2.5.2. Arredondamento e truncamento de dados

(i) Por que arredondar ou truncar

Tanto o arredondamento quanto o truncamento são técnicas para "simplificar" os número, reduzindo a quantidade de decimais ou de algarismos significativos. Esta redução pode ser necessária por várias razões; a mais comum é, provavelmente, descartar o excesso os decimais desnecessários nos resultados produzidos por calculadoras. Se por exemplo os alunos encontram na prova uma questão como "Calcule o lado de um quadrado cuja área seja de 2 m^2", sabe que a resposta é $\sqrt{2}$, o uma calculadora poderá dar como:

$$\sqrt{2} = 1{,}41421356237$$

A maioria dos alunos porém desconfia, com razão, que todos estes decimais são desnecessários, e responde simplesmente 1,41 ou 1,414 – mesmo que não tenham estudado conceitos como *precisão* de uma medida (veja seção 2.5.3), *algarismos significativos*, etc.

Algumas razões técnicas para arredondar os números são:

a) Para ajustar a quantidade de algarismos significativos de um resultado

Para ajustar a quantidade de algarismos significativos no resultado de uma sequência de cálculos (as regras para este ajuste estão na seção 2.5.1), geralmente é preciso simplificar este resultado, eliminando alguns de seus algarismos. Por exemplo, se somamos os quatro valores abaixo, obtidos a partir de medições de distâncias:

$$13{,}81 + 0{,}28 + 3{,}9 + 4{,}168 = 22{,}158 \approx 22{,}2$$

O resultado do cálculo é dado na calculadora com três decimais; como porém uma das parcelas (=3,9) tem apenas uma casa decimal, o resultado final deve ser arredondado, e apenas uma decimal deve ser mantida.

b) Para indicar a precisão de uma medida, quando há mudança de unidade

Quando mudamos a unidade de uma medida, é preciso tomar cuidado para evitar que a quantidade de decimais usada pareça indicar uma precisão na medida maior do que a que existe na realidade. Este problema ocorre com frequência na tradução de textos do inglês (do sistema *imperial* para o sistema *métrico*). Por exemplo, suponha que a altura de um homem tenha sido dada numa publicação americana como sendo 5 pés e 9 polegadas. Para converter este valor para o sistema métrico, temos (1 pé = 30,48 m; 1 polegada = 2,54 cm):

$$(5 \times 30{,}48) + (9 \times 2{,}54) = 175{,}26 \text{ cm}$$

A medida original porém não tem precisão de centésimos de centímetro, como sugere este resultado 175,26 cm. A menor unidade usada foi a polegada (2,54 cm), portanto a margem de erro a ser considerada é de meia polegada (± 1,27 cm). O mais correto seria traduzir a altura como

$$5 \text{ ft } 9 \text{ in} = (175{,}26 \pm 1{,}27) \text{ cm}$$

o que dá alguma coisa entre 174 cm e 176,5 cm. Na prática, a maioria das publicações não-técnicas (como jornais e revistas) não aceitaria mostrar o resultado desta forma; o melhor então é arredondar o número e dar o resultado simplesmente como "175 cm".

c) Para permitir a representação dos números em gráficos

Quando vamos representar graficamente os dados de uma amostra, com frequência é preciso antes reduzir o número de decimais destes dados. Por exemplo, suponha que seja preciso representar num diagrama de ramo-e-folhas os dados desta amostra:

```
2,981   3,284   3,993   4,022   4,158   4,449   4,968
5,336   5,450   5,494   5,606   5,610   5,765   6,040
6,100   6,102   6,167   6,259   6,385   6,405   6,625
```

Não é possível representar no gráfico todas as três casas decimais de cada número; o que podemos fazer arredondar os números, reduzindo-os a apenas uma casa decimal:

```
3,0     3,3     4,0     4,0     4,2     4,4     5,0
5,3     5,5     5,5     5,6     5,6     5,8     6,0
6,1     6,1     6,2     6,3     6,4     6,4     6,6
```

Estes valores podem agora ser representados num diagrama de ramo-e-folhas (Fig. 1).

```
1 |
2 |
3 | 03
4 | 0024
5 | 0355668
6 | 01123446
```

Figura 1. Diagrama de ramo-e-folhas

(ii) A técnica de arredondamento mais usada

Existem várias técnicas diferentes para arredondamento de números; uma comparação entre elas pode ser encontrada na Wikipédia em inglês [1] (a versão em português porém não é tão completa).

É mais fácil explicar as técnicas se usarmos como exemplo o arredondamento de números que têm apenas uma casa decimal. Neste caso, a técnica mais usada no Brasil é a que leva o número a ser arredondado até o valor inteiro mais próximo. Por exemplo, suponha que desejamos arredondar o número $X = 2{,}6$ para um valor inteiro (eliminando a decimal). Como 2,6 está mais próximo de 3 do que de 2, X será arredondado para cima, para $X \approx 3$. O número $X = 2{,}3$, por outro lado, será arredondado para baixo, para $X \approx 2$. Se X for igual a 2,5, estará à igual distância de 2 ou de 3; nestes casos, a regra mais comumente usada (recomendada pelo IEEE e também pela ABNT NBR 5891) é a implementada em R na função `round`: arredondar para cima o algarismo dos inteiros se ele for ímpar, ou mantê-lo inalterado se for par. O número $X = 2{,}5$, portanto, será arredondado para $X \approx 2$.

O mesmo procedimento pode ser aplicado quando desejamos arredondar não para um número inteiro, mas sim para um número com uma quantidade de decimais pré-determinada. Suponha por exemplo que os números da amostra no item (c) acima devam ser arredondados até a segunda casa decimal (centésimo). Para isto, descartamos o valor que está na terceira casa decimal (milésimo); se ele for maior que 5, o algarismo do centésimo será aumentado de uma unidade. Por exemplo:

```
4,158 → 4,16          4,449 → 4,45          4,968 → 4,97
```

Se o valor do milésimo for menor que 5, o algarismo do centésimo não será modificado. Por exemplo:

```
2,981 → 2,98          3,284 → 3,28          3,993 → 3,99
```

Se o valor do milésimo for igual a 5, iremos arredondar para cima o algarismo do centésimo se ele for *ímpar*, ou mantê-lo inalterado se for *par* (isto significa, em essência, que estaremos sempre arredondando o centésimo para o número *par* mais próximo):

```
5,765→5,76   6,385→6,38   6,415→6,42   6,675→6,68
```

Há duas observações importantes a fazer sobre este método de arredondamento. Primeiro, ele não leva em conta os sinais dos números; o valor absoluto do resultado é sempre o mesmo. Por exemplo,

```
+5,765 → +5,76
-5,765 → -5,76
```

(o mesmo não é verdade para os dois métodos que serão vistos na próxima seção).

Segundo, não é correto fazer por este método vários arredondamentos seguidos de um mesmo número. Se temos um número com duas decimais e queremos arredondá-lo para um valor inteiro (sem decimais) o resultado final poderá ser diferente se fizermos dois arredondamentos consecutivos, removendo uma decimal de cada vez, ou se fizermos um arredondamento único, removendo de uma vez as duas decimais. Por exemplo, considere o número X = 5,47. Se o arredondamos duas vezes, removendo a cada vez uma decimal, iremos obter:

```
5,47 → 5,5 → X≈6
```

Se porém o arredondamos de uma vez só, eliminando as duas decimais ao mesmo tempo, obtemos um resultado diferente:

```
5,47 → X≈5
```

O segundo valor é o correto, já que 5,47 está mais próximo de 5 do que de 6.

(iii) Outras técnicas de arredondamento

Na técnica de arredondamento mostrada no item anterior, o algarismo que ocupa uma posição pode ser arredondado para cima ou para baixo, dependendo do valor do algarismo que ocupa a posição seguinte. Existem porém outras técnicas nas quais os algarismos são arredondados sempre numa mesma direção: ou todos para cima, ou todos para baixo.

O arredondamento sempre para cima (isto é, no sentido de $+\infty$) é feito pela função `ceiling` do R, como nestes exemplos:

```
ceiling( 2.4)    →  3
ceiling( 5.5)    →  6
ceiling( 1.7)    →  2
ceiling(-1.7)    → -1
```

O arredondamento sempre para baixo (isto é, no sentido de $-\infty$) é feito pela função `floor`, como nos exemplos:

```
floor( 2.4)      →  2
floor( 5.5)      →  5
floor( 1.7)      →  1
floor(-1.7)      → -2
```

Note que estas duas funções do R arredondam sempre para valores inteiros. As mesmas técnicas podem porém ser aplicadas se quisermos arredondar, por exemplo, até os centésimos (usando como exemplo os mesmos dados do item (c)):

Arredondando sempre para cima:

```
2,981  3,284  3,993  4,022  4,158  4,449  4,968
  ↓      ↓      ↓      ↓      ↓      ↓      ↓
2,99   3,29   4,00   4,03   4,16   4,45   4,97
```

Usando os mesmos números, mas agora com sinal negativo:

```
-2,981 -3,284 -3,993 -4,022 -4,158 -4,449 -4,968
   ↓      ↓      ↓      ↓      ↓      ↓      ↓
-2,98  -3,28  -3,99  -4,02  -4,15  -4,44  -4,96
```

Arredondando sempre para baixo:

```
2,981  3,284  3,993  4,022  4,158  4,449  4,968
  ↓      ↓      ↓      ↓      ↓      ↓      ↓
2,98   3,28   3,99   4,02   4,15   4,44   4,96
```

Usando os mesmos números, com sinal negativo:

```
-2,981 -3,284 -3,993 -4,022 -4,158 -4,449 -4,968
   ↓      ↓      ↓      ↓      ↓      ↓      ↓
-2,99  -3,29  -4,00  -4,03  -4,16  -4,45  -4,97
```

A desvantagem destes dois métodos, em termos estatísticos, é a de introduzirem *viés* na média da amostra; esta média será sempre *maior* do que a média dos dados originais, se o arredondamento for feito para cima, ou *menor*, se o arredondamento for feito para baixo. Por exemplo, a média da amostra no item (c) é de 5.29519; se arredondarmos todos os dados até os centésimos, sempre para baixo, a média ficará igual a 5.290952; se arredondarmos para cima, igual a 5.299048. A média amostral será, portanto, um estimador *tendencioso* da média populacional. Uma maneira de evitar isto é usar a técnica mostrada no item (ii); esta técnica, por arredondar os números às vezes para cima, às vezes para baixo, tende a não alterar muito a média da amostra (no exemplo, se os dados forem arredondados até os centésimos, sua média será igual a 5.294762).

(iv) O truncamento de dados

As regras de arredondamento acima são fáceis de se implementar num programa de computador, mas pode ser trabalhoso demais aplicá-las quando estamos usando calculadoras. O *truncamento* é um procedimento mais simples; para truncar, simplesmente desconsideramos todos os algarismos além do número desejado de decimais e mantemos os restantes. Por exemplo, para truncar os valores da lista de dados no item (c) até os centésimos, simplesmente eliminamos os milésimos:

```
2,981   3,284   3,993   4,022   4,158   4,449   4,968
  ↓       ↓       ↓       ↓       ↓       ↓       ↓
2,98    3,28    3,99    4,02    4,15    4,44    4,96
```

Se quisermos truncar ainda mais, até os décimos, eliminamos também os centésimos:

```
2,9     3,2     3,9     4,0     4,1     4,4     4,9
```

O sinal dos números não faz diferença; o valor absoluto do número é sempre reduzido da mesma forma e o sinal é mantido:

```
-2,981  -3,284  -3,993  -4,022  -4,158  -4,449  -4,968
   ↓       ↓       ↓       ↓       ↓       ↓       ↓
-2,98   -3,28   -3,99   -4,02   -4,15   -4,44   -4,96
```

O truncamento está implementado em R na função `trunc`. Uma comparação entre um diagrama de ramo-e-folhas feito com dados de uma amostra, depois de truncados, e outro feito com os mesmos dados, depois de arredondados, pode ser vista na Fig. 6 da seção 2.1.2.4.

(v) *Considerações finais*

Existem ainda outras técnicas para arredondar números que têm aplicações em certas áreas especiais; por exemplo, as chamadas na Wikipédia de “arredondamento comercial” (*commercial rounding*) e “arredondamento estocástico” (*stochastic rounding).* Não iremos discutir estas técnicas neste texto, pois são de aplicação muito limitada.

Na maioria das aplicações práticas, a técnica mais usada é com certeza a de arredondamento vista no item (ii); às vezes, é usada também a de truncamento, vista no item (iv). (Um exemplo de variável truncada na vida cotidiana é a idade de uma pessoa – ninguém costuma arredondar para cima o valor real). Alguns autores sugerem contudo que o truncamento deveria ser usado com maior frequência, especialmente se estamos trabalhando com calculadoras [**2**]. Truncar os dados é mais fácil do que arredondá-los, e a probabilidade de cometermos erros é menor. Além disso, se precisarmos depois de encontrar o valor original do dado numa planilha, é mais fácil achar o valor original de um dado que foi truncado do que o de um que foi arredondado. Por exemplo, se queremos achar o valor original de um número que foi *truncado* para 55, basta procurar na planilha todos os números 55,__ - números que começam por 55 e têm depois algumas decimais. Se o número foi *arredondado* para 55, contudo, a situação é mais complicada, porque o valor original pode ter sido não apenas 55,__. mas também algum 54, __ que foi arredondado para cima, ou um 56, __ que foi arredondado para baixo.

Numa sequência de cálculos, é sempre melhor manter todas as decimais nos resultados intermediários (ou, pelo menos, manter um grande número de decimais) e arredondar só o resultado final. Se o arredondamento for feito a cada etapa, o erro no resultado final acabará sendo grande, pelo efeito cumulativo dos pequenos erros a cada etapa. Para um exemplo bem simples, suponha que queiramos calcular o volume V de um cilindro, cujo raio seja r = 2,143 e a altura h=4,379. Se queremos que o volume seja expresso com apenas uma decimal, mas calculamos sem arredondar as etapas intermediárias, obtemos (V: volume; S: superfície da base):

$$V = S \times h$$
$$S = \pi r^2 = 3{,}1416 \times 2{,}143^2 = 14{,}42764$$
$$V = Sh = 14{,}42764 \times 4{,}379 = 63{,}17863 \approx 63{,}2$$

Se porém arredondamos os valores do raio, do volume, e da superfície, para apenas uma casa decimal, teremos uma perda desnecessária de precisão no resultado final:

$$r = 2{,}143 \approx 2{,}1 \qquad h{=}4{,}379 \approx 4{,}4$$
$$S = \pi r^2 = 3{,}1416 \times 2{,}1^2 = 13.85446 \approx 13{,}9$$
$$V = Sh = 13{,}9 \times 4{,}4 = 60.95961 \approx 61{,}0$$

Referências

[1] https://en.wikipedia.org/w/index.php?title=Rounding&oldid=992399835"

[2] Hoaglin, Mosteller, Tukey. *Understanding robust and exploratory data analysis*. John Wiley & Sons, 1983, p.13.