

Distribuição de frequências para variáveis quantitativas

Prof. Walmes M. Zeviani

Departamento de Estatística
Universidade Federal do Paraná



- ▶ Tabelas de frequência.
- ▶ Agrupar valores em classes.
- ▶ Tipos de frequência.
- ▶ Cálculo de frequência e densidade.
- ▶ Gráficos para a distribuição de variáveis quantitativas.



Figura 1. Photo by nappy from Pexels. Fonte <https://rb.gy/nigarg>.

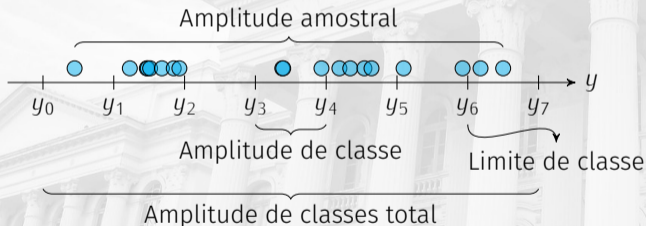


Tabelas de frequência

Definições

Amplitude de classe

Amplitude de classe é o comprimento de cada classe resultante do agrupamento ou compartimentalização de variáveis quantitativas.



Notação

$$\begin{array}{|l}
 a < y \leq b \\
 a \leq y < b
 \end{array}
 \left|
 \begin{array}{l}
 (a, b] \\
 [a, b)
 \end{array}
 \right|
 \begin{array}{l}
 a \longleftarrow b \\
 a \longrightarrow b
 \end{array}
 \left|
 \begin{array}{l}
 \\
 \\
 \end{array}
 \right.$$

Figura 2. Intervalo particionado em classes. Nesta ilustração $y_k (k = 1, \dots, 7)$ representa um limite de classe e não o k -ésimo valor observado da variável y . Fonte: os autores.

Frequência acumulada

Frequência acumulada é a frequência absoluta (ou relativa) acumulada conforme disposição das i) classes de uma variável ordinal ou ii) das classes de uma variável quantitativa que foi agrupada em classes.

Densidade

A densidade de uma classe é o quociente da frequência relativa (f_r) da classe pela amplitude da classe (a). Em gráficos que exibem a densidade no eixo y , a área abaixo da curva ou a soma dos retângulos é igual a 1.

Uma tabela de dados

Tabela 1. Um recorte aleatório de 10 observações da tabela de dados resultante da pesquisa feita no Moodle com alunos matriculados em Estatística que já responderam à enquete. Total de registros: 977 respondentes.

Área	Email	Adequação	Residentes	CH
Saúde	gmail	Tot. adeq.	(2,5]	240
Exatas	hotmail	Tot. adeq.	(2,5]	200
Biolog	outros	Tot. adeq.	(5,15]	270
Eng	gmail	Tot. inadeq.	(2,5]	480
Exatas	gmail	Adeq.	(2,5]	240
Eng	outros	Adeq.	(2,5]	300
Saúde	gmail	Adeq.	(2,5]	500
Exatas	hotmail	Adeq.	(2,5]	180
Saúde	gmail	Adeq.	(2,5]	150
Biolog	gmail	Tot. adeq.	(2,5]	225

Tabela 2. Tabela anterior com registros ordenados pelos valores da variável **CH**.

Área	Email	Adequação	Residentes	CH
Saúde	gmail	Adeq.	(2,5]	150
Exatas	hotmail	Adeq.	(2,5]	180
Exatas	hotmail	Tot. adeq.	(2,5]	200
Biolog	gmail	Tot. adeq.	(2,5]	225
Saúde	gmail	Tot. adeq.	(2,5]	240
Exatas	gmail	Adeq.	(2,5]	240
Biolog	outros	Tot. adeq.	(5,15]	270
Eng	outros	Adeq.	(2,5]	300
Eng	gmail	Tot. inadeq.	(2,5]	480
Saúde	gmail	Adeq.	(2,5]	500

Vetor de dados ordenado

Valores ordenados

150 180 200 225 240 240 270 300 480 500

Informações

- ▶ Número de dados: 10.
- ▶ Menor valor: 150.
- ▶ Maior valor: 500.
- ▶ Amplitude: 350.
- ▶ Amplitude de classe: 43.75?

Tabela de frequência do recorte

Tabela 3. Tabela de frequência considerando o agrupamento da CH em classes usando apenas os registros do recorte de 10 observações para exemplificar o procedimento.

CH	Frequência	Freq. relativa
(120,180]	2	0.20
(180,240]	4	0.40
(240,300]	2	0.20
(420,480]	1	0.10
(480,540]	1	0.10
Total	10	1.00

A tabela de frequência de todos os respondentes

Tabela 4. Tabela de frequência para as classes de CH para todos respondentes.

CH	Frequência	Freq. relativa
[0,60]	74	0.084
(60,120]	97	0.111
(120,180]	137	0.156
(180,240]	136	0.155
(240,300]	142	0.162
(300,360]	105	0.120
(360,420]	88	0.100
(420,480]	67	0.076
(480,540]	31	0.035
Total	877	1.000

Tabela 5. Tabela de frequência para as classes de CH para todos respondentes.

CH	Frequência	Freq. relativa	Freq. acumulada	Freq. rel. acum.
[0,60]	74	0.084	74	0.084
(60,120]	97	0.111	171	0.195
(120,180]	137	0.156	308	0.351
(180,240]	136	0.155	444	0.506
(240,300]	142	0.162	586	0.668
(300,360]	105	0.120	691	0.788
(360,420]	88	0.100	779	0.888
(420,480]	67	0.076	846	0.965
(480,540]	31	0.035	877	1.000
Total	877	1.000		

Formas de agrupar em classes

- ▶ Classes definidas com a mesma amplitude.
 - ▶ É o procedimento mais usual.
 - ▶ A frequência é definida em função do agrupamento em classes.
 - ▶ Evitar número de classes que gerem limites de classe complicados.
- ▶ Classes definidas com frequências aproximadamente iguais.
 - ▶ Os limites de classe são definidos em função da frequência.
 - ▶ Tentar obter frequências aproximadamente iguais.
 - ▶ Quando valores se repetem na amostra, quase sempre não é possível igualar frequências.
 - ▶ Tem relação com o conceito de separatrizes que ainda será visto.

Formas de agrupar em classes

Tabela 6. Frequência com classes de mesmo amplitude.

CH	Frequência	Freq. relativa
[0,100]	89	0.091
(100,200]	236	0.242
(200,300]	261	0.267
(300,400]	144	0.147
(400,500]	147	0.150
NA	100	0.102

Tabela 7. Frequências com classes de amplitude para valores próximos de frequência.

CH	Frequência	Freq. relativa
[0,150]	201	0.206
(150,210]	154	0.158
(210,300]	231	0.236
(300,379]	115	0.118
(379,500]	176	0.180
NA	100	0.102

Gráficos para frequência

Histograma da frequência absoluta

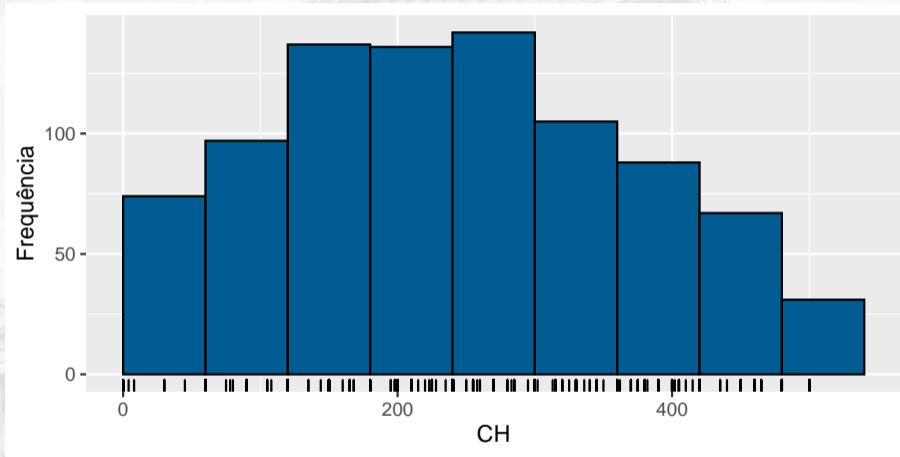


Figura 3. Histograma de frequência da variável CH.

O efeito do número de classes

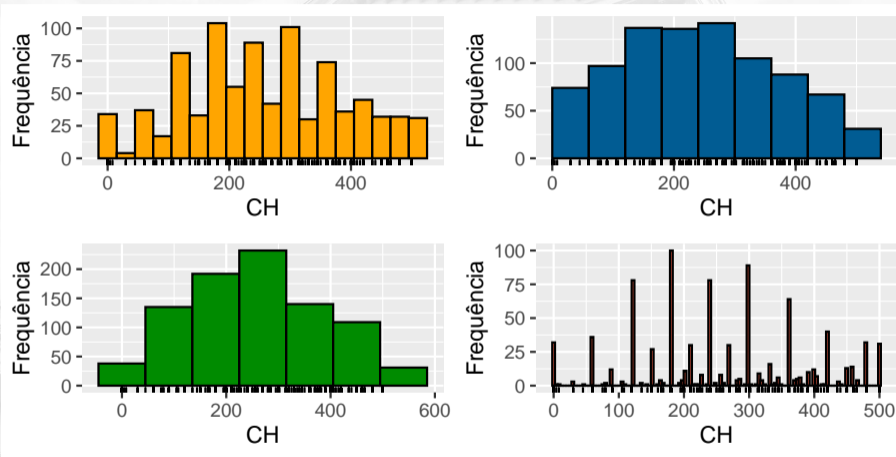


Figura 4. O efeito do número de classes na frequência e o resultado gráfico.

Histograma com amplitude de classe variável

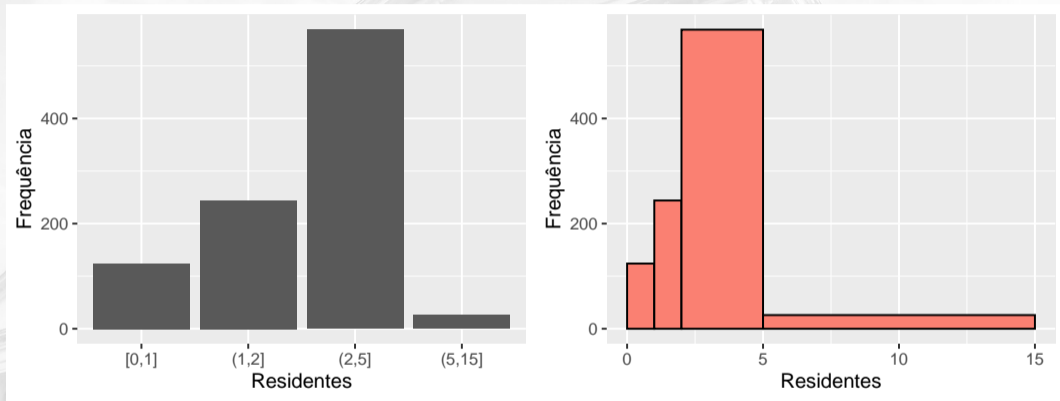


Figura 5. Frequências para a variável Residentes.

Histograma da densidade

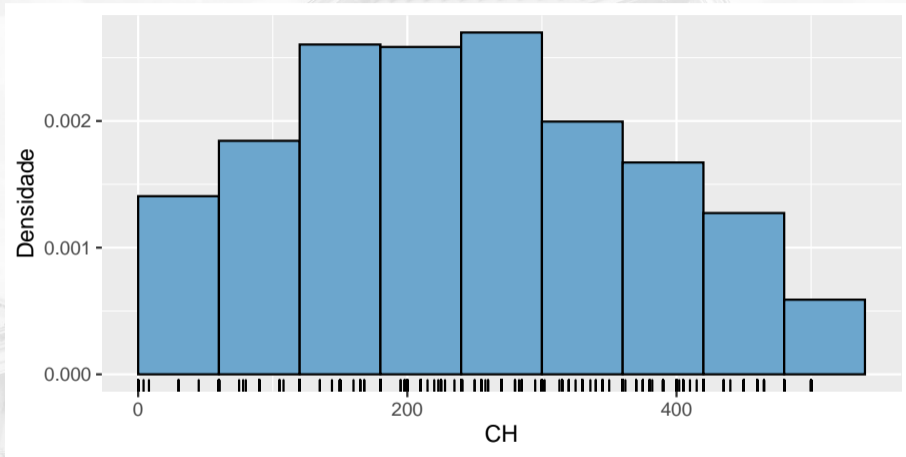


Figura 6. Histograma de densidade da variável CH.

A densidade

A densidade (d) é

$$d = \frac{f_r}{a}, \quad \text{logo} \quad f_r = a \cdot d,$$

em que f_r é a frequência relativa e a é a amplitude de classe.

Tabela 8. Exemplo de cálculo da densidade. Apenas as primeiras linhas da tabela são exibidas.

Residentes	Frequência	Freq. relativa	Amplitude	Densidade
(2,5]	569	0.591	3.000	0.197
(5,15]	26	0.027	10.000	0.003

Histograma da densidade para amplitude de classe variável

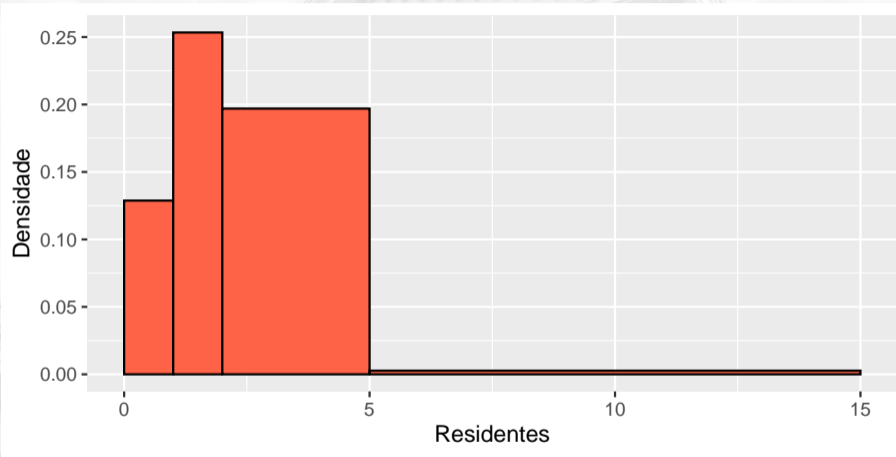


Figura 7. Histograma de densidade da variável Residentes.

A simetria da distribuição

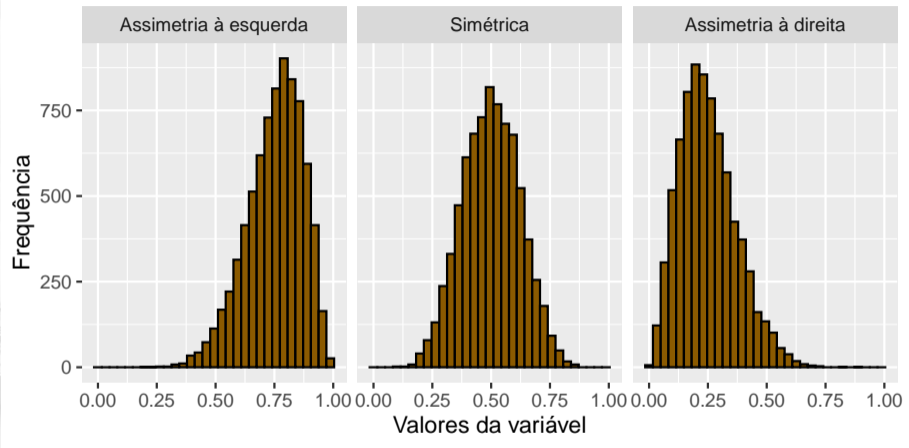


Figura 8. Histogramas exibindo assimetria à esquerda, simetria e assimetria à direita para uma variável contínua hipotética.

Gráfico de frequências acumuladas

- ▶ O comprimento vertical dos degraus é a f_r .
- ▶ O comprimento horizontal é a distância entre valores consecutivos da variável.
- ▶ Quando os valores não se repetem, $f_r = 1/n$.
- ▶ Quando valores se repetem, $f_r(y) = \#y/n$.
- ▶ A imagem do gráfico $\in [0, 1]$.

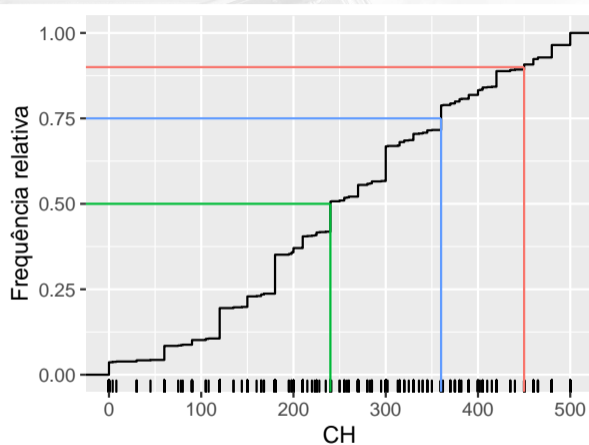


Figura 9. Gráfico de frequências acumuladas para a variável CH. Segmentos dentro do gráfico indicam como interpretar o gráfico.

Gráfico de densidade empírica

- ▶ Computacionalmente intensivo.
- ▶ Depende de escolher:
 - ▶ A função *kernel*.
 - ▶ Largura de banda ou tamanho de vizinhança.
 - ▶ A área abaixo da curva (em azul) é 1.
- ▶ Veja aplicação web:

<http://shiny.leg.ufpr.br/walmes/density/>.

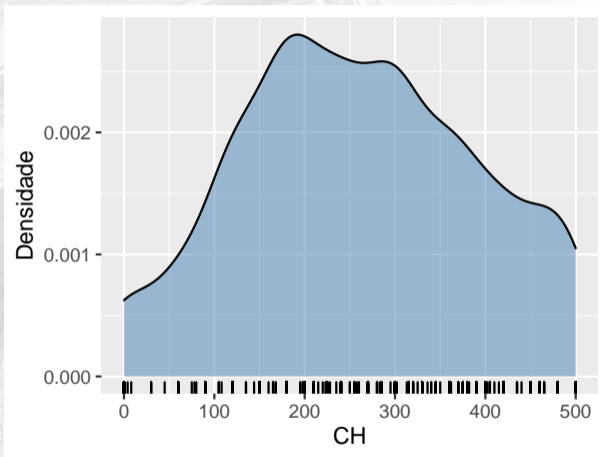


Figura 10. Gráfico de densidade empírica da variável CH.

Quando usar cada gráfico

- ▶ Histograma de frequência.
 - ▶ Para examinar a simetria da distribuição.
 - ▶ Para classes de mesma amplitude.
- ▶ Histograma de densidade.
 - ▶ Para examinar a simetria da distribuição.
 - ▶ Para classes de amplitude variável.
- ▶ Frequência acumulada.
 - ▶ Para examinar frequências relativas e separatrizes.
- ▶ Densidade.
 - ▶ Para examinar a simetria da distribuição.
 - ▶ Para determinar as modas da distribuição.

Alguns cuidados

- ▶ Histograma.
 - ▶ Escolha apropriada do número de classes agrupamento dos dados contínuos em classes.
 - ▶ Não usar barras separadas (como os feitos para variáveis qualitativas).
 - ▶ O histograma é um gráfico de barras mas que enfatiza a continuidade por não apresentar espaço entre barras.
 - ▶ Use densidade quando as classes são de amplitude variável.
- ▶ Densidade.
 - ▶ Escolha apropriadamente a função kernel e principalmente a largura de banda.
- ▶ Esteja ciente dos valores ausentes.



Considerações finais

Revisão

- ▶ Análise descritiva para variáveis quantitativas.
 - ▶ Agrupar valores em classe.
 - ▶ Tabelas de frequência.
 - ▶ Gráficos de frequência.
 - ▶ Histograma.
 - ▶ Frequência acumulada.
 - ▶ Densidade.
- ▶ Função da análise:
 - ▶ Uma variável: descrever o comportamento → distribuição de frequências.
 - ▶ Duas ou mais: investigar relações → será visto em outro vídeo.



Figura 11. Photo by nappy from Pexels. Fonte <https://rb.gy/nigarg>.