

Gráficos

Exemplo

Considere novamente os dados sobre a dureza do alumínio.

53,0	70,2	84,3	69,5	77,8	87,5
53,4	82,5	67,3	54,1	70,5	71,4
95,4	51,1	74,4	55,7	63,5	85,8
53,5	64,3	82,7	78,5	55,7	69,1
72,3	59,5	55,3	73,0	52,4	50,7

Fonte: Hoaglin, Mosteller e Tukey, 1983, apud Morettin & Bussab, Estatística Básica.

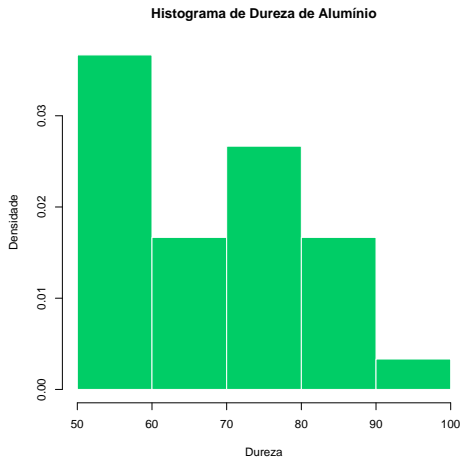
Gráficos

Exemplo

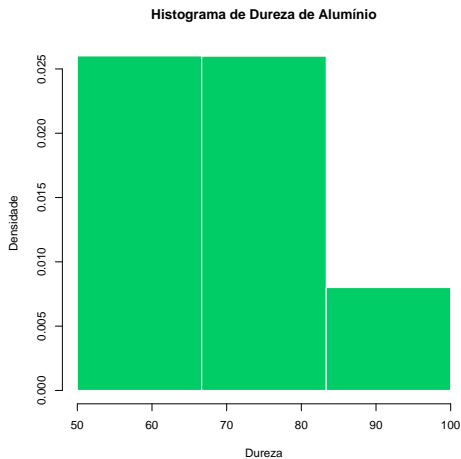
Mostraremos três exemplos de histogramas, além do *boxplot* e do gráfico ramo-e-folhas desses dados. Os histogramas foram gerados com diferentes números de intervalos:

- O primeiro com o padrão do pacote estatístico **R**.
- O segundo com apenas 3 (poucos intervalos).
- O terceiro com 20 (muitos intervalos).

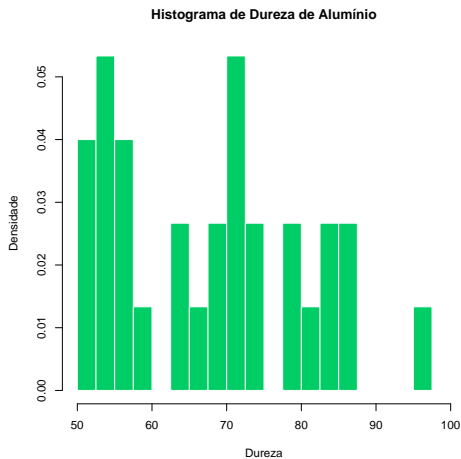
Histograma



Histograma (poucas categorias)



Histograma (muitas categorias)



Box Plot

Box Plot de Dureza de Alumínio

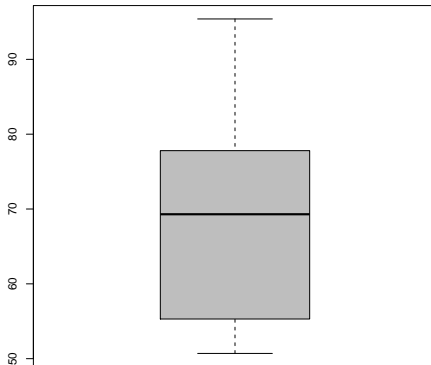
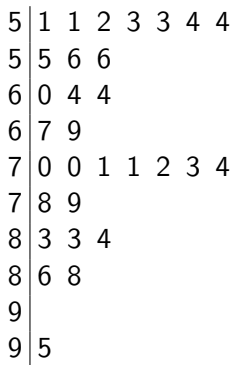


Gráfico de Ramo-e-Folhas

O gráfico de Ramo-e-Folhas foi construído para os valores inteiros (truncados) dos dados:



Construção de um Histograma

Exemplo

A seguinte tabela resume o salário da seção de orçamentos da Companhia MB:

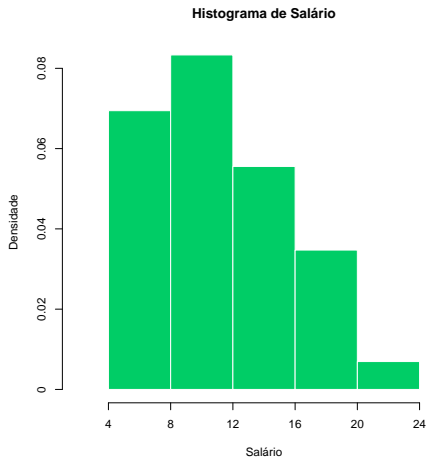
	Ponto Médio	Frequência	Proporção ($100f_i$)
[4,00 – 8,00)	6,00	10	27,78%
[8,00 – 12,00)	10,00	12	33,33%
[12,00 – 16,00)	14,00	8	22,22%
[16,00 – 20,00)	18,00	5	13,89%
[20,00 – 24,00]	22,00	1	2,78%
Total:	–	36	100%

Fonte: Morettin & Bussab, Estatística Básica 6ª edição, pág 18.

Construção de um Histograma

O histograma é um gráfico de barras contíguas, onde as bases são proporcionais aos intervalos de classe, e as alturas são dadas pela frequência relativa. Se um certo intervalo tem amplitude Δ_i , então a altura da barra é dada por f_i/Δ_i , de tal maneira que a área do gráfico seja 1.

Construção de um Histograma



Construção de um Box Plot

Exemplo

Considere a seguinte amostra aleatória de um experimento:

0,5 2,3 8,0 9,8 4,0 15,3 6,4 13,5 12,0

Esses números podem ser ordenados em

0,5 2,3 4,0 6,4 8,0 9,8 12,0 13,5 15,3

Adaptado de: Morettin & Bussab, Estatística Básica 6ª edição.

Construção de um Box Plot

Para construir o Box Plot, devemos determinar algumas estatísticas sobre os dados.

- A *mediana* (ou Q_2) é simplesmente o valor central da amostra ordenada, denotada por $x_{(5)}$ neste caso. Seu valor é de 8,0.
- O *primeiro quartil* é o valor mediano dos dados abaixo da mediana. Ou seja, o valor mediano de

0,5 2,3 4,0 6,4

Temos aí um número par de elementos, então o primeiro quartil é a média entre 2,3 e 4,0, ou seja, 3,15.

Construção de um Box Plot

- O *terceiro quartil* é o valor mediano dos dados acima da mediana. Ou seja, o valor mediano de

9,8 12,0 13,5 15,3

Temos novamente um número par de elementos, então o terceiro quartil é a média entre 12,0 e 13,5, ou seja, 12,75.

- O intervalo interquartilício IQ é simplesmente $Q_3 - Q_1 = 12,75 - 3,15 = 9,6$.

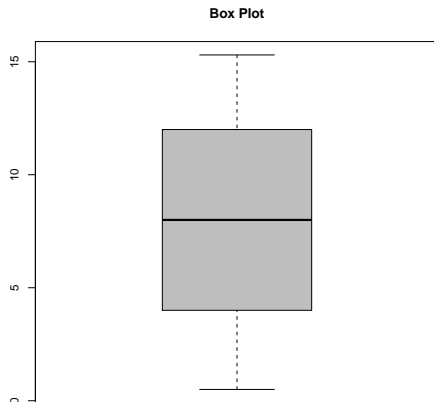
Construção de um Box Plot

A construção do gráfico é imediata: Com os dados no eixo y , o traço horizontal em negrito denota a mediana, a caixa representa a região entre Q_1 e Q_3 , e as linhas pontilhadas denotam o mínimo/máximo dos dados que estiverem na região entre $Q_2 - 1,5I_Q$ e $Q_2 + 1,5I_Q$.

Quaisquer valores fora desse intervalo são marcados com um ponto ou asterisco, e chamados *outliers*.

No nosso caso, como $-1,6 < 0,5 = \min(x)$ e $\max(x) = 15,3 < 17,6$, não temos outliers nos dados.

Construção de um Box Plot

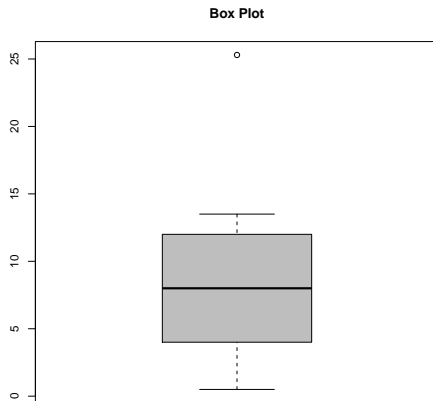


Construção de um Box Plot

Podemos experimentar com esses valores, para observar o efeito de um outlier nos dados (e como o boxplot consegue detectá-los). Outliers comumente são erros de experimentação (mas nem sempre!) e devem ser tratados com cautela.

Suponha que o máximo da amostra tenha sido computado erroneamente, isto é, ao invés de 15,3, computou-se 25,3. As estatísticas Q_1 , Q_2 e Q_3 não se alteram, mas o gráfico resultante exibirá o comportamento patológico dessa observação.

Construção de um Box Plot



Distribuições Bivariadas

Exemplo

Frequentemente os dados serão apresentados em uma tabela, quando lidamos com variáveis discretas, especialmente quando nos interessarem duas ou mais variáveis. Para o caso de duas variáveis X e Y , assumindo valores em $1, 2, \dots, k$ e $1, 2, \dots, r$, respectivamente, temos que a tabela a seguir é a forma mais adequada de resumir estes dados.

Distribuições Bivariadas

		Y				
		1	2	...	r	
X	1	a_{11}	a_{12}	\cdots	a_{1r}	$\sum_{j=1}^r a_{1j}$
	2	a_{21}	a_{22}	\cdots	a_{2r}	$\sum_{j=1}^r a_{2j}$
	\vdots	\vdots		\ddots	\vdots	\vdots
	k	a_{k1}	a_{k2}	\cdots	a_{kr}	$\sum_{j=1}^r a_{kj}$
		$\sum_{i=1}^k a_{i1}$	$\sum_{i=1}^k a_{i2}$	\cdots	$\sum_{i=1}^k a_{ir}$	$\sum_{i=1}^k \sum_{j=1}^r a_{ij}$

Distribuições Bivariadas

Exercício

Considere ainda os dados da companhia Milsa.

- Qual a proporção de indivíduos que tem categoria 1 de Y e categoria 2 de X ?
- Qual a proporção de indivíduos que tem categoria 2 de Y , entre o total?
- Qual a proporção de indivíduos que tem categoria 2 de X , entre o total?
- Entre os elementos que tem a categoria r de Y , que proporção tem a categoria k de X ?

Distribuições Bivariadas

Exemplo

Observe agora os intervalos de classe para salário e a distribuição covariada com a variável procedência. A tabela a seguir mostra a frequência covariada de cada classe:

	Capital	Interior	Outro	Marginal
[4 – 8)	4	3	3	10
[8 – 12)	3	4	6	13
[12 – 16)	1	3	3	7
[16 – 20)	3	1	1	5
[20 – 24]	0	1	0	1
Marginal	11	12	13	36