

## Distribuições de frequências simples e cruzada de dados

- ✓ Com o advento da informática, o mundo encheu-se de dados. As empresas têm dados de suas atividades, de seus funcionários, de seus clientes, etc. Mas para que esses dados sejam informativos, precisamos organiza-los, resumi-los e apresenta-los de forma adequada.
- ✓ Depois de realizado um levantamento de dados, eles são colocados em arquivos, sob a forma de matrizes. As linhas dessas matrizes correspondem ao que se observou em cada elemento pesquisado, enquanto as colunas correspondem às características (variáveis) levantadas.
  - Por exemplo, na atualização das páginas de um *site*, podemos querer avaliar o perfil dos indivíduos que acessam esse *site*. Então, precisamos levantar, junto a cada indivíduo, algumas de suas características, tais como o sexo, a idade, o nível de instrução e o provedor utilizado. Ao realizar a pesquisa, podemos produzir uma matriz de dados, da seguinte forma:

Variáveis					
<i>Usuário</i>	<i>Sexo</i>	<i>Idade (em anos)</i>	<i>Nível de instrução</i>	<i>Provedor</i>	
1	M	35	Superior	A	} <i>Dados</i>
2	F	18	Fundamental	C	
.	.	.	.	.	
.	.	.	.	.	
n	F	23	Médio	B	

*Indivíduos ou casos* →

### 4.1 Distribuição de frequências

- ⇒ Um dos primeiros passos para analisar um arquivo de dados, especialmente quando o número de observações for grande, é a distribuição de frequências de cada variável.
  - Uma distribuição de frequência é uma tabela que mostra categorias, valores ou intervalos de valores de acordo com as ocorrências.
- ⇒ Observe os exemplos de distribuições de frequências abaixo:

Tabela 4.1 – Distribuição de freqüências do provedor usado pelo visitante do site

Provedor	Freqüência	Porcentagem
A	10	25,0
B	17	42,5
C	7	17,5
D	6	15,0
Total	40	100,0

TABELA 4.2 – Distribuição de freqüências do grau de escolaridade dos visitantes do site

Grau de escolaridade	Freqüência	Porcentagem
Fundamental	8	20,0
Médio	14	35,0
Superior	18	45,0

Tabela 4.3 – Distribuição de freqüência do número de defeitos encontrados em escadas no final da linha de produção

Número de defeitos encontrados	Freqüência	Porcentagem	Freqüência acumulada	Porcentagem acumulada
0	13	27,1	13	27,1
1	15	31,3	28	58,3
2	10	20,8	38	79,2
3	7	14,6	45	93,8
4	2	4,2	47	97,9
7	1	2,1	48	100,0
Total	48	100,0		

**Observação importante:** Veja que na Tabela 4.3 há uma coluna contendo as *freqüências acumuladas* e as *porcentagens acumuladas* que são obtidas somando-se os valores menores ou iguais ao valor considerado. Essa coluna é calculada para variáveis quantitativas, mas existem variáveis qualitativas (especialmente as ordinais) em que a freqüência acumulada e a porcentagem acumulada também têm utilização prática.

- ⇒ Há situações em que uma variável discreta ou contínua apresenta uma quantidade muito grande de valores diferentes de modo que, se a tabela fosse construída nos mesmos moldes que as anteriores obteríamos praticamente os valores originais do banco de dados. A alternativa que vamos adotar consiste em construir intervalos de valores e contar o número de ocorrências em cada intervalo.
- ⇒ Esses intervalos devem ser mutuamente exclusivos e, de preferência, ter o mesmo tamanho. Porém, em alguns casos, não é possível usar intervalos com a mesma amplitude. Por exemplo, se formos analisar os salários de uma empresa, provavelmente encontraremos os valores distribuídos em uma grande amplitude e a

maioria deles concentrados na parte inferior da escala. Assim, pode ser conveniente usarmos intervalos menores para os valores iniciais e intervalos maiores (mais amplos) para valores finais.

⇒ O processo de construção de uma tabela de freqüências por intervalos que vamos utilizar envolve os seguintes passos:

1. Cálculo da amplitude:  $A = \max - \min$
2. Cálculo do número de *intervalos* da tabela:  $k = 1 + 3,222 \cdot \log n$
3. Cálculo da amplitude dos *intervalos*:  $A_k = \frac{A}{K}$

**Exemplo:** Os dados, a seguir, representam o tempo (em segundos) que operadores gastam para montar um equipamento na linha de produção de uma empresa:

4,7	5,1	5,7	6,3	7,3
4,8	5,2	5,7	6,4	8,2
4,8	5,3	5,7	6,5	8,2
4,9	5,4	5,9	6,7	8,3
4,9	5,4	6	6,8	8,4
4,9	5,4	6	6,8	8,9
4,9	5,5	6,2	6,9	9,1
5	5,6	6,2	7	9,9
5	5,6	6,2	7,1	14,1
5	5,7	6,3	7,3	15,2

Utilizando os passos acima, temos:

$$A = 15,2 - 4,7 = 10,5$$

$$K = 1 + 3,222 \cdot \log 50 = 1 + 3,222 \cdot 1,699 = 6,474 \text{ (pode-se escolher entre 6 ou 7 classes)}$$

$$A_k = 10,5 / 6 = 1,75 \text{ (nesse caso, escolhendo 6 classes, cada classe deverá ter uma amplitude de 1,8 segundos)}$$

Tabela 4.4 - Distribuição de freqüência do tempo gasto (em segundos) para a montagem de um equipamento na linha de produção

Tempo	Freqüência	Porcentagem	Freqüência acumulada	Porcentagem acumulada
4,7  — 6,5	32	64,0	32	64,0
6,5  — 8,3	11	22,0	43	86,0
8,3  — 10,1	5	10,0	48	96,0
10,1  — 11,9	0	0,0	48	96,0
11,9  — 13,7	0	0,0	48	96,0
13,7  — 15,5	2	4,0	50	100,0
Total	50	100,0	-	-

Observe que o símbolo |— é utilizado para representar intervalo aberto/fechado, ou seja, o valor que se encontra do lado direito do símbolo não está incluído no intervalo (intervalo aberto) e o valor que se encontra do lado esquerdo do intervalo está incluído no intervalo (intervalo fechado). Portanto, a primeira classe de valores 4,7 |— 6,5 indica que o intervalo vai de 4,7 a 6,49.

Os passos para construção de tabelas de frequências não são uma imposição legal, portanto, o pesquisador é livre para criar o número de classes com as amplitudes que lhe forem mais convenientes.

Ao analisar variáveis quantitativas, normalmente, três informações principais são procuradas:

- ✓ Faixa em que os valores ocorrem com maior frequência (faixa de valores típicos);
- ✓ Valores discrepantes, que podem ter sido originados de erros de mensuração ou digitação, mas também podem corresponder a elementos que apresentam comportamento muito diferente dos demais;
- ✓ Forma da distribuição, a fim de compará-la com modelos probabilísticos, o que nos permite usar técnicas mais avançadas de análise.

## 4.2 Tabelas de contingência

- As *tabelas de contingência* são a forma usual de apresentar uma distribuição de frequência conjunta de duas variáveis qualitativas.
- Objetivos: Apresentar sob a forma de uma tabela de dupla entrada a informação de duas variáveis categorizadas.
- Exemplos:

Tabela 4.5 – Distribuição de frequência do número de defeitos de acordo com a posição e o lado do pneu em veículos utilitários

Lado	Posição		Total
	Dianteira	Traseira	
Direito	32	28	60
Esquerdo	35	57	92
Total	67	85	152

Podemos calcular percentuais para cada célula da tabela, seja em relação ao total das linhas (lado do veículo), ao total das colunas (posição do pneu) ou ao total geral da tabela, o que possibilita uma avaliação da qualidade dos processos. Veja a tabela com os percentuais da linha e da coluna calculados a partir dos dados da Tabela 4.5.

Lado	Posição		Total
	Dianteira	Traseira	
Direito	32	28	60
% da linha	53%	47%	100%
% da coluna	48%	33%	39%
% total	21%	18%	39%
Esquerdo	35	57	92
% da linha	38%	62%	100%
% da coluna	52%	67%	61%
% total	23%	38%	61%
Total	67	85	152
% da linha	44%	56%	100%
% da coluna	100%	100%	100%
% total	44%	56%	100%

Tabela 4.6 – Distribuição de freqüência do número de defeitos de acordo com a posição, o lado e a marca do pneu em veículos utilitários

Marca	Lado	Posição		Total
		Dianteira	Traseira	
A	Direito	32	28	60
	Esquerdo	35	57	92
B	Direito	27	23	50
	Esquerdo	43	55	98
Total		137	163	300

## Exercícios de fixação

1. Sejam os dados a seguir, já ordenados do menor para o maior, de 26 observações, em decibéis, do nível de ruído de tráfego em certo cruzamento:

52,0	54,4	54,5	55,7	55,8	56,7	56,8
57,2	57,6	58,9	60,2	60,3	60,5	60,6
60,8	62,1	62,6	62,7	63,1	63,6	65,7
66,2	66,8	67,0	67,1	69,4		

- a) Organize os dados em uma distribuição de frequência.  
b) Escolha um valor da coluna de frequência relativa e um valor da coluna de frequência relativa acumulada e interprete-os.
2. Dados sobre a avaliação da qualidade de 300 restaurantes e preço típico das refeições são apresentados na tabela abaixo:

Avaliação da qualidade	Preço das refeições (em R\$)				Total
	10 a 19	20 a 29	30 a 39	40 a 49	
Bom	<b>42</b>	40	2	0	84
Ótimo	34	<b>64</b>	46	6	150
Excelente	2	14	<b>28<sup>A</sup></b>	22	66
Total	78	118	76	28	300

- a) Para cada um dos valores em negrito, calcule os % da linha, % da coluna e % total.  
b) Para o valor marcado com a letra A, interprete todos os percentuais calculados.

3. Os 208 funcionários de uma empresa são classificados em quatro níveis salariais (I, II, III e IV).

Sabe-se que, dos 140 funcionários que têm nível salarial I, 120 deles têm o nível médio de escolaridade e nenhum tem nível de escolaridade superior.

Cinquenta e dois funcionários pertencem ao nível salarial II. Destes, 10 têm nível médio de escolaridade e 2 tem nível superior.

Onze funcionários têm nível superior sendo que somente 5 deles pertencem ao nível salarial IV.

Dez funcionários pertencem ao nível salarial III, dentre estes, somente um tem nível fundamental de escolaridade.

Dentre os 161 funcionários que têm nível fundamental de escolaridade, nenhum pertence ao nível IV de escolaridade.

- a) A partir das informações fornecidas acima, construa uma distribuição de frequências cruzando o nível salarial com o grau de escolaridade.  
b) Refaça a tabela construída na letra A, substituindo as frequências pelos percentuais das colunas em cada uma das células.  
c) Escreva um comentário a partir dos percentuais obtidos.