

Aula 2 – Apresentação de dados - parte 2

Nesta aula, você aprenderá:

- a construir distribuições de frequências agrupadas para variáveis quantitativas discretas e contínuas;
- a construir histogramas e polígonos de frequência para representar distribuições de frequências agrupadas;
- a construir gráficos de linha e diagramas de ramos e folhas.

Apresentação de dados quantitativos - conclusão

Variáveis quantitativas discretas

Na aula anterior, você aprendeu a construir distribuições de frequências para variáveis discretas com poucos valores. No exemplo ali apresentado, há duas variáveis quantitativas discretas: número de dependentes e idade.

A diferença entre elas é que a idade pode assumir um número maior de valores, o que resultaria em uma tabela grande, caso decidíssemos relacionar todos os valores. Além disso, em geral não é necessário apresentar a informação em tal nível de detalhamento. Por exemplo, para as seguradoras de planos de saúde, as faixas etárias importantes - aquelas em que há reajuste por idade - são 0 a 18; 19 a 23; 24 a 28; 29 a 33; 34 a 38; 39 a 43; 44 a 48; 49 a 53; 54 a 58 e 59 ou mais. Sendo assim, podemos agrupar os funcionários segundo essas faixas etárias e construir uma *tabela de frequências agrupadas* da mesma forma que fizemos para o número de dependentes, só que agora cada frequência corresponde ao número de funcionários na respectiva faixa etária. Na **Tabela 2.1** temos a tabela resultante.

TABELA DE FREQUÊNCIAS
AGRUPADAS

Tabela 2.1: Distribuição de frequência da idade de 500 funcionários

Faixa Etária	Frequência Simples		Frequência Acumulada	
	Absoluta	Relativa %	Absoluta	Relativa %
19 – 23	1	0,2	1	0,2
24 – 28	23	4,6	24	4,8
29 – 33	103	20,6	127	25,4
34 – 38	246	49,2	373	74,6
39 – 43	52	10,4	425	85,0
44 – 48	50	10,0	475	95,0
49 – 53	25	5,0	500	100,0
Total	500	100,0		

Atividade 2.1

Na **Tabela 2.2** temos as informações sobre idade e salário para os 15 funcionários do Departamento de Recursos Humanos. Construa uma tabela de frequências para a idade, levando em conta as mesmas faixas etárias utilizadas acima.

Tabela 2.2: Idade e salário dos funcionários do Departamento de RH

Nome	Idade	Salário
João da Silva	36	6300
Pedro Fernandes	51	5700
Maria Freitas	26	4500
Paula Gonçalves	25	3800
Ana Freitas	29	3200
Luiz Costa	53	7300
André Souza	42	7100
Patrícia Silva	38	5600
Regina Lima	35	6400
Alfredo Souza	45	7000
Margarete Cunha	26	3700
Pedro Barbosa	37	6500
Ricardo Alves	24	4000
Márcio Rezende	31	5100
Ana Carolina Chaves	29	4500

Variáveis quantitativas contínuas

Para as variáveis quantitativas contínuas, devemos também trabalhar com distribuições de frequências agrupadas. O processo de construção é idêntico ao visto para as variáveis discretas, mas aqui devemos tomar um cuidado especial na construção das classes. A escolha dos limites das classes deve ser feita com base na natureza, valores e unidade de medida dos dados. As únicas regras que têm que ser seguidas são as seguintes.

Regra: Definição das classes em uma distribuição de frequências agrupadas.

1. As classes têm que ser exaustivas, isto é, todos os elementos devem pertencer a alguma classe.
2. As classes têm que ser mutuamente exclusivas, isto é, cada elemento tem que pertencer a uma única classe.

O primeiro passo é definir o número de classes desejado; esse número, de preferência, deve estar entre 5 e 25. Em seguida, devemos determinar a *amplitude* dos dados, ou seja, o intervalo de variação dos valores observados da variável em estudo.

Definição

A **amplitude** de um conjunto de dados, representada por Δ_{total} , é definida como a diferença entre os valores máximo e mínimo:

$$\Delta_{total} = V_{Máx} - V_{Mín} \quad (2.1)$$

Se queremos trabalhar com classes de mesmo comprimento (e essa é uma opção bastante comum), para determinar esse comprimento, temos que dividir a amplitude total pelo número de classes desejado. No entanto, para garantir a inclusão dos valores mínimo e máximo, podemos, como regra geral, usar o seguinte procedimento: considere o primeiro múltiplo do número de classes maior que o valor da amplitude e use esse número como a nova amplitude. Por exemplo se a amplitude é 28 e queremos trabalhar com 5 classes, vamos considerar 30 como a nova amplitude. Dividindo esse valor

pelo número de classes, obtemos o comprimento de cada classe e os limites de classe podem ser obtidos somando-se o comprimento de classe a partir do valor mínimo dos dados. Continuando com o nosso exemplo, o comprimento de classe é $30 \div 5 = 6$; se o valor mínimo dos dados é 4, então os limites de classe serão

$$4 + 6 = 10$$

$$10 + 6 = 16$$

$$16 + 6 = 22$$

$$22 + 6 = 28$$

$$28 + 6 = 34$$

e as classes serão:

$$[4, 10)$$

$$[10, 16)$$

$$[16, 22)$$

$$[22, 28)$$

$$[28, 34)$$

Note o tipo de intervalo utilizado: para incluir o valor mínimo (4) na primeira classe, o intervalo tem que ser fechado na parte inferior: [4. Se fechássemos o intervalo no limite superior, o 10 estaria incluído na primeira classe e, portanto, não poderia estar na segunda classe. Isso resultaria em [4, 10] como a primeira classe e (10, 16) como a segunda classe. Ou seja, as duas primeiras classes estariam definidas de forma diferente, o que não é conveniente, pois dificulta a leitura da tabela. Assim, é preferível incluir o 10 na segunda classe, o que resulta nas classes apresentadas anteriormente.

Exemplo 2.1

Suponha que entre os 500 funcionários da nossa empresa, o menor salário seja 2800 e o maior salário seja de 12400. Para agrupar os dados em 5 classes devemos fazer o seguinte:

$$\Delta_{total} = V_{Máx} - V_{Mín} = 12400 - 2800 = 9600$$

$$\text{Próximo múltiplo de } 5 = 9605$$

$$\text{Comprimento de classe} = \frac{9605}{5} = 1921$$

Os limites de classe, então, são:

$$\begin{aligned}
 &2800 \\
 2800 + 1921 &= 4721 \\
 4721 + 1921 &= 6642 \\
 6642 + 1921 &= 8563 \\
 8563 + 1921 &= 10484 \\
 10484 - 1921 &= 12405
 \end{aligned}$$

e as classes podem ser definidas como

[2800, 4721)	(2800 incluído; 4721 excluído)
[4721, 6642)	(4721 incluído; 6642 excluído)
[6642, 8563)	(6642 incluído; 8563 excluído)
[8563, 10484)	(8563 incluído; 10484 excluído)
[10484, 12405)	(10484 incluído; 12405 excluído)

Essa é uma regra que resulta em classes corretamente definidas, mas nem sempre as classes resultantes são apropriadas ou convenientes. No exemplo acima, poderia ser preferível trabalhar com classes de comprimento 2000, definindo o limite inferior dos dados como 2500. Isso resultaria nas classes [2500,4500); [4500,6500); [6500,8500); [8500, 10500) e [10500, 12500), que são classes corretas e mais fáceis de ler.

Atividade 2.2

Construa uma distribuição de frequências agrupadas em 5 classes de mesmo comprimento para os dados de salários da **Tabela 2.2**.

Histogramas e polígonos de frequência

O histograma e o polígono de frequências são gráficos usados para representar uma distribuição de frequências simples de uma variável quantitativa contínua.

Um histograma é um conjunto de retângulos com bases sobre um eixo horizontal dividido de acordo com os comprimentos de classes, centros nos pontos médios das classes e *áreas proporcionais ou iguais às frequências*. Vamos ilustrar a construção de um histograma usando como exemplo a distribuição de frequência dos dados sobre salários da Atividade 2, reproduzida na **Tabela 2.3**.