

118,6 127,4 138,4 130,0 113,7 122,0 108,3 131,5
133,2

- a.** Qual é a mediana dos valores de pressão sanguínea *informados*?
- b.** Suponha que a pressão sanguínea do segundo indivíduo seja 127,6 em vez de 127,4 (uma pequena alteração em um único valor). Como isso afeta a mediana dos valores informados? O que isso diz sobre a sensibilidade da mediana ao arredondamento ou agrupamento dos dados?
- 39.** A propagação de trincas por fadiga em diversas peças de aeronaves tem sido objeto de muitos estudos nos últimos anos. Os dados a seguir consistem dos tempos de propagação (horas de voo/ 10^4) para atingir um determinado tamanho de trinca em furos de fixadores propostos para uso em aeronaves militares (“Statistical Crack Propagation in Fastener Holes under Spectrum Loading,” *J. Aircraft*, 1983, p. 1028-1032):
- 0,736 0,863 0,865 0,913 0,915 0,937 0,983 1,007
1,011 1,064 1,109 1,132 1,140 1,153 1,253 1,394
- a.** Calcule e compare os valores da média e da mediana amostrais.
- b.** Em quanto a maior observação da amostra pode ser diminuída sem afetar o valor da mediana?
- 40.** Calcule a mediana amostral, a média aparada de 25%, a média aparada de 10% e a média amostral para os dados de concentração fornecidos no Exercício 27 e compare essas medidas.
- 41.** Uma amostra de $n = 10$ automóveis foi selecionada e cada um deles foi sujeito a um teste de colisão a 5 mph. Representando um carro sem danos visíveis por S (de sucesso) e um carro com danos por F, os resultados são os seguintes:
- S S F S S S F F S S
- a.** Qual é o valor da proporção amostral de sucessos x/n ?
- b.** Substitua cada S por 1 e cada F por 0. Calcule então \bar{x} para essa amostra codificada numericamente. Como \bar{x} pode ser comparado a x/n ?
- c.** Suponha que se decidiu incluir mais 15 carros no experimento. Quantos deles teriam de ser S para fornecer $x/n = 0,80$ para a amostra de 25 carros?
- 42.** **a.** Se uma constante c é adicionada a cada x_i de uma amostra, resultando em $y_i = x_i + c$, como a média e a mediana amostrais dos y_i se relacionam com a média e a mediana dos x_i ? Verifique suas hipóteses.
- b.** Se cada x_i é multiplicado por uma constante c , resultando em $y_i = cx_i$, responda à questão da parte (a). Verifique novamente suas hipóteses.
- 43.** Um experimento para estudar a vida útil (em horas) de certo tipo de componente consiste em colocar dez componentes em operação e observá-los por 100 horas. Oito dos componentes apresentaram falhas nesse período e esses valores de vida útil foram registrados. Represente os tempos de vida útil dos dois componentes que ainda funcionam após 100 horas por 100+. As observações resultantes foram
- 48 79 100+ 35 92 86 57 100+ 17 29
- Que medidas de tendência central discutidas nesta seção podem ser calculadas e quais são os valores dessas medidas? (Nota: Os dados deste experimento são “censurados pela direita”.)

1.4 | Medidas de Dispersão

Informar apenas a medida de tendência central fornece apenas informações parciais sobre um conjunto de dados ou uma distribuição. Diferentes amostras ou populações podem ter medidas de tendência central idênticas e apresentar diferenças entre si em outros aspectos importantes. A Figura 1.18 apresenta gráficos de pontos de três amostras com a mesma média e a mesma mediana, mas com dispersões diferentes ao redor do centro. A primeira amostra é a que apresenta maior dispersão, a terceira a menor e a segunda é intermediária em relação a elas.

Medidas de dispersão para dados amostrais

A medida de dispersão mais simples de uma amostra é a **amplitude**, a diferença entre o maior e o menor valores da amostra. Observe que o valor da amplitude da amostra 1 da Figura 1.18 é muito maior do que o da amostra 3, o que reflete maior dispersão na primeira amostra do que na terceira. Um defeito da amplitude, entretanto, é que ela depende apenas das duas observações mais extremas e não considera as posições dos $n - 2$ valores restantes. As amostras 1 e 2 na Figura 1.18 possuem amplitudes idênticas mas, se levarmos em conta as observações entre os dois extremos, há muito menos dispersão na segunda amostra do que na primeira.

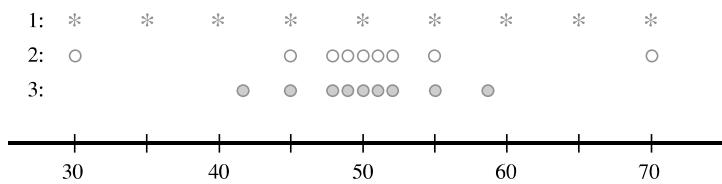


Figura 1.18 Amostras com medidas centrais idênticas mas com variabilidade diferente

Nossa principal medida de dispersão envolve os **desvios em relação à média**, $x_1 - \bar{x}$, $x_2 - \bar{x}$, ..., $x_n - \bar{x}$. Ou seja, os desvios da média são obtidos pela subtração de \bar{x} de cada uma das n observações da amostra. Um desvio será positivo se a observação for maior que a média (à direita da média no eixo das medidas) e negativo se a observação for menor que a média. Se todos os desvios forem pequenos em magnitude, todos os x_i estarão próximos à média e haverá pouca dispersão. Por outro lado, se alguns desvios forem grandes, alguns x_i estarão distantes de \bar{x} , indicando maior dispersão. Uma forma simples de combinar os desvios em uma única quantidade é calcular a sua média (somá-los e dividi-los por n). Infelizmente, há um problema grave com essa sugestão:

$$\text{somatória dos desvios} = \sum_{i=1}^n (x_i - \bar{x}) = 0$$

de forma que o desvio será sempre zero. A demonstração usa diversas regras-padrão de somatória e o fato de que $\sum \bar{x} = \bar{x} + \bar{x} + \dots + \bar{x} = n\bar{x}$:

$$\sum(x_i - \bar{x}) = \sum x_i - \sum \bar{x} = \sum x_i - n\bar{x} = \sum x_i - n\left(\frac{1}{n}\sum x_i\right) = 0$$

Como podemos alterar os desvios para quantidades não-negativas de forma que os desvios positivos e negativos não se cancelam quando forem somados? Uma possibilidade é trabalhar com os valores absolutos e calcular o desvio médio absoluto $\sum |x_i - \bar{x}| / n$. Como a operação em valor absoluto conduz a diversas dificuldades teóricas, considere os quadrados dos desvios $(x_1 - \bar{x})^2$, $(x_2 - \bar{x})^2$, ..., $(x_n - \bar{x})^2$. Em vez de usar o quadrado do desvio médio $\sum(x_i - \bar{x})^2 / n$, por diversos motivos, dividiremos a soma dos quadrados dos desvios por $n - 1$ em vez de n .

DEFINIÇÃO

A **variância amostral**, representada por s^2 , é dada por

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1} = \frac{S_{xx}}{n - 1}$$

O **desvio padrão amostral**, representado por s , é a raiz quadrada (positiva) da variância:

$$s = \sqrt{s^2}$$

A unidade de s é a mesma de cada um dos x_i . Se, por exemplo, as observações forem consumo de combustível em milhas por galão, podemos ter $s = 2,0$ milhas/galão. Uma interpretação grosseira do desvio padrão da amostra é que ele é o tamanho de um desvio típico ou representativo da média amostral para a amostra selecionada. Dessa forma, se $s = 2,0$ milhas/galão, então alguns x_i da amostra estão dentro do intervalo \bar{x} mais de que 2,0, enquanto outros estão mais distantes; 2,0 é um desvio representativo (ou “padrão”) do consumo médio de combustível. Se $s = 3,0$ para uma segunda amostra de carros de outro tipo, um desvio típico dessa amostra é cerca de 1,5 vez maior do que na primeira amostra, indicando maior dispersão da segunda.

Exemplo 1.16

A resistência é uma importante característica de materiais usados em casas pré-fabricadas. Cada um dos $n = 11$ elementos de placas pré-fabricadas foi submetido a um teste de tensão severo e a largura máxima (mm) das trincas resultantes foi registrada. Os dados fornecidos (Tabela 1.3) foram relatados no artigo “Prefabricated Ferrocement Ribbed Elements for Low-Cost Housing” (*J. Ferrocement*, 1984, p. 347-364). Conseqüências do arredondamento influenciam para a soma dos desvios não ser exatamente nula. O numerador de s^2 é 11,9359, portanto $s^2 = 11,9359/(11 - 1) = 11,9359/10 = 1,19359$ e $s = \sqrt{1,19359} = 1,0925$ mm.

Tabela 1.3 Dados do Exemplo 1.16

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
0,684	-0,9841	0,9685
2,540	0,8719	0,7602
0,924	-0,7441	0,5537
3,130	1,4619	2,1372
1,038	-0,6301	0,3970
0,598	-1,0701	1,1451
0,483	-1,1851	1,4045
3,520	1,8519	3,4295
1,285	-0,3831	0,1468
2,650	0,9819	0,9641
1,497	-0,1711	0,0293
$\sum x_i = 18,349$	$\sum (x_i - \bar{x}) = -0,0001$	$S_{xx} = \sum (x_i - \bar{x})^2 = 11,9359$
$\bar{x} = \frac{18,349}{11} = 1,6681$		

Dedução de s^2

Para explicar por que s^2 deve ser usado em vez do quadrado do desvio médio para medir a dispersão, observe primeiro que, enquanto s^2 mede a dispersão da amostra, há uma medida de dispersão da população denominada *variância da população*. Usaremos σ^2 (o quadrado da letra grega minúscula sigma) para representar a variância da população e σ para representar o desvio padrão da população (a raiz quadrada de σ^2). Quando a população é finita e consiste de N valores,

$$\sigma^2 = \sum_{i=1}^N (x_i - \mu)^2 / N$$

que é a média de todos os quadrados dos desvios da média da população (no caso de população, o divisor é N e não $N - 1$). Definições mais gerais de σ^2 são apresentadas nos capítulos 3 e 4.

Da mesma forma que \bar{x} é usado para fazer inferências sobre a média da população μ , devemos definir a variância da amostra de maneira que possa ser usada para fazer inferências sobre σ^2 . Observe que σ^2 envolve os quadrados dos desvios em torno da média da população μ . Se conhecêssemos o valor real de μ , poderíamos então definir a variância da amostra como o quadrado do desvio médio da amostra x_i em torno de μ . Entretanto, o valor de μ quase nunca é conhecido, de modo que deve ser usada a soma dos quadrados dos desvios em torno de \bar{x} . Todavia, os x_i tendem a estar mais próximos de sua média do que da média da população μ , assim, para compensar tal fato, é usado o divisor $n - 1$ em vez de n . Em outras palavras, se usássemos o divisor n na fórmula da variância da amostra, quantidade resultante tenderia a subestimar σ^2 (gerar valores na média muito pequenos para a estimativa), enquanto a divisão pelo valor, ligeiramente menor, $n - 1$ corrige a subestimativa.

É costume se referir a s^2 como tendo $n - 1$ **graus de liberdade** (gl) como base. Essa terminologia resulta do fato de que, apesar de s^2 ter como base as n quantidades $x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}$, sua soma é nula. Assim, especificar os valores de qualquer $n - 1$ das quantidades determina o valor restante. Por exemplo: se $n = 4$ e $x_1 - \bar{x} = 8, x_2 - \bar{x} = -6$ e $x_4 - \bar{x} = -4$, automaticamente $x_3 - \bar{x} = 2$, de forma que apenas três dos quatro valores de $x_i - \bar{x}$ serão determinados livremente (3 gl).

Uma fórmula para o cálculo de s^2

O cálculo dos quadrados dos desvios é tedioso, especialmente se for usada para \bar{x} precisão decimal suficiente para prevenir os efeitos de arredondamento. Uma fórmula alternativa para o numerador de s^2 evita a necessidade de todas as subtrações para obter os desvios. A fórmula envolve $(\sum x_i)^2$, somando e depois obtendo os quadrados e $\sum x_i^2$, obtendo os quadrados e depois somando.

Uma expressão alternativa para o numerador de s^2 é

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

Demonstração Como $\bar{x} = \sum x/n, n\bar{x}^2 = (\sum x_i)^2/n$. Então,

$$\begin{aligned} \sum (x_i - \bar{x})^2 &= \sum (x_i^2 - 2\bar{x} \cdot x_i + \bar{x}^2) = \sum x_i^2 - 2\bar{x} \sum x_i + \sum (\bar{x})^2 \\ &= \sum x_i^2 - 2\bar{x} \cdot n\bar{x} + n(\bar{x})^2 = \sum x_i^2 - n(\bar{x})^2 \end{aligned}$$

Exemplo 1.17

A quantidade de reflexão de luz pelas folhas foi usada para diversos propósitos, incluindo a avaliação da cor da plantação de grama, a estimativa do *status* de nitrogênio e a medida de biomassa. O artigo “Leaf Reflectance-Nitrogen-Chlorophyll Relations in Buffel-Grass” (*Photogrammetric Engr. and Remote Sensing*, 1985, p. 463-466) forneceu as observações a seguir, obtidas usando espectrofotometria, sobre a reflexão das folhas sob condições experimentais especificadas.

Observação	x_i	x_i^2	Observação	x_i	x_i^2
1	15,2	231,04	9	12,7	161,29
2	16,8	282,24	10	15,8	249,64
3	12,6	158,76	11	19,2	368,64
4	13,2	174,24	12	12,7	161,29
5	12,8	163,84	13	15,6	243,36
6	13,8	190,44	14	13,5	182,25
7	16,3	265,69	15	12,9	166,41
8	13,0	169,00			
				$\sum x_i = 216,1$	$\sum x_i^2 = 3168,13$

A fórmula de cálculo fornece

$$\begin{aligned} S_{xx} &= \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 3168,13 - \frac{(216,1)^2}{15} \\ &= 3168,13 - 3113,28 = 54,85 \end{aligned}$$

de onde $s^2 = S_{xx}/(n - 1) = 54,85/14 = 3,92$ e $s = 1,98$. ■

O método alternativo pode resultar em valores de s^2 e s diferentes dos valores calculados, usando as definições. Essas diferenças se devem a efeitos de arredondamento e não serão importantes na maioria das amostras. Para minimizar os efeitos de arredondamento ao usar a fórmula alternativa, os cálculos intermediários devem ser feitos usando-se vários dígitos significativos a mais do que serão mantidos na resposta final. Como o numerador de s^2 é a soma de quantidades não-negativas (quadrados desvios), s^2 é com certeza não-negativo. Ainda assim, se for usada alternativa, particularmente com dados de pouca dispersão, um pequeno erro numérico pode resultar em um numerador negativo ($\sum x_i^2$ menor que $(\sum x_i)^2/n$). Se o valor de s^2 for negativo, foi cometido um erro de cálculo.

Diversas outras propriedades de s^2 podem facilitar esse cálculo.

PROPOSIÇÃO

Considere x_1, x_2, \dots, x_n como uma amostra e c como uma constante qualquer diferente de zero.

1. Se $y_1 = x_1 + c, y_2 = x_2 + c, \dots, y_n = x_n + c$, então $s_y^2 = s_x^2$, e
2. Se $y_1 = cx_1, \dots, y_n = cx_n$, então $s_y^2 = c^2 s_x^2, s_y = |c| s_x$,

onde s_x^2 é a variância amostral dos x e s_y^2 é a variância amostral dos y .

Em palavras, o resultado 1 informa que, se uma constante c for adicionada a (ou subtraída de) cada valor dos dados, a variância não será alterada. Isso é intuitivo, já que a adição ou a subtração de c altera a localização do conjunto de dados, mas mantém as distâncias entre os valores inalteradas. De acordo com o resultado 2, a multiplicação de cada x_i por c resulta em s^2 , sendo multiplicado por um fator de c^2 . Essas propriedades podem ser demonstradas pela observação de que $\bar{y} = \bar{x} + c$ no resultado 1 e de que $\bar{y} = c\bar{x}$ no resultado 2.

Boxplots

Diagramas de caule e folha e histogramas conduzem a impressões gerais sobre um conjunto de dados, enquanto um único valor como a média ou o desvio padrão enfoca apenas um aspecto dos dados. Nos últimos anos, um resumo esquemático denominado *boxplot* vem sendo usado para descrever as características mais proeminentes de conjuntos de dados. Essas características incluem (1) centro, (2) dispersão, (3) a extensão e a natureza de qualquer desvio em relação à simetria e (4) a identificação de *outliers*, observações que normalmente estão distantes da maior parte dos dados. Como apenas um *outlier* pode afetar drasticamente os valores de \bar{x} e s , um *boxplot* é baseado em medidas “resistentes” à presença de alguns *outliers*: a mediana e uma medida de dispersão denominada *dispersão entre os quartos*.

DEFINIÇÃO

Ordene as n observações da menor para a maior e então separe a metade menor da maior. A mediana \tilde{x} estará incluída em ambas as partes se n for ímpar. Então o **quarto inferior** será a mediana da metade menor e o **quarto superior** será a mediana da metade maior. Uma medida de dispersão resistente a *outliers* é a **dispersão entre os quartos** f_s , dada por

$$f_s = \text{quarto superior} - \text{quarto inferior}$$

Grosso modo, a dispersão entre os quartos não será alterada pelas posições das observações nos menores 25% ou nos maiores 25% dos dados.

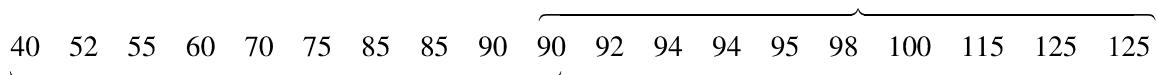
O *boxplot* mais simples tem base no seguinte resumo de cinco números:

menor x_i	quarto inferior	mediana	quarto superior	maior x_i
-------------	-----------------	---------	-----------------	-------------

Primeiro, desenhe um eixo de medida horizontal. Então, coloque um retângulo sobre o eixo; a extremidade esquerda do retângulo estará no quarto inferior e a extremidade direita estará no quarto superior (de forma que a largura da caixa = f_s). Trace um segmento de reta vertical ou outro símbolo dentro do retângulo na posição da mediana. A posição do símbolo da mediana em relação às duas extremidades indica informações sobre o desvio nos 50% centrais dos dados. Por fim, desenhe “bigodes” saindo de cada extremidade do retângulo para as observações maior e menor. Um *boxplot* com uma orientação vertical também pode ser desenhado, fazendo-se modificações óbvias no processo de construção.

Exemplo 1.18

O ultra-som foi usado para obter informações sobre dados de corrosão na espessura da chapa do assoalho de um reservatório elevado usado para armazenar óleo bruto (“Statistical Analysis of UT Corrosion Data from Floor Plates of a Crude Oil Aboveground Storage Tank,” *Materials Eval.*, 1994, p. 846-849). Cada observação é a maior profundidade do orifício na placa, expressa em milipolegadas.



O resumo de cinco números segue:

$$\begin{array}{llll} \text{menor } x_i = 40 & \text{quarto inferior} = 72,5 & \tilde{x} = 90 & \text{quarto superior} = 96,5 \\ \text{maior } x_i = 125 & & & \end{array}$$

A Figura 1.19 exibe o *boxplot* resultante. A extremidade direita da caixa está muito mais próxima da mediana do que a esquerda, indicando um desvio substancial na parte central dos dados. A largura da caixa (f_s) também é relativamente grande em relação à amplitude dos dados (distância entre as pontas dos bigodes).

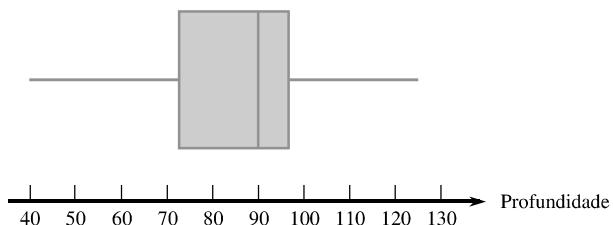


Figura 1.19 Um *boxplot* dos dados de corrosão

A Figura 1.20 mostra o resultado do MINITAB de uma solicitação de descrição dos dados de corrosão. A média aparada é a média das 17 observações que permanecem após a exclusão do maior e do menor valores (porcentagem de truncamento $\approx 5\%$). Q1 e Q3 são os quartis inferior e superior. Eles são similares aos quartos, mas calculados de uma forma ligeiramente diferente. A Média SE é s/\sqrt{n} , quantidade que será importante no trabalho subsequente em inferências de μ .

Variável	N	Média	Mediana	Média Ap	Desv Padrão	Média SE
profundidade	19	86,32	90,00	86,76	23,32	5,35
profundidade	Mínimo	Máximo	Q1	Q3		
	40,00	125,00	70,00	98,00		

Figura 1.20 Descrição em MINITAB dos dados de profundidade do orifício

Boxplots com Outliers

Um *boxplot* pode ser formatado para indicar explicitamente a presença de *outliers*.

DEFINIÇÃO

Qualquer observação distante mais do que $1,5f_s$ do quarto mais próximo é um *outlier*. Um *outlier* é **extremo** se estiver a mais de $3f_s$ do quarto mais próximo. Caso contrário, é **moderado**.

Muitos procedimentos inferenciais se baseiam na suposição de que a amostra é proveniente de uma distribuição normal. Mesmo um único *outlier* extremo na amostra avisa o investigador de que tais procedimentos não devem ser usados e a presença de diversos *outliers* moderados passa a mesma mensagem.

Modifiquemos agora nossa construção anterior de um *boxplot*, desenhando um “bigode” em cada extremidade da “caixa”, para que a menor e a maior observações *não* sejam *outliers*. Cada *outlier* moderado é representado por um círculo cheio e cada *outlier* extremo é representado por um círculo vazio. Alguns softwares estatísticos não fazem distinções entre *outliers* extremos e moderados.

Exemplo 1.19

Os efeitos de descargas parciais na degradação de materiais de cavidades isolante têm importantes implicações na vida útil de componentes de alta voltagem. Consideremos a seguinte amostra de $n = 25$ larguras de pulso de descargas lentas em uma cavidade cilíndrica de polietileno. (Esses dados são consistentes com um histograma de 250 observações no artigo “Assessment of Dielectric Degradation by Ultrawide-band PD Detection,” *IEEE Trans. on Dielectrics and Elec. Insul.*, 1995, p. 744-760.) O autor do artigo nota o impacto de diversas ferramentas estatísticas na interpretação dos dados de descarga.

5,3	8,2	13,8	74,1	85,3	88,0	90,2	91,5	92,4	92,9	93,6	94,3	94,8
94,9	95,5	95,8	95,9	96,6	96,7	98,1	99,0	101,4	103,7	106,0	113,5	

Os indicadores relevantes são

$$\begin{array}{lll} \bar{x} = 94,8 & \text{quarto inferior} = 90,2 & \text{quarto superior} = 96,7 \\ f_s = 6,5 & 1,5f_s = 9,75 & 3f_s = 19,50 \end{array}$$

Dessa forma, qualquer observação menor que $90,2 - 9,75 = 80,45$ ou maior que $96,7 + 9,75 = 106,45$ é um *outlier*. Há um *outlier* na extremidade superior da amostra e quatro *outliers* na inferior. Como $90,2 - 19,5 = 70,7$, as três observações: 5,3, 8,2 e 13,8 são *outliers* extremos. Os outros dois *outliers* são moderados. Os “bigodes” se estendem até 85,3 e 106,0, as observações mais extremas que não são *outliers*. O *boxplot* resultante está na Figura 1.21. Há desvio negativo grande na parte central da amostra, assim como na amostra inteira.

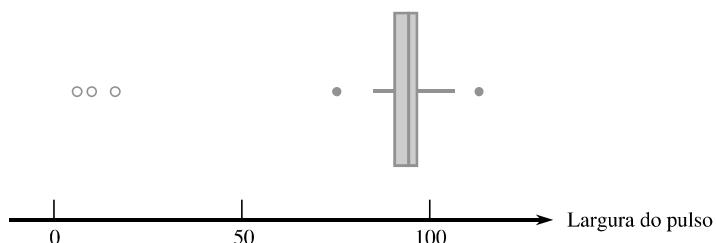


Figura 1.21 Um *boxplot* dos dados de largura de pulso que exibem *outliers* extremos e moderados ■

Boxplots Comparativos

Um *boxplot* comparativo ou lado a lado é uma forma muito eficiente de revelar semelhanças e diferenças entre dois ou mais conjuntos de dados consistindo de observações da mesma variável.

Exemplo 1.20

Nos últimos anos, algumas evidências sugerem que a alta concentração de radônio em ambientes fechados pode estar ligada ao desenvolvimento de cânceres infantis, mas muitos profissionais da saúde não estão convencidos. Um artigo recente (“Indoor Radon and Childhood Cancer”, *The Lancet*, 1991, p. 1537-1538) apresentou os seguintes dados sobre a concentração de radônio (Bq/m^3) em duas amostras diferentes de casas. A primeira amostra consistia de casas em que havia residido uma criança com diagnóstico de câncer. As casas da segunda amostra não possuíam casos registrados de câncer infantil. A Figura 1.22 apresenta um diagrama de caule e folha dos dedos.

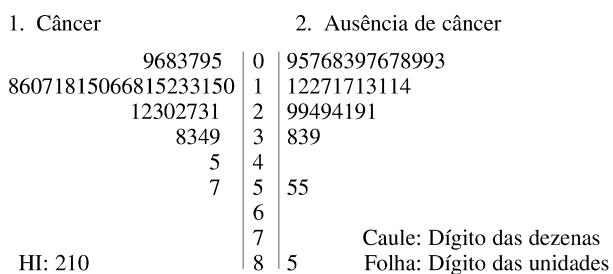


Figura 1.22 Diagrama de caule e folha do Exemplo 1.20

As quantidades numéricas resumidas são as seguintes:

	\bar{x}	\tilde{x}	s	f_s
Câncer	22,8	16,0	31,7	11,0
Sem câncer	19,2	12,0	17,0	18,0

Os valores da média e da mediana sugerem que a amostra de câncer está centrada um pouco à direita da amostra de ausência de câncer na escala de medida. A média, entretanto, exagera a grandeza do desvio, principalmente devido ao valor 210 na amostra de câncer. Os valores de s sugerem dispersão maior na amostra de câncer do que na amostra de ausência de câncer, mas essa impressão é contrariada pela quarta dispersão. Novamente, a observação 210, um *outlier* extremo, é a culpada. A Figura 1.23 mostra um *boxplot* comparativo do pacote de software S-Plus. A caixa dos indivíduos com ausência de câncer é mais extensa se comparada à caixa de câncer ($f_s = 18$ vs. $f_s = 11$), além disso, as posições das retas das medianas mostram desvio maior na metade da amostra de ausência de câncer do que na amostra de câncer. Os *outliers* são representados por segmentos de reta horizontais e não há distinção entre *outliers* moderados e extremos.

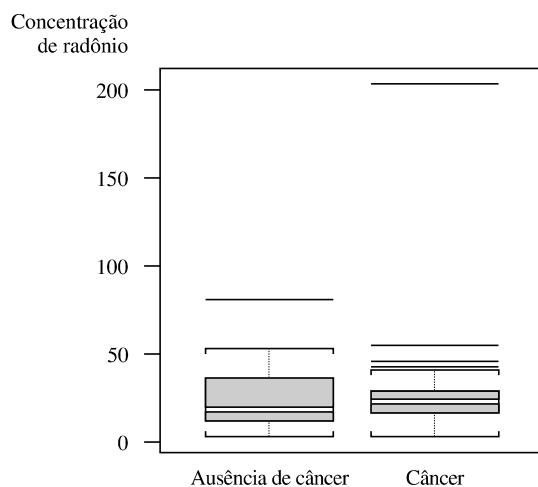


Figura 1.23 Um *boxplot* dos dados do Exemplo 1.20, no S-Plus