtidos a partir de dados levantados junto a um subconjunto da população – uma *amostra*. Há várias razões para se trabalhar com *pesquisas por amostragem* – custo e tempo, em geral, são as mais comuns. Mas, além de serem mais baratas e rápidas, as pesquisas por amostragem, se bem planejadas, podem fornecer resultados quase tão precisos quanto aqueles fornecidos por *pesquisas censitárias*, em que todos os elementos da população são investigados. Exemplos clássicos de pesquisa censitária são os Censos Demográficos realizados a cada

POPULAÇÃO

AMOSTRA

CENSO

dez anos no Brasil e em outros países. O objetivo desses censos é levantar informações sobre toda a população do país, de modo a fornecer subsídios para os governantes definirem as políticas públicas.

Variáveis qualitativas e quantitativas

Nas pesquisas estatísticas, as características sobre as quais queremos obter informação são chamadas *variáveis*. Em uma pesquisa domiciliar sobre emprego e renda, algumas variáveis de interesse são sexo, raça, grau de instrução e valor dos rendimentos do morador. Em uma pesquisa sobre o estado nutricional dos brasileiros, o peso e a altura dos moradores de cada domicílio da amostra foram medidos. Para o acompanhamento da atividade industrial no Rio de Janeiro, a FIRJAN obtém informações junto às empresas industriais sobre tipo de atividade econômica, número de empregados, número de horas trabalhadas, valor da folha de pagamento. É importante diferenciar entre variáveis qualitativas e variáveis quantitativas. Sexo, raça, religião e atividade econômica de uma empresa são exemplos de *variáveis qualitativas*. Já valor dos rendimentos, peso, altura, número de empregados, valor da folha de pagamento são exemplos de *variáveis quantitativas*. Podemos ver, então, que as variáveis qualitativas *descrevem* características dos elementos de uma população, enquanto as variáveis quantitativas *mensuram* características desses elementos.

As variáveis quantitativas, por sua vez, podem ser discretas ou contínuas. Quando a variável puder assumir qualquer valor numérico em um determinado intervalo de variação, ela será uma variável *contínua*. Essas variáveis resultam normalmente de medições: peso, altura, dosagem de hemoglobina, renda etc. A interpretação desse tipo de variável leva à noção de valor aproximado, pois não existe instrumento de medição capaz de fornecer precisão absoluta na informação. Assim, quando uma balança mostra o peso de uma pessoa como 65,5 kg, esse valor, na verdade, é uma aproximação para qualquer valor entre, digamos, 65,495 e 65,505 kg. Por outro lado, a variável quantitativa *discreta* só poderá assumir valores pertencentes a um conjunto enumerável; os valores normalmente são obtidos através de algum processo de contagem. Alguns exemplos são: número de filhos de um casal, número de empregados de uma firma de contabilidade etc.

VARIÁVEIS QUALITATIVAS

VARIÁVEIS QUANTITATIVAS

Atividade 1.1

O texto a seguir foi extraído da página do IBOPE na Internet: www.ibope.com.br. Aí temos parte da descrição da pesquisa sociodemográfica realizada por esse instituto. Identifique as variáveis pesquisadas, classificando-as como qualitativas ou quantitativas.

“O Levantamento Socioeconômico (LSE) é a pesquisa do IBOPE Mídia que mapeia as características sociais, demográficas e econômicas das famílias das principais regiões metropolitanas do país. Oferece também outros dados essenciais para traçar a estratégia de marketing para um produto. Com uma base de dados estendida em relação às outras pesquisas do IBOPE Mídia, o LSE serve de base para outros estudos.

São levantados dados sobre a condição do domicílio entrevistado (condição da rua, tipo de imóvel) e sobre a condição socioeconômica do domicílio (informações sobre renda e classificação econômica). Também são pesquisados o número de pessoas no domicílio, a presença e a quantidade de crianças e adolescentes, a idade, grau de instrução e condição de atividade do chefe da casa e da dona-de-casa. A pesquisa levanta também dados sobre a posse de bens, como geladeira, máquina de lavar, automóvel, rádio, computador, telefone, entre outros, e acesso a serviços de mídia, como TV por Assinatura, Internet, etc.”

Apresentação de dados qualitativos

Vamos considerar o seguinte exemplo fictício, mas verossímil. A direção de uma empresa está estudando a possibilidade de fazer um seguro saúde para seus funcionários e respectivos familiares. Para isso, ela faz um levantamento junto a seus 500 funcionários, obtendo informação sobre sexo, estado civil, idade, número de dependentes e salário. Como são 500 funcionários, temos que achar uma forma de resumir os dados. Nesta aula você irá aprender a resumir dados qualitativos em forma de uma distribuição (ou tabela) de frequência e também em forma gráfica. Você verá que os gráficos complementam a apresentação tabular.

Distribuições de frequência

Consideremos inicialmente a variável qualitativa sexo. O que interessa saber sobre essa variável não é que João é do sexo masculino e Maria é do sexo feminino, mas, sim, quantos funcionários e quantas funcionárias há na empresa. Esse resultado pode ser resumido em uma tabela ou distribuição de frequências da seguinte forma:

Sexo	Número de Funcionários
Masculino	270
Feminino	230
Total	500

Os números 270 e 230 resultaram da contagem das frequências de ocorrência de cada uma das categorias da variável sexo. Essa contagem é também chamada de *frequência simples absoluta* ou simplesmente *frequência*. O total de 500 é obtido somando-se o número de homens e de mulheres.

É interessante também expressar esses resultados em forma relativa, ou seja, considerar a frequência de cada categoria em relação ao total:

$$\frac{270}{500} = 0,54$$

ou seja, 54% dos funcionários da empresa são do sexo masculino e

$$\frac{230}{500} = 0,46$$

ou seja, 46% dos funcionários são mulheres. A **Tabela 1.1** apresenta a versão completa.

Tabela 1.1: Distribuição do número de funcionários por sexo

Sexo	Frequência Simples	
	Absoluta	Relativa
Masculino	270	0,54
Feminino	230	0,46
Total	500	1,00

Note que a soma das frequências relativas é sempre 1, enquanto a soma das frequências absolutas deve ser igual ao número total de elementos sendo investigados.

De maneira análoga, obteríamos a **Tabela 1.2** para a variável estado civil. Note que, aí, a frequência relativa está apresentada em forma percentual, ou seja, multiplicada por 100. Por exemplo, para os casados temos:

$$\frac{280}{500} \times 100 = 0,56 \times 100 = 56\%$$

Em geral, essa é a forma mais usual de se apresentarem as frequências relativas e neste caso, a soma deve dar 100%.

Tabela 1.2: Distribuição do número de funcionários por estado civil

Estado Civil	Frequência Simples	
	Absoluta	Relativa %
Solteiro	125	25,0
Casado	280	56,0
Divorciado	85	17,0
Viúvo	10	2,0
Total	500	100,0

Exemplo 1.1

Consideremos que, na situação descrita anteriormente, os dados tenham sido levantados por departamento, para depois serem totalizados. Para o Departamento de Recursos Humanos, foram obtidas as seguintes informações:

Nome	Sexo	Estado civil	Número de dependentes
João da Silva	M	Casado	3
Pedro Fernandes	M	Viúvo	1
Maria Freitas	F	Casada	0
Paula Gonçalves	F	Solteira	0
Ana Freitas	F	Solteira	1
Luiz Costa	M	Casado	3
André Souza	M	Casado	4
Patrícia Silva	F	Divorciada	2
Regina Lima	F	Casada	2
Alfredo Souza	M	Casado	3
Margarete Cunha	F	Solteira	0
Pedro Barbosa	M	Divorciado	2
Ricardo Alves	M	Solteiro	0
Márcio Rezende	M	Solteiro	1
Ana Carolina Chaves	F	Solteira	0

Para pequenos conjuntos de dados, podemos construir a tabela à mão e para isso precisamos contar o número de ocorrências de cada categoria de cada uma das variáveis. Varrendo o conjunto de dados a partir da primeira linha, podemos ir marcando as ocorrências da seguinte forma:

Masculino		Solteiro	
Feminino		Casado	
		Divorciado	
		Viúvo	

Obtemos, então, as seguintes tabelas:

Sexo	Frequência Simples		Estado Civil	Frequência Simples	
	Absoluta	Relativa %		Absoluta	Relativa %
Masculino	8	53,33	Solteiro	6	40,00
Feminino	7	46,67	Casado	6	40,00
Total	15	100,0	Divorciado	2	13,33
			Viúvo	1	6,67
			Total	15	100,00

Arredondamento de números

No exemplo anterior, a divisão de algumas frequências absolutas pelo total de 15 resultou em dízimas. Nesses casos, torna-se necessário arredondar os resultados, mas esse arredondamento deve ser feito com cautela para se evitarem problemas tais como a soma não ser igual a 1 ou 100%.

A primeira etapa no processo de arredondamento consiste em se decidir o número de casas decimais desejado. Em geral, frequências relativas percentuais são apresentadas com, no máximo, 2 casas decimais. Isso significa que temos que descartar as demais casas decimais. Existe a seguinte regra de arredondamento:

Regra: Arredondamento de números

Quando o primeiro algarismo a ser suprimido é menor ou igual a 4 (ou seja, é igual a 0, 1, 2, 3 ou 4), o último algarismo a ser mantido permanece inalterado. Quando o primeiro algarismo a ser suprimido é igual a 5, 6, 7, 8 ou 9, o último algarismo a ser mantido é acrescido de 1.

Na distribuição de frequências da variável sexo, temos os seguintes resultados:

$$\begin{aligned}\frac{8}{15} \times 100 &= 53,33333\dots \\ \frac{7}{15} \times 100 &= 46,66666\dots\end{aligned}$$

No primeiro caso, o primeiro algarismo a ser suprimido é 3; logo, o último algarismo a ser mantido (3) não se altera e o resultado é 53,33. No segundo caso, o primeiro algarismo a ser suprimido é 6; logo, o último algarismo a ser mantido (6) deve ser acrescido de 1 e o resultado é 46,67. Tente sempre usar essa regra em seus arredondamentos; com ela, você evitará erros grosseiros.

Na apresentação de tabelas de frequências relativas, é possível que essas frequências não somem 100%, ou seja, é possível que, ao somarmos as frequências relativas, obtenhamos resultados como 99,9% ou 100,01%. Esses pequenos erros são devidos a arredondamentos e nem sempre é possível evitá-los; no entanto, aceita-se implicitamente que a soma das frequências seja 100%. Veja a tabela de frequências apresentada na **Figura 1.7**, relativa à solução do Exercício 1- a soma das frequências relativas é 99,99%. Se trabalhássemos com 3 casas decimais, obedecendo à regra de arredondamento, a soma daria 100,001. Isso não significa que as contas estejam erradas!

Atividade 1.2

Para o Departamento Financeiro, obteve-se a seguinte informação sobre o sexo dos 23 funcionários:

M F F M M M F F M M M M
M F M M F F M M M F F

onde M = Masculino e F = Feminino. Construa uma tabela de frequências para esses dados.

Gráficos

As distribuições de frequência para dados qualitativos também podem ser ilustradas graficamente através de gráficos de colunas ou gráficos de setores, também conhecidos como gráficos de pizza. Na **Figura 1.1** temos os gráficos de coluna e de setores para os dados da **Tabela 1.2**, referentes ao estado civil dos funcionários. No *gráfico de colunas*, a altura de cada coluna representa a frequência da respectiva classe e o gráfico pode ser construído com base nas frequências absolutas ou relativas. Para diferenciar um do outro, coloca-se no título do eixo o tipo de frequência utilizada. Note que, no eixo horizontal, não há escala, uma vez que aí se representam as categorias da variável, que devem ser equi-espaçadas.

No *gráfico de setores*, a frequência de cada categoria é representada pelo tamanho (ângulo) do setor (ou fatia da pizza). Para construir um gráfico de setores à mão, você precisa de um compasso para fazer um círculo de raio qualquer. Em seguida, trace um raio qualquer no círculo e a partir daí, comece a marcar os raios de acordo com os ângulos de cada setor, utilizando um transferidor. Para determinar o ângulo de cada setor, você deve usar a seguinte regra de proporcionalidade: o ângulo total – 360° – corresponde ao número total de observações; o ângulo de cada setor corresponde à frequência da respectiva classe. Dessa forma, você obtém a seguinte regra de três para os solteiros:

$$\frac{360^\circ}{500} = \frac{x}{125} \Rightarrow x = 90^\circ$$

Esses gráficos podem ser construídos facilmente com auxílio de programas de computador, como, por exemplo, o programa de planilhas Excel da Microsoft ®.

GRÁFICO DE COLUNAS

GRÁFICO DE SETORES

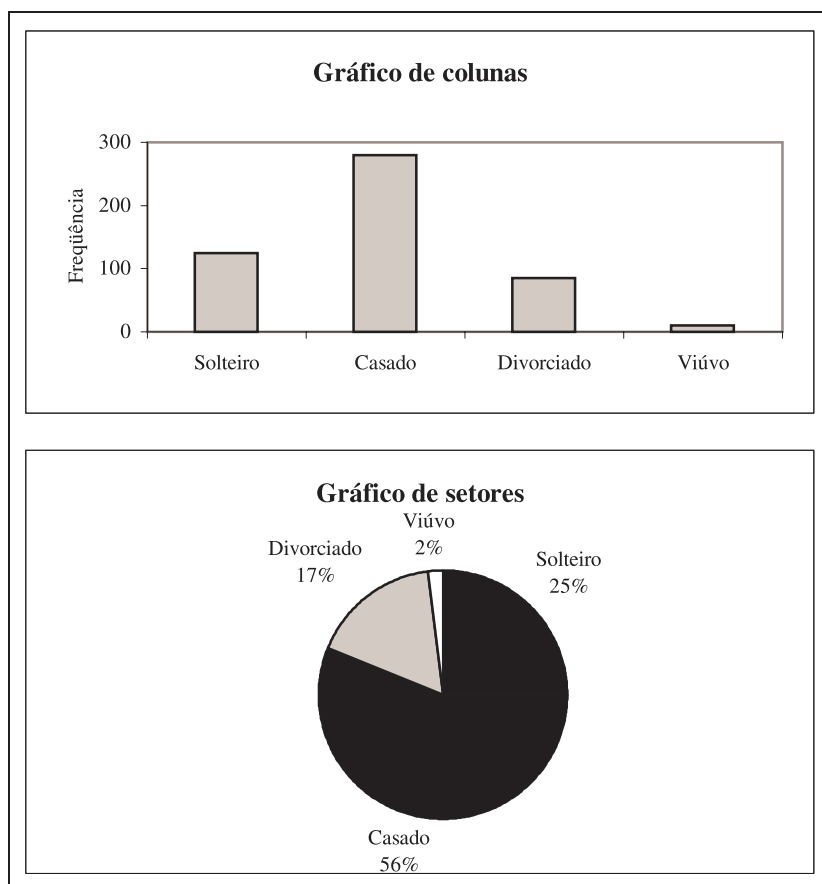


Figura 1.1: Distribuição do número de funcionários por estado civil.

Atividade 1.3

Construa os gráficos de setores e de colunas para os dados da Atividade 2.

Apresentação de dados quantitativos discretos

Quando uma variável quantitativa discreta assume poucos valores distintos, é possível construir uma distribuição de frequências da mesma forma que fizemos para as variáveis qualitativas. A diferença é que, em vez de termos categorias nas linhas da tabela, teremos os distintos valores da variável. Continuando com o nosso exemplo, vamos trabalhar agora com a variável número de dependentes. Suponha que alguns funcionários não tenham dependentes e que o número máximo de dependentes seja 7. Obteríamos, então, a seguinte distribuição de frequências:

Número de dependentes	Frequência Simples	
	Absoluta	Relativa %
0	120	24,0
1	95	19,0
2	90	18,0
3	95	19,0
4	35	7,0
5	30	6,0
6	20	4,0
7	15	3,0
Total	500	100,0

O processo de construção é absolutamente o mesmo mas, dada a natureza quantitativa da variável, é possível acrescentar mais uma informação à tabela. Suponha, por exemplo, que a empresa esteja pensando em limitar o seu projeto a 4 dependentes, de modo que funcionários com mais de 4 dependentes terão que arcar com as despesas extras. Quantos funcionários estão nessa situação? Para responder a perguntas desse tipo, é costume acrescentar à tabela de frequências uma coluna com as *frequências acumuladas*. Essas frequências são calculadas da seguinte forma: para cada valor da variável (número de dependentes), contamos quantas ocorrências correspondem a valores menores ou iguais a esse valor. Por exemplo, valores da variável menores ou iguais a 0 correspondem aos funcionários sem dependentes. Logo, a frequência acumulada para o valor 0 é igual à frequência simples: 120. Analogamente, valores da variável menores ou iguais a 1 correspondem aos funcionários sem dependentes mais os funcionários com 1 dependente. Logo, a frequência acumulada para o valor 1 é igual a $120 + 95 = 215$. Para o valor 2, a frequência acumulada é igual a $120 + 95 + 90 = 215 + 90 = 305$. Repetindo esse procedimento, obtemos a **Tabela 1.3**. Note que aí acrescentamos também as frequências acumuladas em forma percentual. Essas frequências são calculadas como a proporção da frequência acumulada em relação ao total; por exemplo,

$$87,0 = \frac{435}{500} \times 100$$

FREQUÊNCIA ACUMULADA

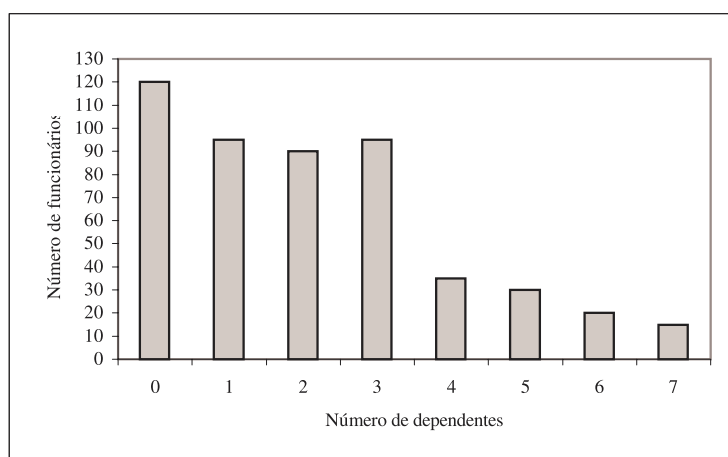
Tabela 1.3: Distribuição de frequências para o número de dependentes

Número de dependentes	Frequência Simples		Frequência Acumulada	
	Absoluta	Relativa %	Absoluta	Relativa %
0	120	24,0	120	24,0
1	95	19,0	215	43,0
2	90	18,0	305	61,0
3	95	19,0	400	80,0
4	35	7,0	435	87,0
5	30	6,0	465	93,0
6	20	4,0	485	97,0
7	15	3,0	500	100,0
Total	500	100,0		

Atividade 1.4

Construa a distribuição de frequência para o número de dependentes dos funcionários do Departamento de Recursos Humanos, conforme dados no Exemplo 1.

A representação gráfica da distribuição de frequências de uma variável quantitativa discreta pode ser feita através de um gráfico de colunas. A única diferença nesse caso é que no eixo horizontal do gráfico é representada a escala da variável quantitativa e tal escala deve ser definida cuidadosamente de modo a representar corretamente os valores. Na **Figura 1.2** temos o gráfico de colunas para o número de dependentes dos 500 funcionários.

**Figura 1.2:** Distribuição do número de dependentes de 500 funcionários.

Embora não seja incorreto, não é apropriado representar dados quantitativos discretos em um gráfico de setores, uma vez que, nesse gráfico, não é possível representar a escala dos dados.

Atividade 1.5

Construa o gráfico de colunas para representar a distribuição de frequências obtida na Atividade 4.

Resumo da Aula

Ao final desta aula, você deve ser capaz de compreender os seguintes conceitos:

- População – conjunto de elementos para os quais se deseja estudar determinada(s) característica(s).
- Amostra – subconjunto de uma população.
- Pesquisa censitária - pesquisa em que toda a população é investigada.
- Pesquisa por amostragem – pesquisa em que apenas uma amostra da população é investigada.
- Variável – característica de uma população que desejamos estudar.
- Variável qualitativa – variável que descreve uma característica dos elementos de uma população.
- Variável quantitativa – variável que mensura uma característica dos elementos de uma população.
- Variável quantitativa discreta – variável cujos possíveis valores formam um conjunto enumerável.
- Variável quantitativa contínua – variável cujos possíveis valores pertencem a um intervalo $[a, b]$.

Nesta aula, você também aprendeu a resumir dados de variáveis qualitativas e de variáveis quantitativas discretas através de tabelas de frequência e gráficos de setores e de colunas. É importante saber os seguintes conceitos:

- Freqüência simples absoluta – é a contagem do número de elementos pertencentes a uma determinada categoria de uma variável qualitativa ou número de elementos que assumem determinado valor de uma variável quantitativa discreta.
- Freqüência simples relativa – representa a participação percentual de cada categoria ou valor no total de observações.
- Freqüência acumulada absoluta – para cada valor de uma variável quantitativa discreta, é o número de ocorrências (elementos) correspondentes a valores menores ou iguais a esse valor.
- Freqüência acumulada relativa – é a freqüência acumulada em forma percentual, calculada como uma participação no total de observações.

Exercícios

1. Na **Tabela 1.4** temos informações sobre o sexo, a matéria predileta (**P**ortuguês, **M**atemática, **H**istória, **G**eografia ou **C**iências) no 2º grau e a nota (número de questões certas) em um teste de múltipla escolha com 10 questões de matemática, ministrado no primeiro dia de aula dos calouros de Administração de uma universidade (dados fictícios).
 - (a) Classifique as variáveis envolvidas.
 - (b) Construa a tabela de freqüências apropriada para cada uma das variáveis.
 - (c) Construa gráficos apropriados para ilustrar as distribuições de freqüência.

Tabela 1.4: Dados sobre sexo, matéria predileta e nota de alunos

Sexo	Predileta	Nota	Sexo	Predileta	Nota	Sexo	Predileta	Nota	Sexo	Predileta	Nota
F	H	5	M	M	2	M	H	3	F	M	8
M	M	8	M	G	4	M	M	5	M	P	5
F	P	8	M	G	9	F	P	5	M	G	6
F	H	6	M	M	7	F	G	5	F	M	7
M	C	5	M	M	1	M	C	7	M	P	5
M	H	6	F	P	8	M	H	4	F	M	5
F	M	8	F	G	5	F	M	7	F	M	5
F	P	4	M	G	9	F	P	7	F	P	9
F	H	2	M	P	5	F	M	6	M	M	8
M	C	6	F	M	8	M	G	6			
F	P	8	F	G	6	M	H	9			

2. Na **Tabela 1.5** temos dados sobre o consumo de refrigerantes no Brasil em 2005, segundo dados da Associação Brasileira das Indústrias de Refrigerantes e de Bebidas Não Alcoólicas. Construa um gráfico apropriado para ilustrar esses dados.

Tabela 1.5: Refrigerantes – Participação dos sabores – 2005

Refrigerantes	%
Colas	51,1
Guaraná	24,4
Laranja	10,9
Limão	5,9
Uva	3,2
Tuti Fruti	1,1
Tônica	0,7
Cítrico	0,1
Maçã	0,5
Outros sabores	2,1
Total	100,0

Fonte: ABIR - www.abir.org.br

3. Na **Tabela 1.6** temos as freqüências acumuladas do número de sinistros por apólice de seguro do ramo Automóveis. Complete a tabela, calculando as freqüências simples absolutas e relativas e também as freqüências acumuladas relativas.

Tabela 1.6: Número de sinistros por apólice, para o Exercício 3

Número de sinistros	Número de apólices
0	2913
≤ 1	4500
≤ 2	4826
≤ 3	4928
≤ 4	5000

4. Para a seguinte notícia, extraída do jornal *Folha de S. Paulo*, construa um gráfico para ilustrar o texto no segundo parágrafo da notícia.

Dentro de dez anos, 90% do mercado automobilístico mundial estará nas mãos de meia dúzia de conglomerados. A previsão consta de estudo produzido pela consultoria especializada britânica Autopolis, que dá assessoria técnica a montadoras que estão instaladas no Reino Unido.

Dados levantados pela Autopolis mostram que, hoje, a concentração de mercado já é grande. Cerca de 75% do setor é dominado por somente seis conglomerados, liderados por General Motors (22,8%), Ford (16,8%), Volkswagen (9,4%), Toyota (9,2%, incluindo Daihatsu), Renault-Nissan (8,7%) e Daimler-Chrysler (8,3%). Os outros 24,8% do mercado são dominados por uma infinidade de empresas pequenas e médias, como Fiat, BMW, Peugeot e Honda, entre outras.

Solução das Atividades

Atividade 1.1

É possível identificar as seguintes variáveis:

- Condição do domicílio – variável qualitativa.
- Condição da rua – variável qualitativa.