

3

Lista 2

Cap. 3 – Descrição e Apresentação de Dados

Ex. 1:

Deve-se montar um banco de dados em que cada *linha* corresponde a uma *unidade amostral* (um pãozinho) e cada *coluna* contém um *atributo* dessa unidade amostral.

Assim, a primeira coluna indica o dia de fabricação, a segunda, a identificação do padeiro, a terceira, a máquina, e, finalmente, a última, para os pesos. Veja exemplo ao lado, com as 20 primeiras linhas do banco de dados.

Dia	Padeiro	Máquina	Peso
1	A	1	209,2
1	A	1	209,5
1	A	1	210,2
1	A	1	212,0
1	A	2	214,3
1	A	2	221,8
1	A	2	214,6
1	A	2	214,4
2	A	1	208,5
2	A	1	208,7
2	A	1	206,2
2	A	1	207,8
2	A	2	215,3
2	A	2	216,7
2	A	2	212,3
2	A	2	212,0
3	A	1	204,2
3	A	1	210,2
3	A	1	210,5
3	A	1	205,9

Cap. 3 – Descrição e Apresentação de Dados (cont.)

Ex. 2:

- a) É uma afirmação falsa. A média tem valor menor que pelo menos um dos dados. A mediana é que está sempre abaixo da metade dos dados.
- b) Também é falso. A média é representativa do conjunto de dados apenas se esse conjunto for simétrico.
- c) Depende. Se for a média for entendida aqui como o valor para aprovação, 60 pontos, é verdade. Se for a média referida for a média da turma, então isso é falso, pois sempre haverá alguém abaixo da média do conjunto de dados, a menos que todos tenham a mesma nota.

Cap. 3 – Descrição e Apresentação de Dados (cont.)

Ex. 3 - Seja o conjunto de dados:

6	12	12	14	15	15	15	15	16
17	18	18	19	19	19	20	21	21
22	22	22	23	23	23	23	23	23
24	25	25	25	27	27	28	32	N/A

a) A média é

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{6 + 12 + \dots + 32}{35} = \frac{709}{35} \cong 20,26$$

Em um banco de dados com número ímpar de elementos, a mediana é o elemento central, $\frac{1}{2} \times (35+1) = 18$, portanto

$$\tilde{x} = x_{18} = 21$$

b) O primeiro quartil é o elemento que supera 25% dos dados, ou seja, é o x_i tal que $i = \frac{1}{4} \times 36 = 9$. Logo $Q_1 = x_9 = 16$. O terceiro quartil, que supera 75% dos elementos, é $Q_3 = x_{27} = 23$.

Cap. 3 – Descrição e Apresentação de Dados (cont.)

Ex. 3 (cont.) – Sejam as turmas A e B:

Turma	Média	Desvio-padrão
A	22,5	4,5
B	24,0	5,4

c) A turma C tem desvio-padrão igual a

$$s = \sqrt{\frac{[\sum_{i=1}^n (x_i)^2] - n\bar{x}^2}{n-1}} = \sqrt{\frac{[6^2 + 12^2 + \dots + 32^2] - 35 \times 20,26^2}{35-1}} \cong 5,31$$

Calculando-se os coeficientes de variação das três turmas, tem-se

$$CV_A = \frac{s_A}{\bar{x}_A} = \frac{4,5}{22,5} = 0,20; \quad CV_B = \frac{5,4}{24,0} \cong 0,23 \text{ e } CV_C = \frac{5,31}{20,26} \cong 0,26$$

Assim, a turma A é a mais homogênea e a C é aquela que apresenta a maior variabilidade.

Cap. 3 – Descrição e Apresentação de Dados (cont.)

Ex. 4 – Seja o conjunto de dados do Ex. 3.2 (dados ausentes iguais a 9):

278	182	247	227	277	194	196	276	244	192
118	219	255	201	9	209	219	228	209	209
171	213	233	226	209	200	200	363	209	200
179	167	192	277	317	146	217	292	217	255
212	233	250	243	150	209	184	199	250	479
175	194	221	233	9	184	217	150	167	265
242	180	255	170	209	161	196	165	234	179
248	184	291	185	242	276	243	229	242	250

Seguindo os passos das notas de aula do Cap. 3 (Estatística Descritiva, p. 16), temos:

Menor valor = **118**

Maior valor = **479**

Número de classes = $\sqrt{78} \cong 9$

Amplitude = $479 - 118 = 361$

Tamanho de cada classe = $361 \div 8 \cong 50$

Cap. 3 – Descrição e Apresentação de Dados (cont.)

Ex. 4 (cont.) :

Dessa forma, tem-se a seguinte tabela de frequências:

Nível de colesterol	Frequência absoluta			Frequência relativa	
	simples	acumulada		simples	acumulada
100 † 150	1	1		1,28%	1,28%
150 † 200	25	26		32,1%	33,3%
200 † 250	35	61		44,9%	78,2%
250 † 300	14	75		17,9%	96,2%
300 † 350	1	76		1,28%	97,4%
350 † 400	1	77		1,28%	98,7%
400 † 450	0	77		0,00%	98,7%
450 † 500	1	78		1,28%	100%
Total	78	-		100%	-

Cap. 3 – Descrição e Apresentação de Dados (cont.)

Ex. 5:

- i. Como a dosagem da albumina dessa pessoa foi o percentil 95, isso significa que essa dosagem supera as dosagens encontradas em 95% das pessoas saudáveis. É, portanto, elevada.
- ii. Já a dosagem do colesterol foi o percentil 5. Isso significa que o valor encontrado supera apenas 5% das dosagens das pessoas saudáveis. É, portanto, bastante baixo.

Cap. 3 – Descrição e Apresentação de Dados (cont.)

Ex. 6:

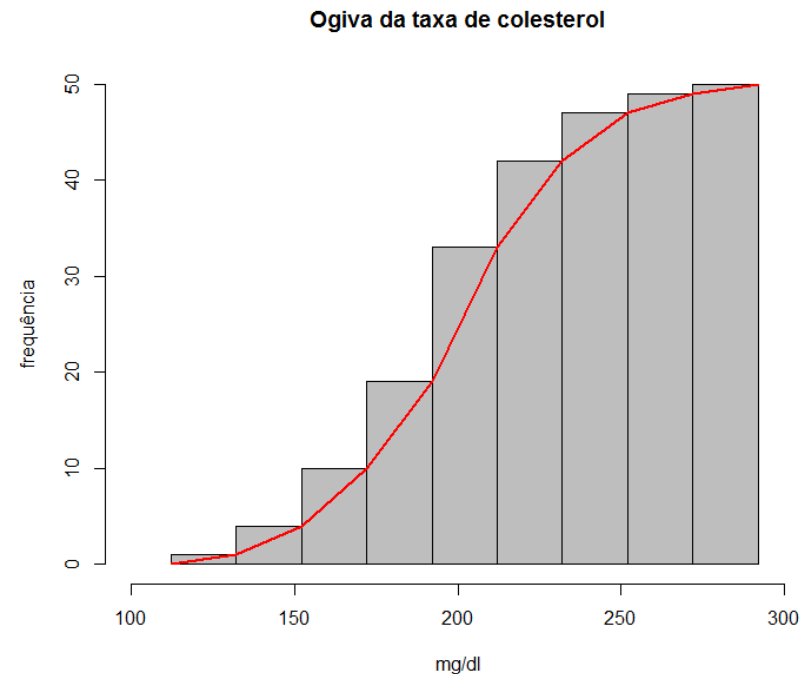
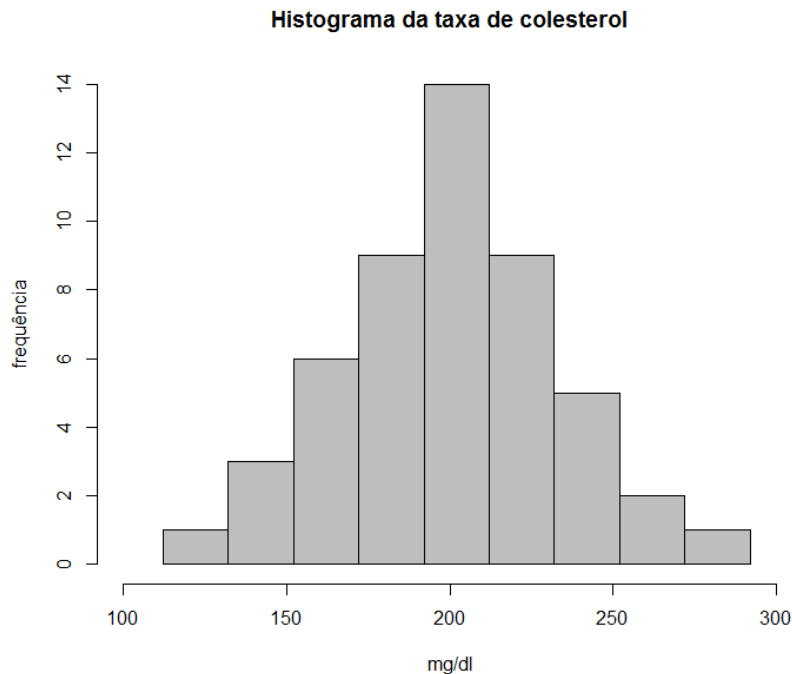
a) Completando-se a tabela, tem-se:

Taxa de colesterol	Frequência absoluta		pontos médios	Frequência relativa	
	simples	acumulada		simples	acumulada
112 † 132	1	1	122	2,00%	2,00%
132 † 152	3	4	142	6,00%	8,00%
152 † 172	6	10	162	12,0%	20,0%
172 † 192	9	19	182	18,0%	38,0%
192 † 212	14	33	202	28,0%	66,0%
212 † 232	9	42	222	18,0%	84,0%
232 † 252	5	47	242	10,0%	94,0%
252 † 272	2	49	262	4,00%	98,0%
272 † 292	1	50	282	2,00%	100%
Total	50	-		100%	-

Cap. 3 – Descrição e Apresentação de Dados (cont.)

Ex. 6 (cont.):

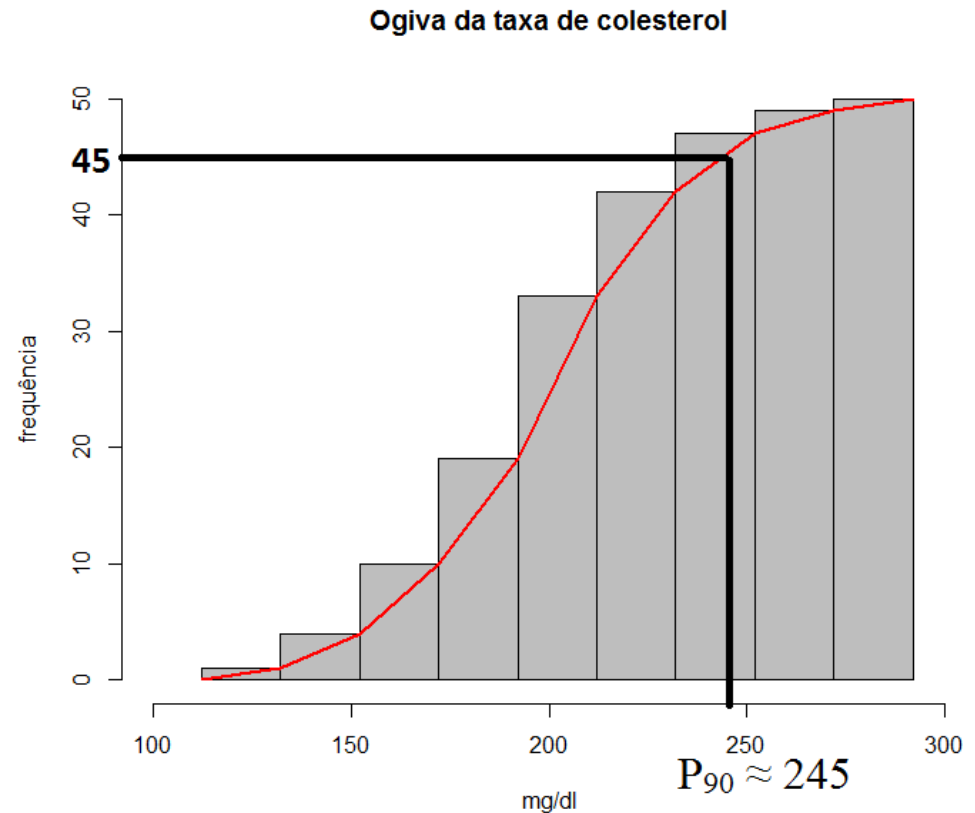
b) Construindo-se o histograma e a ogiva, tem-se:



Cap. 3 – Descrição e Apresentação de Dados (cont.)

Ex. 6 (cont.):

- c) O P_{90} é um valor que supera 90% das observações do banco de dados. É o elemento x_i , $i = 0,90 \times 50 = 45$. Graficamente, vale aproximadamente $P_{90} \approx 245$.



Cap. 3 – Descrição e Apresentação de Dados (cont.)

Ex. 6 (cont.):

- d) Para se comparar os indivíduos do grupo B, calcula-se a média e o desvio-padrão do grupo A, pela expressões de **dados agrupados** a seguir, em que x_i é o ponto médio das classes e f_i sua frequência:

$$\bar{x} = \frac{\sum_{i=1}^n (x_i \times f_i)}{\sum_{i=1}^n f_i} = \frac{(122 \times 1) + (142 \times 3) + \dots + (282 \times 1)}{1 + 3 + \dots + 1} = \frac{10000}{50} = 200,0$$

$$\begin{aligned} s &= \sqrt{\frac{[\sum_{i=1}^n (x_i^2 \times f_i)] - n\bar{x}^2}{n - 1}} = \sqrt{\frac{[(122^2 \times 1) + (142^2 \times 3) + \dots + (282^2 \times 1)] - 50 \times 200,0^2}{50 - 1}} \\ &= \sqrt{\frac{2055400 - 2000000}{50 - 1}} = \sqrt{\frac{55400}{49}} \cong 33,62 \end{aligned}$$

Portanto, calculando-se

$$CV_A = \frac{s}{\bar{x}} = \frac{33,62}{200} = 0,17 \text{ e } CV_B = \frac{s}{\bar{x}} = \frac{48}{254} = 0,19,$$

pode-se afirmar que o grupo A é mais homogêneo por apresentar menor CV .

- e) Pela comparação das médias, verifica-se um aumento do colesterol com a idade.

Cap. 3 – Descrição e Apresentação de Dados (cont.)

Ex. 7:

Sejam os seguinte dados obtidos dos laboratórios:

Colesterol	284	295	330	228	365
	335	361	333	294	318
Triglicérides	-	166	158	202	162
	135	82	150	86	121

- a) Quantificam-se as variações encontradas, calculando-se as médias, os desvios-padrões e os coeficientes de variação:

$$\bar{x}_C = \frac{\sum_{i=1}^n x_i}{n} = \frac{284 + \dots + 318}{10} \cong 314,3,$$

$$s_C = \sqrt{\frac{[\sum_{i=1}^n (x_i)^2] - n\bar{x}^2}{n-1}} = \sqrt{\frac{(284^2 + \dots + 318^2) - 10 \times 314,3^2}{10-1}} \cong 40,7 \text{ e}$$

$$CV_C = \frac{s}{\bar{x}} = \frac{40,7}{314,3} = 0,129.$$

Cap. 3 – Descrição e Apresentação de Dados (cont.)

Ex. 7 (cont.):

a) (cont.):

$$\bar{x}_T = \frac{\sum_{i=1}^n x_i}{n} = \frac{166 + \dots + 121}{9} \cong 140,2,$$

$$s_T = \sqrt{\frac{[\sum_{i=1}^n (x_i)^2] - n\bar{x}^2}{n-1}} = \sqrt{\frac{[166^2 + \dots + 121^2] - 10 \times 140,2^2}{9-1}} \cong 38,9 \text{ e}$$

$$CV_T = \frac{s}{\bar{x}} = \frac{38,9}{140,2} = 0,389.$$

Verifica-se, portanto, um CV de 12,9% nos resultados do colesterol e 38,9% no triglicérides, um valor alto.

b) Não foi usada uma expressão apropriada, pois o que se observou foram variações nos resultados das medições. Sem uma medida de referência (exata), não se pode dizer que se tratam de erros.