



Sintetizando dados qualitativos e quantitativos

Universidade Estadual de Santa Cruz

Ivan Bezerra Allaman

Cronograma

1. Dados qualitativos
2. Dados quantitativos

INTRODUÇÃO

- Para sintetizarmos as informações provenientes de dados qualitativos podem ser utilizados tanto métodos tabulares quanto métodos gráficos.
- Com relação ao método tabular utiliza-se a **distribuição de frequência**.
- Uma **distribuição de frequência** é um sumário tabular de dados que mostra o número (frequência) de itens em cada uma das diversas classes não sobrepostas.

Itens de uma tabela de distribuição de frequências

- Categoria ou classe no qual se está estudando;
- Frequência absoluta: corresponde ao número de itens de cada categoria estudada;
- Frequência acumulada: corresponde a frequência acumulada na classe anterior mais a frequência absoluta da classe atual;

- Frequência relativa: equivale a fração ou proporção dos itens pertencentes aquela classe

$$\text{Frequência relativa} = \frac{\text{Frequência absoluta}}{n}$$

- Frequência relativa (%): é a frequência relativa multiplicada por 100;
- Frequência acumulada (%): é a frequência relativa acumulada multiplicada por 100;

• Visualizando-a:

Classes	Freq. absoluta	Freq. relativa	Freq. relativa(%)	Freq. acumulada	Freq. acumulada(%)
...

SINTETIZANDO DADOS QUALITATIVOS

Aplicação

1. Uma pesquisa foi realizada com o intuito de levantar conhecimentos sobre o *nível de instrução do chefe da casa*, nas famílias residentes no bairro Saco Grande II, Florianópolis - SC (Barbetta, 2012). Os dados se encontram no link <https://lec.pro.br/download/R/dados/barbetta.txt>. Faça uma síntese das variáveis qualitativas.

- Percebam que nesta base de dados há duas variáveis qualitativas: *p.a.p* (*programa de alimentação popular*) e *instr.* (*grau de instrução*).
- A variável *p.a.p* está na escala nominal e *instr.* na escala ordinal.
- Logo, a tabela de distribuição de frequências se resume a:

Classe	Freq. absoluta	Freq. relativa	Freq. relativa (%)	Freq. acumulada	Freq. acumulada (%)
Usa	78	0,65	65	78	65
Não usa	42	0,35	35	120	100

- Quando a variável está na escala ordinal, é coerente que as classes sejam apresentadas em sua ordem natural.
- No caso da variável grau de escolaridade tem-se:

Classe	Freq. absoluta	Freq. relativa	Freq. relativa(%)	Freq. acumulada	Freq. acumulada(%)
nenhum	38	0,3167	31,67	38	31,67
prgrau	38	0,3167	31,67	76	63,33
seggrau	44	0,367	36,67	120	100,00

2. O proprietário de uma cabana que se destina a criação de carneiros, apresentou ao técnico uma ficha de campo referente aos principais problemas de mortalidades de cordeiros nos últimos 3 meses ocorridos em sua propriedade. O proprietário quer saber quais foram os principais problemas ocorridos no período considerado. Segue os dados:

Aborto	Aborto	Aborto	Distoquia	Aborto
Aborto	Frio/Desnutrição	Distoquia	Aborto	Frio/Desnutrição
Frio/Desnutrição	Frio/Desnutrição	Distoquia	Aborto	Aborto
Aborto	Distoquia	Aborto	Aborto	Infecção
Distoquia	Predação/Acidente	Aborto	Aborto	Outros
Aborto	Aborto	Aborto	Distoquia	Aborto
Infecção	Frio/Desnutrição	Aborto	Frio/Desnutrição	Frio/Desnutrição
Infecção	Distoquia	Outros	Frio/Desnutrição	Infecção

APRESENTAÇÃO GRÁFICA

- Dentre os diversos tipos de gráficos existentes, será abordado apenas os mais utilizados e, cujas informações advêm de uma tabela de distribuição de frequências.
- Logo, serão apresentados os seguintes gráficos: o gráfico de *barras*, de *setores*, de *pontos* e o gráfico de *pareto*.

Gráfico de barras

- Normalmente utilizado para variáveis qualitativas.
- A altura da coluna pode corresponder a frequência absoluta, frequência relativa ou frequência percentual.
- A disposição das colunas pode ser tanto no sentido horizontal quanto vertical.

Aplicação

3. Aproveitando os dados da aplicação 1 elabore um gráfico de barras para a variável "p.a.p.".

Segue o gráfico.

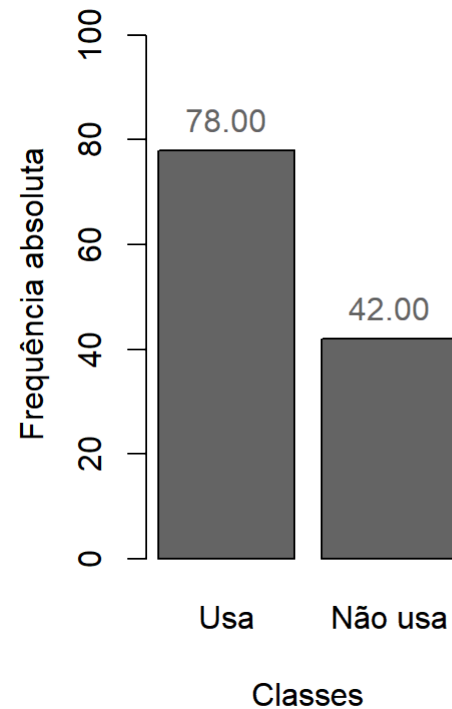


Gráfico de setores

- É um tipo de gráfico útil quando se tem poucas categorias. Em geral, quando o número de categorias ultrapassa 6 a visualização das categorias se torna difícil.
- A confecção do gráfico manualmente nos dias atuais é inviável devido a grande quantidade de recursos computacionais. No entanto, caso deseje fazer manualmente, deverá ter o auxílio de um transferidor. O ângulo de cada setor, é calculado multiplicando a frequência relativa por 360° .

Aplicação

4. Utilizando ainda os dados da aplicação 1, elabore um gráfico de setores para a variável "grau de instrução".

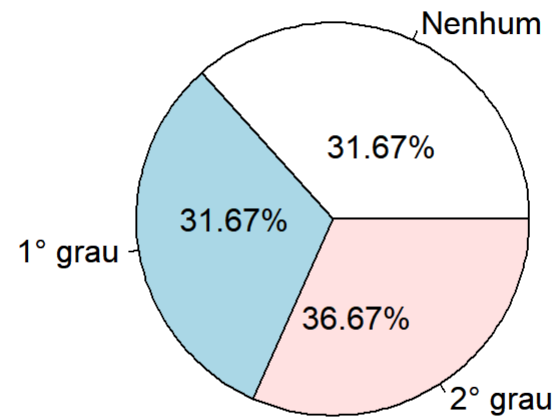


Gráfico de pontos (Dot Chart)

- É um tipo de gráfico pouco conhecido e conseqüentemente pouco utilizado.
- No entanto, embora visualmente seja simples, é extremamente útil e eficaz quando comparado aos gráficos tradicionais de barra e de setores.

Aplicação

5. Considerando ainda a aplicação 1, elabore o gráfico de pontos para a variável "p.a.p.".

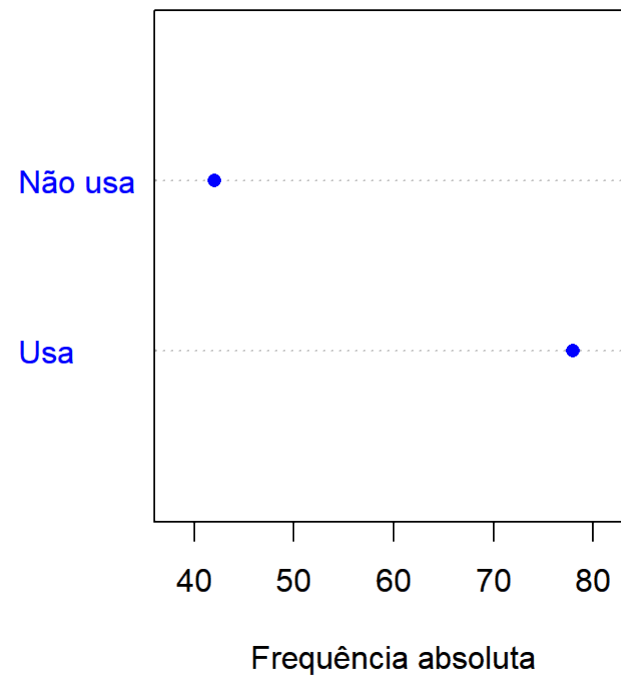
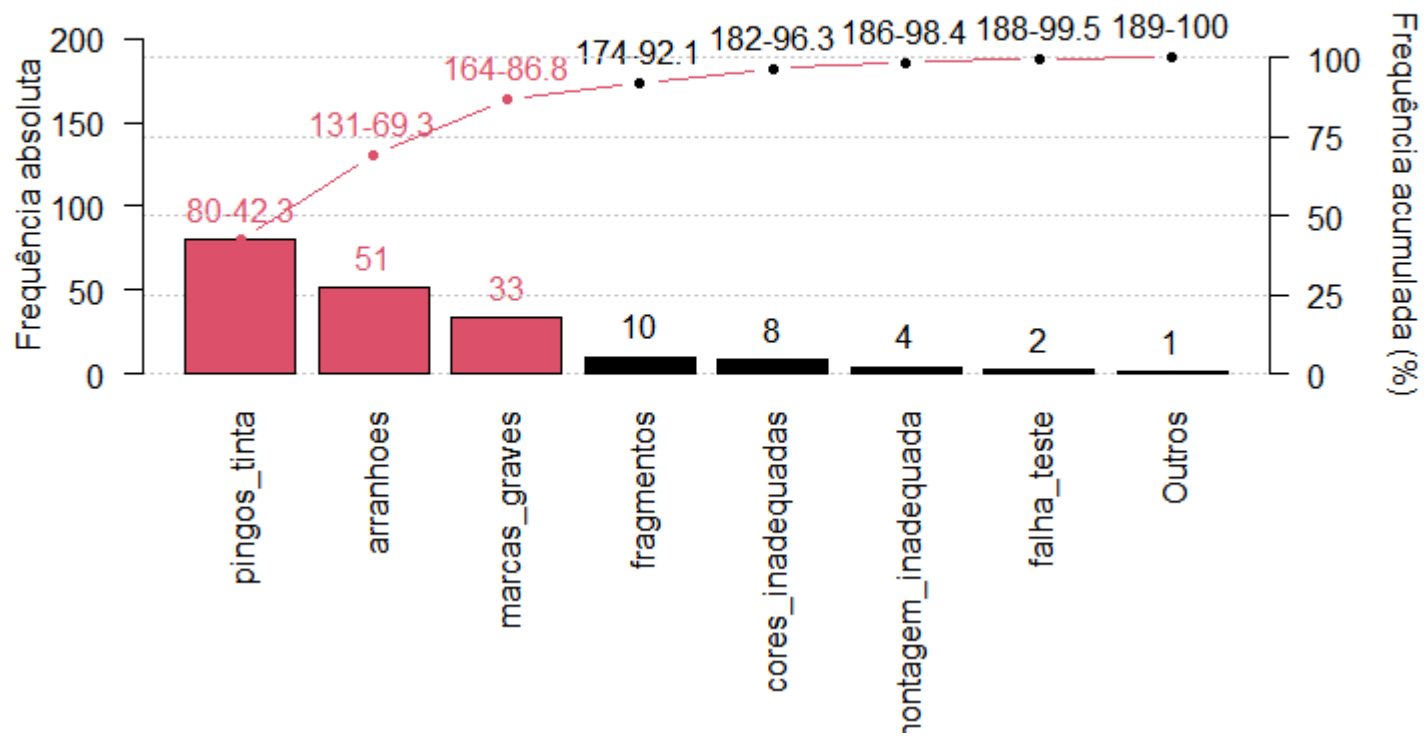


Gráfico de Pareto

- É uma junção entre os gráficos de barra e um gráfico de linhas correspondente a frequência acumulada. Neste caso, as categorias devem ser colocadas em ordem decrescente pela frequência (absoluta ou relativa).
- É muito utilizado em situações que envolvem controle de qualidade cujo o intuito é a priorização dos problemas.
- Segundo Pareto, 80% das consequências advêm de 20% das causas, ou seja, há muitos problemas sem importância diante de outros mais graves.

Aplicação

6. Em uma determinada fábrica de brinquedos foi feito um levantamento durante uma semana a cerca de um mesmo tipo de brinquedo com o intuito de verificar quais eram os principais defeitos que ocorriam durante a fabricação. Os dados se encontram no seguinte link:
<https://lec.pro.br/download/R/dados/salvendy.txt>. Elabore um gráfico de pareto.



SÍNTESE DE DADOS QUANTITATIVOS

- No caso de variáveis discretas, há duas possibilidades de apresentação tabular.
 - Do mesmo modo de variáveis qualitativas,
 - Utilizando a técnica para variáveis contínuas,

Quando há muitas categorias, a apresentação tabular do mesmo modo que variáveis qualitativas não será muito útil, já que, o objetivo é sintetizar, como podemos observar no exemplo a seguir:

Categoria	Freq. absoluta	Freq. relativa	Freq. relativa(%)	Freq. acumulada	Freq. acumulada (%)
11	1	0.02	2	1	2
13	3	0.06	6	4	8
16	4	0.08	8	8	16
17	6	0.12	12	14	28
18	7	0.14	14	21	42
19	2	0.04	4	23	46
20	1	0.02	2	24	48
21	5	0.10	10	29	58
22	6	0.12	12	35	70
23	5	0.10	10	40	80
24	3	0.06	6	43	86
25	3	0.06	6	46	92
26	2	0.04	4	48	96
27	1	0.02	2	49	98
28	1	0.02	2	50	100

- No caso de variáveis contínuas, são necessários alguns passos para elaborar a tabela de distribuição de frequências:
 - Ordenar os dados em ordem crescente
 - Determinar o número de classes da tabela (k). As metodologias são:
 - Critério de Sturges:

$$k = 1 + \log_2 n,$$
 em que n é o tamanho da amostra.
 - Critério de Scott:

$$k = \frac{A}{3.49 \cdot s \cdot n^{-1/3}},$$
 em que A é a amplitude da classe (maior valor menos o menor) e s é o desvio padrão.
 - Critério de Freedman-Diaconis:

$$k = \frac{A}{2 \cdot IQR \cdot n^{-1/3}},$$
 em que IQR é a amplitude inter-quartílica (3º quartil menos o 1º quartil)

- Determinar o limite inferior da primeira classe: $Li = X_1 - |X_1|/100$
 - Determinar o limite superior da última classe: $Ls = X_n + |X_n|/100$
 - Determinar o comprimento das classes: $h = (Ls - Li)/k$
- O pacote `fdth` ([Faria et al. 2014](#)) elabora uma tabela de distribuição de frequências utilizando a teoria abordada anteriormente.

Aplicação

7. Uma característica importante da qualidade da água é a concentração de material sólido suspenso. A seguir, estão 20 medições de sólidos suspensos de um certo lago. Construa uma tabela de distribuição de frequências. 42,4 65,7 29,8 58,7 52,1 55,8 57,0 68,7 67,3 67,3 54,3 54,0 73,1 81,3 59,9 56,9 62,2 69,9 66,9 59,0

• Ordenando os dados de modo crescente.

29.8	42.4	52.1	54.0	54.3	55.8	56.9	57.0	58.7	59.0
59.9	62.2	65.7	66.9	67.3	67.3	68.7	69.9	73.1	81.3

- Determinando o número de classes pelo critério de Sturges.

$$\begin{aligned}n &= 20 \\k &= 1 + \log_2(n) \\k &= 5,32 \approx 6\end{aligned}$$

- Determinando os limites inferior e superior.

$$\begin{aligned}Li &= 29,8 - \frac{29,8}{100} = 29,50 \\Ls &= 81,3 + \frac{81,3}{100} = 82,11\end{aligned}$$

- Determinando o comprimento de cada classe.

$$h = \frac{82,11 - 29,5}{6} = 8,77$$

- Portanto teremos a seguinte tabela:

Intervalo de classes	Freq. absoluta	Freq. relativa	Freq. relativa(%)	Freq. acumulada	Freq. acumulada(%)
[29.5,38.27)	1	0.05	5	1	5
[38.27,47.04)	1	0.05	5	2	10
[47.04,55.81)	4	0.20	20	6	30
[55.81,64.58)	6	0.30	30	12	60
[64.58,73.34)	7	0.35	35	19	95
[73.34,82.11)	1	0.05	5	20	100

APRESENTAÇÃO GRÁFICA

- Dentre os gráficos existentes para variáveis contínuas será apresentado os mais importantes que são o histograma, o polígono de frequências e a ogiva.

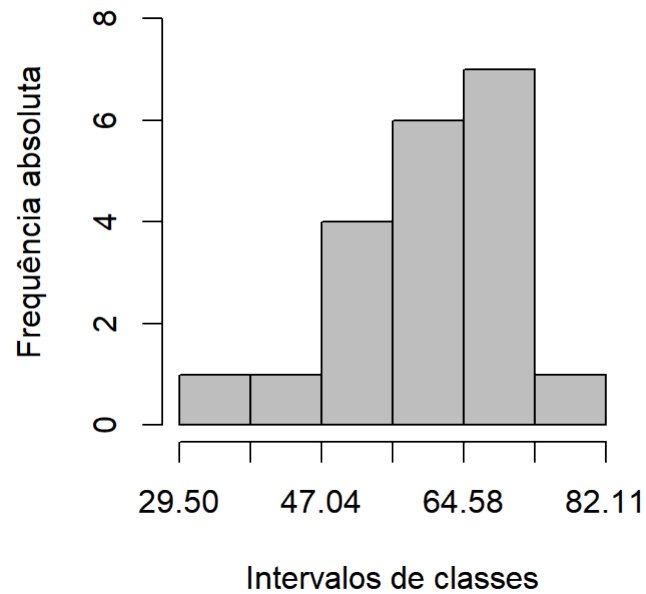
Histograma

- O histograma é um gráfico utilizado para plotar as frequências absolutas em função das classes provenientes de uma tabela de distribuição de frequências.
- O histograma é também um estimador da distribuição de probabilidade da variável em estudo. Neste caso, utiliza-se a **densidade** ao invés da frequência absoluta ou relativa, uma vez que, a área correspondente aos retângulos devem somar um, que é o espaço amostral.
 - A densidade é calculada como: $\text{densidade} = \text{frequência relativa} \div h$

- Neste gráfico, as barras são justapostas. Destaca-se:
 - A forma da variável, ou seja, sua distribuição de probabilidade.
 - O grau de simetria da variável, que pode ser classificada em simétrica, assimétrica à direita ou assimétrica à esquerda.

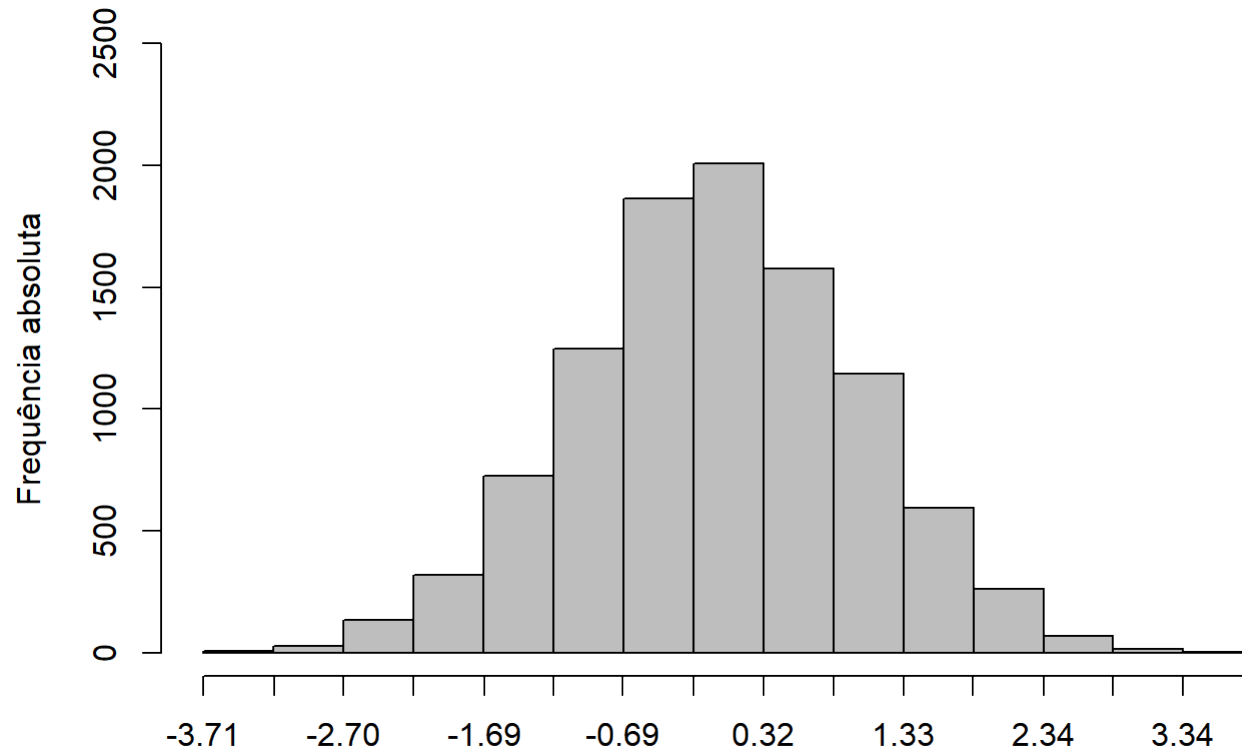
Aplicação

8. Aproveitando os dados da aplicação 6 elabore um histograma.

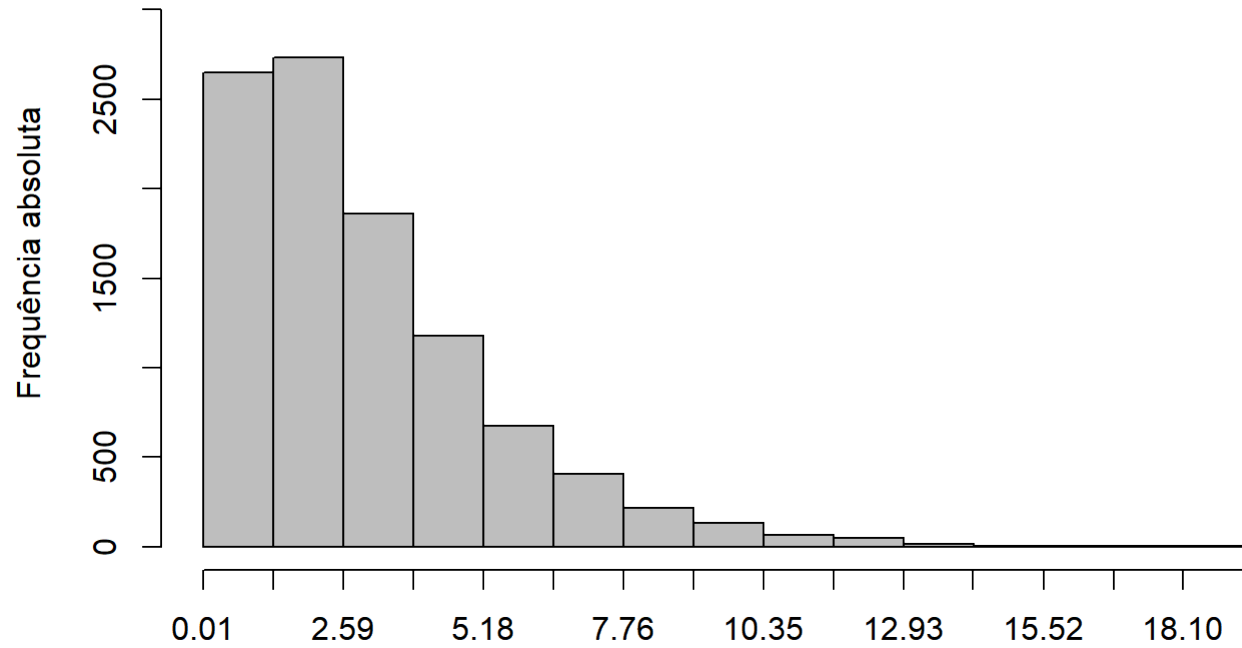


Percebamos que na variável sp a distribuição é classificada como assimétrica à esquerda.

- No exemplo abaixo, tem-se um tipo de distribuição simétrica.



- Segue um exemplo de assimetria à direita.



Polígono de frequências

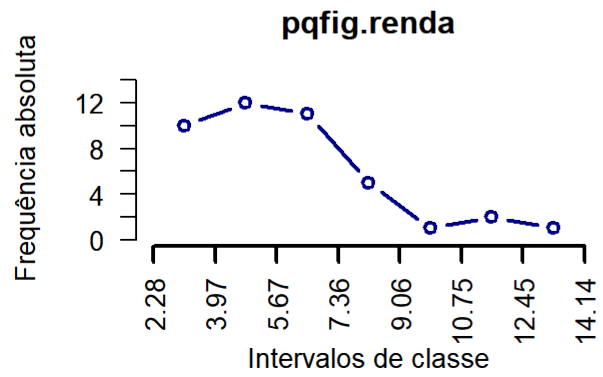
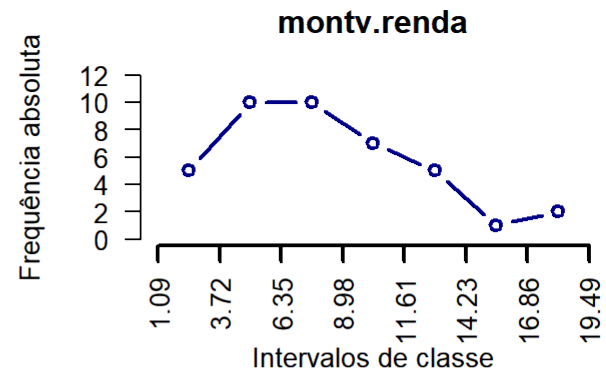
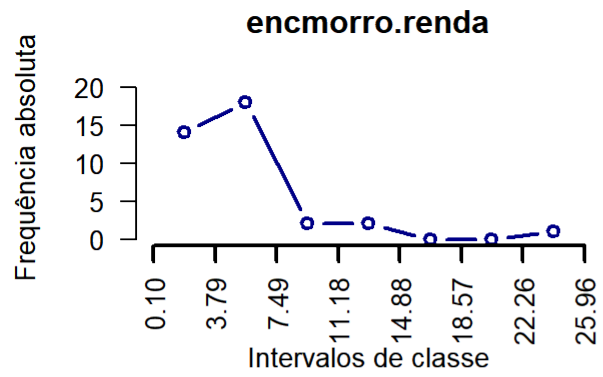
- É um gráfico de linhas interligados por meio do ponto médio de cada barra do histograma.
- É útil quando se deseja comparar mais de uma base de dados, evitando deste modo, a sobreposição de histogramas.

Se o intuito for comparar bases de dados, as mesmas devem ter o mesmo intervalo de classe.

Aplicação

9. Aproveitando os dados da aplicação 1, compare a renda familiar entre as localidades.

Neste caso vamos utilizar o polígono de frequência separado para cada localidade com diferentes intervalos de classe para percebemos a dificuldade na comparação.



- Vamos agora colocar os polígonos em um mesmo gráfico.
- Primeiramente vamos ajustar todos os dados para um mesmo intervalo de classe.
 - Para isso, faça uma tabela em separado para cada localidade e pegue o menor valor dos limites inferiores;
 - Em tal exemplo temos: Enc. morro = 0,099; Mont. verde = 1,089; Pq. da figueira = 2,3. Portanto vamos utilizar 0,099. Neste caso é mais prático se arredondarmos para um inteiro. Logo, será 0.

- - Pegue o maior valor dos limites superiores. Utilize um múltiplo do limite que foi escolhido no passo anterior.
 - Em tal exemplo temos: Enc. morro = 25,96; Mont. verde = 19,49; Pq. da figueira = 14. Portanto vamos utilizar 25,96. Arredondando vamos utilizar 26.
- O comprimento de classe (h) utilizado em todas as classes foi em torno de 3. No entanto, deve ser múltiplo também dos limites escolhidos. Logo, será o número 2.

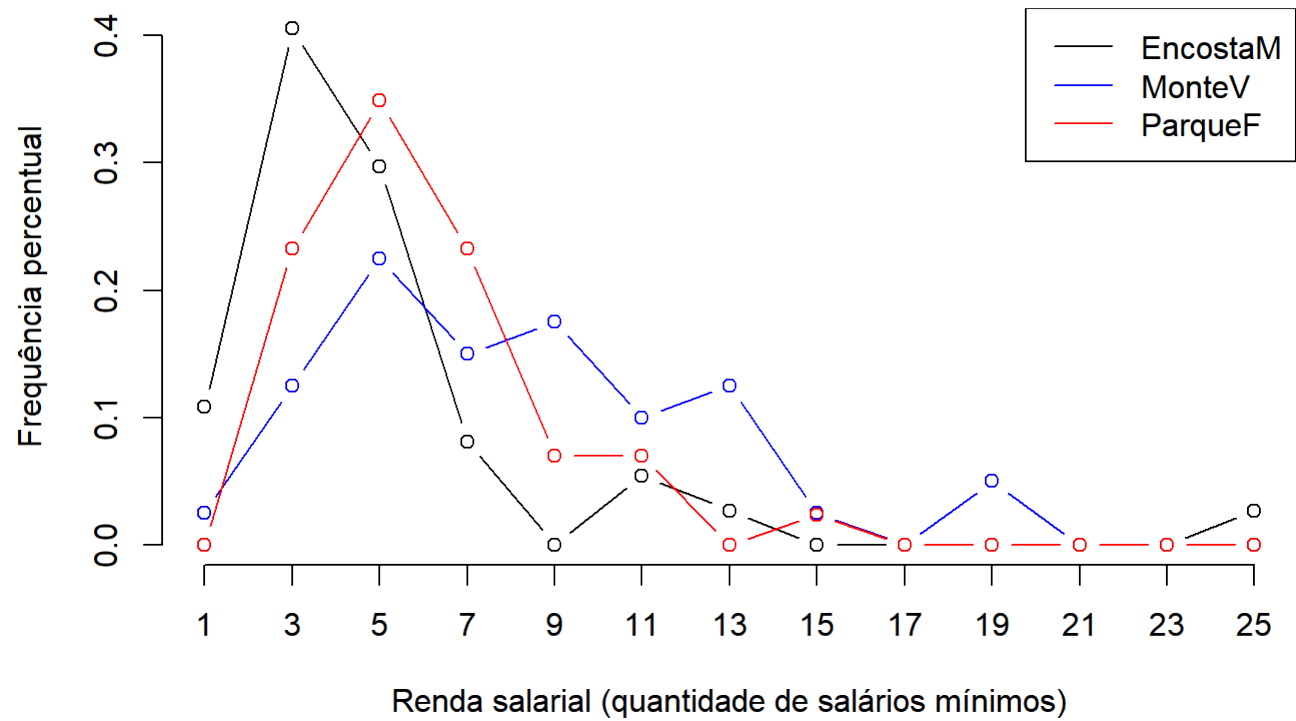
• Logo, temos as seguintes tabelas:

Intervalo de classes	Freq. relativa(%)		
	Encosta do morro	Monte verde	Parque da figueira
[0,2)	10.81	2.5	0.0
[2,4)	40.54	12.5	23.81
[4,6)	29.73	22.5	35.71
[6,8)	8.11	15	23.81
[8,10)	0.00	17.5	7.14
[10,12)	5.41	10.0	7.14
[12,14)	2.70	12.5	0.0
[14,16)	0.00	2.5	2.38
[16,18)	0.00	0.00	0.00
[18,20)	0.00	5.00	0.00
[20,22)	0.00	0.00	0.00
[22,24)	0.00	0.00	0.00
[24,26)	2.70	0.00	0.00

• Os valores no eixo x é o ponto médio de cada classe. Logo, os valores no eixo x foram:

1	5	9	13	17	21	25
3	7	11	15	19	23	

- Plotando em um mesmo gráfico.



Ogiva

- Utiliza-se as frequências acumuladas absolutas, acumuladas relativas ou acumuladas percentuais para elaboração do gráfico.

Aplicação

10. Aproveitando os dados da aplicação 9, elabore uma ogiva e compare as três localidades.

