## 2.1.2. Diagrama de ramo-e-folhas (*stem-and-leaf*)



---

### 2.1.2.1. Introdução

O *diagrama de ramo-e-folhas* foi desenvolvido nos anos 1970s pelo estatístico John Tukey. É um gráfico que faz um *agrupamento* simples dos dados, mas tem a vantagem de permitir que os valores originais sejam recuperados (isto é, se tivermos apenas o gráfico, podemos a partir dele recuperar os valores dos dados que foram usados para fazê-lo). O *histograma* (Seção **2.1.5**) também serve para agrupar os dados, mas não permite que os valores originais sejam recuperados depois. O nome *stem-and-leaf* (literalmente, *caule-e-folhas*) foi dado porque se supõe que o gráfico se pareça com um caule vertical, a partir do qual saem as folhas (números) para os dois lados (por exemplo, na Fig. **3**).

### 2.1.2.2. Como fazer o diagrama de ramo-e-folhas básico

No gráfico de ramo-e-folhas, os valores das observações são representados por meio de seus próprios algarismos. Por exemplo, suponha que os dados abaixo representem a distribuição de idades numa amostra de pacientes de uma clínica:

**Amostra A:** 13 17 21 22 24 25 26 29 32 32 34 37 40 40 46 52 73

O eixo do gráfico, colocado na posição vertical, marca os algarismos das dezenas, da seguinte forma:

```
1|
2|
3|
4|
5|
6|
7|
```

A seguir, cada observação é representada por seu algarismo das unidades, colocado à direita de uma traço vertical, na linha correspondente à dezena. A idade do paciente mais jovem (13 anos) é representada na primeira linha do gráfico como: **1|3**. A idade seguinte (17 anos) é representada na mesma linha; o algarismo das unidades (7) é acrescentado à direita do valor já existente. A primeira linha fica portanto como: **1|37**. Se acrescentamos os valores restantes, obtemos o gráfico da Fig. **1**:

```
1|37
2|124569
3|2247
4|006
5|2
6|
7|3
```

**Figura 1. Idades dos pacientes, amostra A**

Portanto, ao invés de empilharmos pontos sobre um eixo horizontal (como no gráfico de pontos), acrescentamos algarismos à direita do eixo vertical, numa mesma linha, para representar as ocorrências de valores que pertencem a uma mesma dezena.

Note, em primeiro lugar, que neste tipo de gráfico não nos interessa muito verificar a freqüência com que cada valor individual ocorre (por exemplo, não nos interessa contar quantos vezes o valor "13" se repete entre os dados), mas sim a freqüência com que valores de cada dezena ocorrem. No exemplo, podemos ver que nesta amostra as idades na dezena de 20 a 29 anos são as mais freqüentes; as da dezena 60 a 69 anos não foram observadas nenhuma vez. Em termos estatísticos, dizemos que estes dados estão "agrupados", ou "classificados" em dezenas; o diagrama de ramo-e-folhas é, por isso, um parente do *histograma* (seção **2.1.5**).

Em segundo lugar, note que a maneira de descrever a forma deste gráfico é exatamente igual a do *gráfico de pontos* visto antes, apesar de os dados terem sido agrupados e de o eixo estar na posição vertical. Girando o gráfico de 90° e observando seu contorno, podemos dizer que a distribuição é unimodal, com leve assimetria positiva e com um ponto discrepante à direita (Fig. **2**).

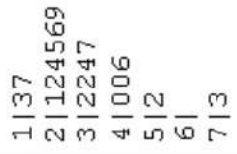


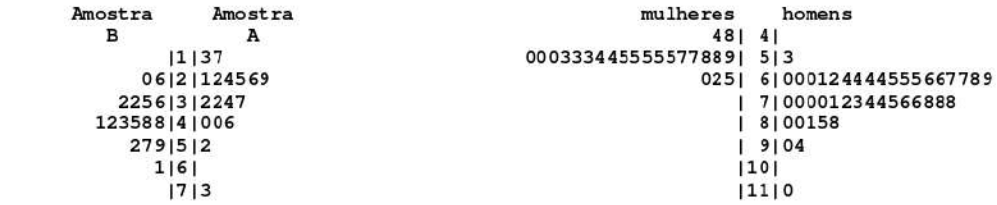

**Figura 2. O mesmo diagrama da Fig. 1 com o eixo na horizontal**

### 2.1.2.3. Diagrama de ramo-e-folhas duplo (*back-to-back*)

O ramo-e-folhas pode ser usado para comparar as distribuições de duas amostras. Suponha por exemplo que tenhamos uma outra amostra de pacientes, com idades:

**Amostra B:** 20 26 32 32 35 36 41 42 43 45 48 48 52 57 59 61

Se desejarmos comparar suas idades com as da Amostra A, podemos colocar os dados das duas amostras em lados opostos do "caule" vertical. O gráfico resultante seria o da Fig. **3A**. Note que a os pacientes da amostra B tendem a ser mais velhos que os da amostra A. Além disso, a distribuição das idades na amostra B é mais próxima da simetria, e não há valores discrepantes.




A. Idades dos pacientes em duas amostras

B. Pesos dos estudantes em uma turma

**Figura 3. Diagrama de ramo-e-folhas duplo**

Como outro exemplo, comparamos na Fig. **3B** os pesos dos alunos e alunas de uma turma de estudantes de Medicina. Como era de se esperar, os alunos são em geral mais pesados do que as alunas; a maioria dos alunos estão na faixa dos 60–70 kg, a maioria da alunas na faixa dos 50 kg. As duas distribuições são unimodais e razoavelmente simétricas, mas a dos alunos tem um valor discrepante, igual a 110 kg.

### 2.1.2.4. Diagramas com dezenas ou centenas subdivididas

Às vezes, precisamos de subdividir as linhas do gráfico para evitar que fiquem sobrecarregadas, ou para ampliar a escala vertical. Se os dados se referem à altura de pessoas de uma amostra, por exemplo, é bem provável que todas estejam entre 100 e 199 cm; se não subdividirmos este intervalo, o gráfico terá apenas uma linha. Outro exemplo: suponha que as idades de uma terceira amostra de pacientes sejam:

**Amostra C:** 33 37 37 38 39 40 40 41 42 44 44 46 48 48 48 49 52 52 52 54 55 58 58

As idades variam entre 33 e 58. Se as representarmos num gráfico com uma dezena por linha, obteremos o gráfico da Fig. **4**, com apenas 3 linhas, o que é muito pouco:

```
3|37789
4|00124468889
5|222458
```

**Figura 4 – Gráfico com número insuficiente de linhas**

Um gráfico mais interessante pode ser conseguido se representarmos cada dezena em duas linhas: por exemplo, a dezena 30-39 será subdividida de forma que os valores entre 30 e 34 fiquem em uma linha, e os entre 35 e 39 fiquem na linha seguinte. No eixo vertical, a primeira linha será marcada pelo algarismo 3; a segunda, apenas por um traço. O gráfico resultante (Fig. **5**) terá o dobro do número de linhas do gráfico original (Fig. **4**). A ampliação da escala vertical é útil neste caso para deixar mais evidente a forma da distribuição.

```
3|3
-|7789
4|001244
-|68889
5|2224
-|588          legenda: 5|2 representa 52 anos
```

**Figura 5. Diagrama de ramo-e-folhas**
**Amostra C - escala vertical ampliada**

Às vezes, é preciso subdividir ainda mais as linhas e ampliar a escala vertical. A única alternativa é subdividir cada linha em cinco, de forma que cada uma represente apenas dois algarismos diferentes. A linha que representa a dezena 10-19, por exemplo, pode ser subdividida na forma:

```
1| 0 ou 1
1| 2 ou 3
1| 4 ou 5
1| 6 ou 7
1| 8 ou 9
```

Por exemplo, suponha que as observações tenham valores expressos por três algarismos, como neste exemplo abaixo (níveis de glicose no sangue em mg%, numa amostra de 100 homens adultos com doença cardíaca).

```
 58, 73, 76, 78, 84, 87, 88, 91, 91, 95, 97, 98, 98,100,
104,105,106,106,107,109,109,110,111,113,116,117,118,119,
...
231,231,237,237,257,265,298,442
```

Neste caso, é melhor *simplificar* estes dados, arredondando os valores até as dezenas mais próximas. Os valores passarão a ser:

```
 60, 70, 80, 80, 80, 90, 90, 90, 90,100,100,100,100,100,
100,110,110,110,110,110,110,110,110,110,120,120,120,120,
...
230,230,240,240,260,270,300,440
```

No eixo vertical serão marcadas as centenas, e os algarismos das dezenas constituirão as "folhas"; os algarismos das unidades serão desprezados. Cada centena será subdividida em cinco linhas. Por exemplo, o valor 227 seria arredondado para 230, e representado como 2|3. O valor 248 será arredondado para 250, e marcado com 2|5, numa linha diferente. O resultado está mostrado na Fig. **6A**; a legenda ao pé do gráfico indica que os dados foram arredondados. Se *truncarmos* os dados (veja seção **2.5.2**), ao invés de arredondá-los, o gráfico resultante será o da Fig. **6B**.

```
0 | 67
0 | 88899999
1 | 00000011111111
1 | 222222223333333333
1 | 444444444445555555
1 | 66666666777
1 | 8888888999999
2 | 00011
2 | 22233
2 | 44
2 | 66
2 |
3 | 0
3 |
3 |
3 |
3 |
4 |
4 |
4 | 4

legenda: 2|1 representa 206-215 mg%
```

A. Valores arredondados

```
0 | 5
0 | 777
0 | 888999999
1 | 000000001111111
1 | 22222222233333333333
1 | 44444444445555
1 | 6666666677777
1 | 88888899999
2 | 011111
2 | 3333
2 | 5
2 | 6
2 | 9
3 |
3 |
3 |
3 |
3 |
4 |
4 |
4 | 4

legenda: 2|1 representa 210-219 mg%
```

B. Valores truncados

**Figura 6. Nível de glicose de 100 homens cardíacos; dados arredondados × truncados**

Em cada um destes gráficos (Figuras **6A** e **B**) é acrescentada uma legenda, e através dela o leitor poderá deduzir se os dados foram arredondados ou truncados, e quais foram as unidades usadas. Note que não há muita diferença entre arredondar ou truncar os dados; os gráficos resultantes são praticamente idênticos. O truncamento contudo é mais fácil, se os gráficos estão sendo feitos à mão, sem um computador.

Se por alguma razão preferirmos não arredondar ou truncar os dados, e sim mantê-los com seus valores originais, será ainda possível representá-los num gráfico de ramo-e-folhas. Há duas maneiras de fazer isto. A primeira, é a de colocar no "ramo" o algarismo

das centenas e o das dezenas, e nas "folhas" o algarismo das unidades. O resultado seria o da Fig. 7. Esta forma pode parecer um pouco confusa, porque há duas dezenas em cada linha, e isto não está indicado na legenda. Assim, por exemplo, na quarta linha, marcada

```
10 | 045667990136789
```

estão os valores de 100 a 109, e em seguida, os de 110 a 119. Poderíamos também fazer uma escala onde cada dezena ocupa uma linha, mas o gráfico então ficaria longo demais, pois teria mais de 40 linhas.

```
 4 | 8
 6 | 368
 8 | 478115788
10 | 045667990136789
12 | 023467788001246788 99
14 | 13455678890368
16 | 0111347936677
18 | 45568834488
20 | 855888
22 | 1177
24 | 7
26 | 5
28 | 8
30 |
32 |
34 |
36 |
38 |
40 |
42 |
44 | 2                  legenda: 44|2 representa 442 mg%
```

**Figura 7. Nível de glicose de 100 homens cardíacos**

A segunda maneira de representar estes dados, sem arredondá-los ou truncá-los, é fazer com que cada folha contenha 2 algarismos (o das dezenas e o das unidades), e seja separada das outras folhas por vírgulas, como na Fig. 8. A desvantagem desta representação é tornar o gráfico mais complicado e gastar mais espaço, se comparada aos das Figs. 6 e 7. Os detalhes trazidos pelo uso de dois algarismos em cada folha na maior parte das vezes não são muito importantes, e em geral não justificam a complicação adicional do gráfico.

Na maioria dos exemplos de diagramas de ramo-e-folhas que vimos até agora, as distribuições eram aproximadamente simétricas, ou tinham leve assimetria positiva, com pontos discrepantes nos valores mais altos da escala. Este tipo de distribuição é o mais comum em variáveis naturais, como a altura de pessoas ou o nível de glicose. Estas variáveis são limitadas à esquerda (os valores não podem ser iguais ou menores que zero), mas não à direita; se houver grande variação entre os valores, eles tenderão a se dispersar à direita do gráfico, não à esquerda. Na Fig. 9 há mais dois exemplos: o gráfico **A** mostra a distribuição do nível de colesterol no sangue da mesma amostra de 100 homens adultos; a distribuição é praticamente simétrica, exceto por um valor muito (estes dados foram usados antes na Fig. 8, seção 2.1.1.5). O gráfico **B** mostra os pesos de uma amostra de 189 mulheres; neste gráfico não há valores discrepantes, mas a assimetria positiva é bem evidente (o *sobrepeso*, aliás, é considerado hoje o problema de saúde número 1 nos EUA).

```
0 | 58
0 | 73,76,78
0 | 84,87,88,91,91,95,97,98,98
1 | 00,04,05,06,06,07,09,09,10,11,13,16,17,18,19
1 | 20,22,23,24,26,27,27,28,28,30,30,31,32,34,36,37,38,38,39,39
1 | 41,43,44,45,45,46,47,48,48,49,50,53,56,58
1 | 60,61,61,61,63,64,67,69,73,76,76,77,77
1 | 84,85,85,86,88,88,93,94,94,98,98
2 | 08,15,15,18,18,18
2 | 31,31,37,37
2 | 57
2 | 65
2 | 98
3 |
3 |
3 |
3 |
3 |
4 |
4 |
4 | 42                          legenda: 4|42 representa 442 mg%
```

**Figura 8. Nível de glicose de 100 homens cardíacos**

O gráfico **C** da Fig. **9**, em contraste, mostra uma distribuição que tem clara assimetria negativa: a das expectativas de vida das mulheres nos 193 países do mundo. À medida que um país se desenvolve, cresce a expectativa de vida dos habitantes (considerada um indicador da saúde da população); contudo, este valor não pode crescer indefinidamente, pois existem barreiras impostas pela biologia. Neste gráfico, a maior expectativa registrada é a do Japão (86 anos); a do Brasil é de 78 anos.

```
13 |4
14 |7
15 |7
16 |12456
17 |1366899
18 |011345666799
19 |0024568999
20 |134569
21 |011235667788999
22 |1133458
23 |001112448899
24 |000368
25 |15569
26 |178
27 |29
28 |679
29 |068
30 |
31 |
32 |
33 |
34 |
35 |
36 |
37 |
38 |2
```

```
 3 |699
 4 |0111123333333344
 4 |55555666777888888888999
 5 |000000000011111112222223333344444444444444444444444
 5 |55555566666777888899999999999999
 6 |0000011111113233333444
 6 |777888889
 7 |000002222
 7 |5667777779
 8 |2344
 8 |55666
 9 |12
 9 |7
10 |4
10 |69
11 |3     legenda: 11|3 representa 113 kg
```

```
46 |x
48 |
50 |xxxx
52 |xxxxxx
54 |xxxxx
56 |xxxxx
58 |xxxxx
60 |xxxxxxxxxx
62 |xxx
64 |xxxxxxxxx
66 |xxxxxxxx
68 |xxxxxxx
70 |xxxxxxx
72 |xxxxxxxxx
74 |xxxxxxxxxxxxx
76 |xxxxxxxxxxxxxxxxxxxxxxxxxxxx
78 |xxxxxxxxxxxxxxxxxxxxxxxxx
80 |xxxxxxxxxxxxxxxx
82 |xxxxxxxxxxxxxx
84 |xxxxxxxxxxxxxxxxx
86 |x
```

A B C

A. Nível de colesterol no sangue de cem pacientes cardíacos, EUA
B. Pesos de 189 mulheres adultas (EUA, 1985)
C. Expectativa de vida das mulheres em 193 paises (Org. Mundial da Saúde, 2011)

**Figura 9.** Exemplos de diagrama de ramo-e-folhas

A assimetria nas distribuições de variáveis naturais quase nunca é extrema. Assimetrias exageradas e pontos discrepantes muito afastados do aglomerado são, por outro lado, comuns nas variáveis sociais e econômicas. O gráfico da Fig. **10**, por exemplo, mostra as rendas percapita de 194 países (dados do Banco Mundial, 2018, disponíveis na *Wikipedia*). Esta renda (o produto interno bruto anual em dólares, dividido pelo tamanho da população) é um indicador da riqueza de um país. Sua distribuição tem uma assimetria

positiva muito mais evidente do que a de qualquer dos gráficos vistos anteriormente, o que evidencia a desigualdade econômica que existe entre os países do mundo.

```
 0 | 00000000111111111111111111111111111122222222222222222223333333333333444444444444444444
 0 | 55555555566666666667777778888899999
 1 | 00001122223444
 1 | 5555666677789
 2 | 0012224
 2 | 569
 3 | 01144
 3 | 5
 4 | 111112
 4 | 567789
 5 | 013
 5 | 59
 6 | 0113
 6 | 68
 7 | 1
 7 | 9
 8 | 14
 8 |
 9 |
 9 |
10 |
10 |
11 |
11 | 6
```

**Figura 10. Renda percapita de 193 países em 2018 ($ 1000) (dados: Banco Mundial)**

Nos dados, o país de maior renda é Liechtenstein, que tem 38 mil habitantes ($ 116 mil), seguido pela Suíça ($ 84 mil). A renda dos EUA é $ 64 mil; os principais países europeus, Alemanha, Reino Unido e França, estão na faixa $ 40-50 mil. A renda percapita de Portugal é de $ 22 mil, a do Brasil e da China em torno de $ 10 mil.

---

**Resumo**

- O diagrama de ramo-e-folhas é provavelmente a melhor ferramenta para analisar graficamente a forma da distribuição de uma amostra (localizar aglomerados, modas, pontos discrepantes, etc.), se a amostra não for muito grande. Se amostra for grande (mais de 100-200 dados), pode ser melhor usar um *histograma*.
- Pode ser usado para comparar duas distribuição (gráfico *back-to-back*).
- Pode se necessário simplificar os dados (*arredondá-los* ou *truncá-los*); deve haver uma legenda explicando o que foi feito.
- Pode ser feito pelo R ou outros pacotes estatísticos.
- Desvantagem: a maior parte do público leigo não conhece este diagrama, e provavelmente não vai entender o que ele significa.