UNIVERSIDADE UNIGRANRIO

Estatística e Probabilidade

Medidas de Posição

Núcleo de Educação a Distância
www.unigranrio.com.br
Rua Prof. José de Souza Herdy, 1.160
25 de Agosto – Duque de Caxias - RJ

Produção: Gerência de Desenho Educacional - NEAD

Desenvolvimento do material: Gregório Dalle Vedove Nosaki

1ª Edição

Sumário

Medidas de Posição

Para início de conversa...

Os dados coletados em uma pesquisa, medição ou experimento são também chamados de observáveis e cada variável possui um espectro de valores que podem ser atingidos. Vamos trabalhar com as medidas de posição, que visam a dar uma noção geral e sintética da coleção de valores que estão sendo considerados para uma determinada variável. As medidas de posição trabalhadas aqui também são conhecidas como medidas de tendência central, pois acompanham, de certa forma, a maioria dos valores que fazem parte da nossa coleção de dados considerados. Estudaremos a média aritmética, média ponderada, mediana e moda. Apresentaremos diversos exemplos para ilustrar as aplicações de tais conceitos tão importantes para a análise estatística. Além disso, trabalharemos também com as separatrizes, que podem ser adaptadas de acordo com o tamanho da amostra que estamos considerando para particionar esse grupo em subgrupos de mesmo tamanho e, dessa forma, podermos ter uma ideia da variação dos valores que determinada variável assume na nossa base de dados.

Objetivos

- Calcular e interpretar as medidas de tendência central: média, moda e mediana.
- Entender como as medidas de posição influenciam a forma da distribuição dos dados.
- Calcular e interpretar resultados de medidas separatrizes.

1. Medidas de tendência central

Além dos gráficos e das tabelas apresentados anteriormente, existem diversas informações que podem ser obtidas a partir de uma colação de dados que podem auxiliar na compreensão de determinadas características e comportamentos da nossa população ou amostra. Vamos explorar as medidas de tendência central que visam descrever de maneira geral uma série de dados. Começaremos estabelecendo a notação de somatório e produtório que serão utilizadas neste e nos próximos capítulos.

Definição: dada uma coleção de valores $x_1, x_2, \dots, x_n$, denotaremos o **somatório** dessa coleção como sendo

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \dots + x_n .$$

Considere um grupo de 5 pessoas cujas idades são 20, 23, 21, 18 e 21 anos. Esses valores representam uma coleção finita que pode ser representada como

$$\begin{aligned} x_1 &= 20 \\ x_2 &= 23 \\ x_3 &= 21 \\ x_4 &= 18 \\ x_5 &= 21 . \end{aligned}$$

O somatório da idade de todas pessoas do grupo é

$$\sum_{i=1}^{5} x_i = 20 + 23 + 21 + 18 + 21 = 103 .$$

A representação do somatório pode ser utilizada para qualquer conjunto de valores, por exemplo, a soma de todos os números naturais pode ser representada como sendo

$$\sum_{n \in \mathbb{N}} n = 1 + 2 + 3 + 4 + 5 + \dots$$

ou ainda

$$\sum_{n=1}^{+\infty} n = 1 + 2 + 3 + 4 + 5 + \dots .$$

Note que, neste caso, o somatório não é um valor finito, mas sempre que o conjunto de valores for finito, o somatório será finito. Utilizando essa notação, apresentaremos a primeira medida de tendência central que é a média.

Definição: dada uma coleção finita de valores $x_1, x_2, ..., x_n$, definimos a **média aritmética** ou apenas **média** como sendo

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad .$$

No exemplo anterior temos que a média da idade no grupo de cinco pessoas é

$$\overline{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{103}{5} = 20,6 \quad ,$$

ou seja, a idade média das pessoas nesse grupo é de 20,6 anos. A média nos auxilia a identificar de maneira mais generalizada o valor de uma determinada variável estudada, mas ela pode nem sempre representar a realidade. Considere ainda um grupo com 5 pessoas onde estamos interessados na idade das pessoas, mas dessa vez as idade são como apresentada a seguir:

$$\begin{aligned} y_1 &= 80 \\ y_2 &= 5 \\ y_3 &= 2 \\ y_4 &= 12 \\ y_5 &= 4 \quad . \end{aligned}$$

Calculando a idade média nesse grupo, obtemos

$$\sum_{i=1}^{5} y_i = 80 + 5 + 2 + 12 + 4 = 103$$

$$\overline{y} = \frac{1}{5}\sum_{i=1}^{5} y_i = \frac{103}{5} = 20,6 \quad .$$

A idade média no primeiro e no segundo grupo é a mesma! Devemos ter cuidado ao calcularmos a média em determinadas coleções de dados pois elas podem induzir a resultados que não correspondem à realidade. Dizer que em ambos os grupos a média das idades de cada uma das pessoas é de 20 anos não faz jus a realidade dos dados. Estudaremos nos próximos capítulos como medir esse tipo de variação entre os valores obtidos em uma amostra.

Outro tipo de média bastante utilizada é a **média ponderada**. Neste tipo de média, há uma diferença entre os pesos que cada um dos dados considerados. Um dos exemplos mais comuns no nosso cotidiano são as médias ponderadas para calcular a nota final de uma disciplina ou de uma prova dividida em diferentes etapas. Vejamos alguns exemplos desse tipo de média e como ela deve ser calculada.

Exemplo 1: três pessoas prestaram um mesmo processo seletivo para um concurso que foi dividido em três etapas distintas. A primeira etapa corresponde a uma prova escrita com peso 3, a segunda etapa corresponde uma entrevista com peso 2 e a terceira e última etapa corresponde à análise de currículo dos candidatos com peso 1. Observe a tabela a seguir com as notas dos três candidatos em cada uma das etapas.

Candidato	1ª etapa	2ª etapa	3ª etapa
A	7	8	6,5
B	8	6	6
C	6,5	9	8,5

As notas de cada uma das etapas varia de 0 a 10. Calculando a média ponderada do desempenho de cada um dos candidatos obtemos:

$$N_A = \frac{3 \cdot 7 + 2 \cdot 8 + 1 \cdot 6,5}{6} = \frac{43,5}{6} = 7,25$$

$$N_B = \frac{3 \cdot 8 + 2 \cdot 6 + 1 \cdot 7}{6} = \frac{42}{6} = 7$$

$$N_C = \frac{3 \cdot 6,5 + 2 \cdot 9 + 1 \cdot 8,5}{6} = \frac{46}{6} = 7,6666...$$

Note que para calcular a média ponderada devemos dividir pelo total da soma dos pesos de cada uma das etapas e não apenas pelo número de etapas.

Exemplo 2: considere a seguinte tabela que mostra o número de estudantes de acordo com o número de faltas em uma determinada disciplina.

Número de faltas	Quantidade de alunos
0	13
1	8
2	3
3	5
4	1

Para determinar qual a média de faltas para cada aluno desta disciplina devemos utilizar uma média ponderada. Multiplicaremos o número de faltas pelo número de alunos que correspondem na tabela e dividiremos pelo total de alunos, ou seja,

$$\overline{F} = \frac{0 \cdot 13 + 1 \cdot 8 + 2 \cdot 3 + 3 \cdot 5 + 4 \cdot 1}{30} = \frac{33}{30} = 1,1.$$

A média de faltas por alunos nessa disciplina é de pouco mais de uma falta.

Importante

As médias ponderadas também podem nos fornecer informações que não correspondem à realidade. No exemplo 2, por exemplo, a média de faltas por aluno foi de 1,1 falta, no entanto, mais de um terço dos alunos dessa turma não apresentou nenhuma falta. A partir da média não podemos inferir muitas informações quanto a distribuição dos dados dentro de nossa amostra.

Nos exemplos anteriores, trabalhamos com tabelas e distribuições de frequência para **dados não agrupados**, isto é, cada categoria considerada corresponde a apenas um valor. Esse tipo de representação geralmente é utilizado quando estamos trabalhando com variáveis numéricas discretas, como número de faltas, número de filhos, quantidade de televisores em uma residência e assim por diante. Outro tipo de distribuição de frequência pode ser descrita por meio de **dados agrupados**, geralmente utilizada quando trabalhamos com uma variável numérica contínua como, por exemplo, peso, altura, temperatura...

Veja, por exemplo, a tabela abaixo, que mostra a distribuição de frequências das alturas dos alunos de uma classe.

Altura (m)	Número de alunos
Menos de 1,60	3
1,60 ⊢ 1,70	12
1,70 ⊢ 1,80	16
1,80 ⊢ 1,90	8
Maior ou igual a 1,90	2

O símbolo "⊢" significa o **intervalo da classe** que estamos considerando. Por exemplo, as pessoas que estão na classe 1,60 ⊢ 1,70 têm entre 1,60m e 1,70. Nesta classe estão as pessoas que tem exatamente 1,60m, mas somente as pessoas que têm menos de 1,70m. Podemos considerar que se trata de um intervalo fechado à esquerda e aberto à direita, pois estamos trabalhando com uma variável numérica contínua e, portanto, os valores podem ser números reais.

Para calcular as medidas de posição que foram apresentadas e as indicadas a seguir, devemos determinar um valor dentro do intervalo da classe que irá representá-la e que utilizaremos nas fórmulas para distribuições de frequência de dados não-agrupados. Uma das formas mais simples de lidar com dados agrupados é determinar o valor médio do intervalo de cada classe e realizar o cálculo para aqueles valores. No exemplo anterior, podemos utilizar os valores apresentados na tabela a seguir.

Altura (m)	Número de alunos	Valor médio (m)
Menos de 1,60	3	1,60
1,60 ├ 1,70	12	1,65
1,70 ├ 1,80	16	1,75
1,80 ├ 1,90	8	1,85
Maior ou igual a 1,90	2	1,90

Utilizando o valor médio de cada uma das classes podemos proceder com o cálculo da média aritmética como sendo

$$\frac{3 \cdot 1,60 + 12 \cdot 1,65 + 16 \cdot 1,75 + 8 \cdot 1,85 + 2 \cdot 1,90}{41} = \frac{71,2}{41} \approx 1,73.$$

Apresentaremos mais exemplos de como trabalhar com dados agrupados no decorrer do capítulo.

Além das médias aritmética e ponderada, existem também outras médias que podem ser calculadas a partir de uma coleção de dados. A **média geométrica** de um conjunto de dados $X = \{x_1, x_2, \ldots, x_n\}$ será denotada por $m_g(X)$ e é dada pela fórmula

$$m_g(X) = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n},$$

ou seja, é o produto de todas as observáveis e deste resultado é extraída a raíz n-ésima onde n é o número de observáveis. Por exemplo, a média geométrica do conjunto $X = \{2, 4, 8\}$ é igual a

$$m_g(X) = \sqrt[3]{2 \cdot 4 \cdot 8} = \sqrt[3]{64} = 4.$$

A **média harmônica** de um conjunto de dados $X = \{x_1, x_2, \ldots, x_n\}$, denotada aqui por $m_h(X)$ é obtida pela seguinte fórmula

$$m_h(X) = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}.$$

Tomamos o inverso de cada uma das observáveis e realizamos a soma deles. A seguir efetuamos o quociente entre o número de observáveis e o total do somatório anterior. Utilizando o mesmo conjunto $X = \{2, 4, 8\}$, a média harmônica deste conjunto de observáveis é

$$m_h(X) = \frac{3}{\frac{1}{2} + \frac{1}{4} + \frac{1}{8}} = \frac{3}{\frac{7}{8}} = \frac{24}{7}.$$

Mais exemplos do cálculo dessas duas médias podem ser encontrados em Spiegel e Stephens (2009).

Apresentaremos agora duas outras medidas de posição que são muito utilizadas para sintetizar um conjunto de dados.

Definição: a **mediana** de uma coleção de valores é o valor que ocupa a posição central quando os dados são dispostos em ordem crescente.

Por exemplo, se nossos valores observados para uma determinada variável forem: 2, 3, 7, 5, 1, 2 e 6, devemos ordenar os valores e determinar o valor que está exatamente no meio de todos os valores em ordem crescente:

1 2 2 3 5 6 7

No caso do conjunto de observações apresentado anteriormente, a mediana é igual a 3. Caso o número de observações seja par, então a mediana é igual a média aritmética entre os dois valores que ocupam a posição central. Por exemplo, se temos os seguintes valores observados:

1 2 3 5 7 8 8 9

então a mediana é igual a

$$\frac{5+7}{2} = 6.$$

Muitas vezes, encontramos tabelas com a **distribuição de frequência acumulada,** que faz referência à soma de todas as ocorrências que são anteriores àquele determinado valor ou intervalo de classe. Observe a tabela a seguir, em que apresentamos a distribuição de frequência acumulada juntamente com a distribuição de frequências em um caso com dados agrupados.

Peso (kg)	Número de indivíduos	Distribuição de frequência acumulada
Menos de 50	2	4%
50 ⊢ 60	4	12%
60 ⊢ 70	11	34%
70 ⊢ 80	19	72%
80 ⊢ 90	8	88%
90 ⊢ 100	3	94%
100 ⊢ 110	2	98%
Acima de 110	1	100%

Para calcular a mediana para dados agrupados, devemos proceder da seguinte maneira: primeiro devemos determinar em qual intervalo de classe está o valor que divide as observáveis ao meio, no caso do exemplo anterior é a classe 70 ⊢ 80, pois pela coluna da frequência acumulada sabemos que 72% dos valores das observáveis são estritamente menores que 80 enquanto que 34% das observáveis são menores que

70. Determinada a classe onde o valor que divide as observáveis ao meio, utilizamos a seguinte fórmula:

$$md = \ell_{inf} + \frac{\frac{n}{2} - F_{ant}}{f} \cdot h$$

onde ℓ_{inf} é o limite inferior da classe determinada anteriormente; n é o número total de observáveis; F_{ant} é o número total de observáveis que pertencem às classes menores que a classe onde está localizada a mediana; f é o número de observáveis que pertencem ao intervalo onde está localizada a mediana e h é o comprimento do intervalo de classe.

No exemplo anterior, temos que $n = 50$, $\ell_{inf} = 70$, $F_{ant} = 2 + 4 + 11 = 17$, $f = 19$ e $h = 10$. Portanto, a mediana md é igual a

$$md = 70 + \frac{25 - 17}{19} \cdot 10 \approx 74,21kg.$$

Definição: a **moda** de uma coleção de valores é o valor com maior número de ocorrências dentro da coleção.

Por exemplo, dada a coleção de dados

1 2 2 2 3 3 5 6 6 7

então, 2 é a moda dessa coleção. No caso de dados agrupados, a moda é igual ao ponto médio do intervalo de classe de maior frequência dentro da amostra. No exemplo da tabela com o peso de 50 indivíduos divididos em intervalos de classe, o intervalo com maior frequência é o 70 ⊢ 80 e, portanto, a moda é igual a 75kg.

Vejamos um exemplo com o cálculo das medidas de posição em um experimento no qual as observáveis são as faces superiores de um dado em uma sequência de lançamentos.

Exemplo 3: Considere 10 lançamento de um dado de 6 faces cujos resultados são apresentados na lista a seguir:

2 5 4 5 3 6 1 5 3 6

Vamos determinar a média, a mediana e a moda desses lançamentos. Para calcular a média devemos somar todos os valores e dividir pelo número total de lançamentos. Dessa forma, obtemos que

$$\frac{2+5+4+5+3+6+1+5+3+6}{10} = \frac{40}{10} = 4.$$

Para calcular a mediana e a moda, organizaremos os dados observados em ordem crescente:

1 2 3 3 4 5 5 5 6 6

Como estamos trabalhando com um conjunto de números pares de observações, a mediana será a média entre os dois valores centrais, ou seja,

$$\frac{4+5}{2} = \frac{9}{2} = 4,5 \quad .$$

Observando os resultados obtidos, concluímos que a moda desses 10 lançamentos foi 5. Portanto, para esses 10 lançamentos realizados com um dado de seis faces, a média é igual a 4, a mediana é igual a 4,5 e a moda é igual a 5.

Exemplo 4: os jogos olímpicos são competições de diversas modalidades esportivas que ocorrem de 4 em 4 anos em diferentes países do mundo. Em cada modalidade são premiados os três primeiros lugares com medalhas de ouro, prata e bronze, para o primeiro, segundo e terceiro colocado, respectivamente.

Observe o quadro de medalhas do Brasil nas últimas seis edições dos jogos olímpicos apresentado a seguir.

Edição do jogos	Ouro	Prata	Bronze
Rio 2016	7	6	6
Londres 2012	3	5	9
Pequim 2008	3	4	10
Atenas 2004	5	2	3
Sydney 2000	0	6	6
Atlanta 2000	3	3	9

Tabela 1: Quadro de medalhas do Brasil nas últimas seis edições dos jogos olímpicos. Fonte: COB.

Vamos determinar qual a média de cada um dos três tipos de medalhas que o Brasil teve nas últimas seis edições dos jogos olímpicos.

Média de medalhas de ouro: $\frac{7+3+3+5+0+3}{6} = \frac{21}{6} = 3,5$

Média de medalhas de prata: $\frac{6+5+4+2+6+3}{6} = \frac{26}{6} = 4,33$

Média de medalhas de bronze: $\frac{6+9+10+3+6+9}{6} = \frac{43}{6} = 7,16$

Considere agora que cada medalha de ouro seja equivalente a 5 pontos, cada medalha de prata seja equivalente a 3 pontos e cada medalha de bronze seja equivalente a 1 ponto. Vamos determinar a média de pontos que o Brasil teve nas 6 últimas edições dos jogos olímpicos.

Primeiro devemos calcular os pontos totais que o Brasil teve em cada uma das edições dos jogos olímpicos apresentada na tabela. Para calcular o total de pontos devemos levar em consideração além do número de medalhas, também a pontuação correspondente a cada uma de acordo com o seu tipo. Para calcular o total de pontos que o Brasil teve na edição de 2016, procedemos da seguinte maneira:

$$7 \cdot 5 + 6 \cdot 3 + 6 \cdot 1 = 59 \, .$$

Repetindo o processo para todas as outras linhas da tabela apresentada anteriormente, obtemos os seguintes resultados:

Edição dos jogos	Pontuação do Brasil
Rio 2016	59
Londres 2012	39
Pequim 2008	37
Atenas 2004	34
Sydney 2000	24
Atlanta 2000	33

Agora podemos calcular a média aritmética da pontuação do Brasil nas seis últimas edições e assim obtemos

$$\frac{59+39+37+34+24+33}{6} = \frac{226}{6} = 37,66 \, .$$

Note que a média que estamos tomando aqui é a média aritmética. Para o cálculo da pontuação do Brasil por cada edição, usamos um sistema de pontuação com pesos distintos para as medalhas que se assemelha à média ponderada, mas neste caso não dividimos pela soma total dos pesos, pois estamos apenas determinando a quantidade de pontos. Logo, a média de pontos do Brasil seguindo o sistema de pontuação descrito anteriormente foi de 37,66.

Curiosidade

A pandemia causada pelo coronavírus, que teve início no final de 2019, acarretou o adiamento dos jogos olímpicos de Tóquio que ocorreriam no ano de 2020. Essa não foi a primeira vez que eventos globais afetaram a realização dos jogos olímpicos. As edições dos jogos de 1916, 1940 e 1944 foram canceladas em razão das duas grandes guerras mundiais.

2. Separatrizes

As separatrizes podem ser vistas como generalizações do conceito de mediana. Enquanto a mediana divide os valores de uma coleção ao meio, as separatrizes podem dividir os valores da coleção em quantas partes forem estipuladas. Os **quartis** são um exemplo de separatrizes que dividem uma dada coleção de dados em quatro partes: devemos ordenar os dados em ordem crescente e depois dividir os dados observados de modo a encontrar os valores que dividem a coleção em quatro partes. Vejamos um exemplo numérico para ilustrar essa definição.

Exemplo 5: Considere a tabela a seguir com a temperatura mínima e máxima mensal registrada na cidade de São Paulo:

Mês	Temperatura mínima (ºC)	Temperatura máxima (ºC)
Janeiro	19,4	26,3
Fevereiro	19,4	26,8
Março	18,7	25,8
Abril	17,1	24,6
Maio	14,3	22
Junho	13,1	21,7
Julho	12,3	21,5
Agosto	12,9	22,9
Setembro	14,6	24,1
Outubro	16,2	24,8
Novembro	16,9	24,5
Dezembro	18,4	25,8

Na tabela acima, temos dados referentes às temperaturas mínimas e máximas mensais registradas em um determinado ponto da cidade de São Paulo. Temos portanto, um conjunto de 12 observáveis para cada variável (mínima e máxima). Vamos determinar os quartis de cada uma dessas variáveis. Devemos ordenar os valores em ordem crescente, assim como fizemos para determinar a mediana. Para as temperaturas mínimas temos:

12,3
12,9
13,1
14,3
14,6
16,2
16,9
17,1
18,4
18,7
19,4
19,4

Como no total temos 12 observáveis, ao dividirmos nossos dados em 4 subgrupos com o mesmo número de elementos, cada subgrupo terá 3 observáveis. Para determinar os valores dos quartis, podemos determinar primeiro a mediana e depois a mediana de cada um dos subgrupos de dados obtidos.

12,3

12,9

13,1

14,3

14,6

16,2

16,9

$$\Rightarrow \frac{16,2+16,9}{2} = 16,55$$

17,1

18,4

18,7

19,4

19,4

Portanto, a mediana é igual a 16,55. Temos agora duas coleções de 6 dados cada e, portanto, basta proceder da mesma maneira:

12,3

12,9

13,1

14,3

$$\Rightarrow \frac{13,1+14,3}{2} = 13,7$$

14,6

16,2

16,9

17,1

18,4

18,7

$$\Rightarrow \frac{18,4+18,7}{2} = 18,55$$

19,4

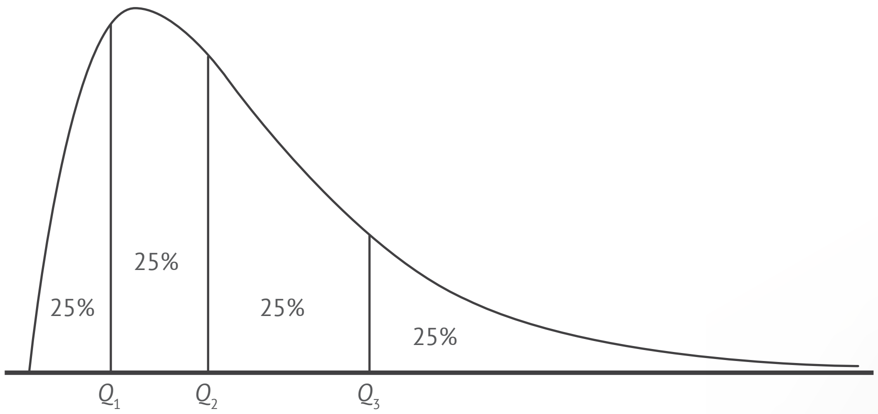
19,4

Os quartis da temperatura mínima são: 13,7; 16,55 e 18,55. Procedendo da mesma maneira para as temperatura máxima obtemos os quartis: 22,45; 24,3 e 25,8.

Observe que os quartis de uma distribuição de frequência são 3 valores que dividem os dados em quatro subgrupos de mesmo tamanho. Se chamarmos os quartis de Q_1, Q_2 e Q_3, então os quatro quartis são determinados como sendo:

1. O primeiro é formado pelas observáveis que são menores que Q_1;
2. O segundo é formado pelas observável que estão entre Q_1 e Q_2;
3. O terceiro é formado pelas observáveis que estão entre Q_2 e Q_3;
4. O quarto é formado pelas observáveis que são maiores que Q_3.

O gráfico a seguir mostra uma representação de uma distribuição assimétrica de dados dividida pelos quartis. Note que cada subconjunto definido pelos quartis tem 25% do total das observáveis que a distribuição representa.



De forma análoga, podemos dividir nosso subconjunto de dados em quantas parcelas forem necessárias. Quanto maior o volume de dados considerados, dividi-lo em parcelas menores torna mais fácil compreender o comportamento e a distribuição dos valores registrados por meio das separatrizes.

Alguns autores usam a denominação **quantil** para significar os valores que indicam dividem o conjunto das observáveis de acordo com uma proporção q com $0 < q < 1$, onde os valores menores que P_q correspondem a uma fração q do total de observáveis. Por exemplo, o quantil tomando $q = 0,5$ é a mediana, pois esse quantil divide o conjunto de observáveis de modo que 50% dos valores está abaixo dele e 50% dos valores está acima dele. O quantil P_q é tal que uma proporção de q de todas as observáveis está abaixo dele e uma proporção de $1 - q$ de todas as observáveis são maiores que ele.

Um procedimento simples para poder determinar os quantis de um conjunto com n observáveis começa com a ordenação de todos os dados em ordem crescente. Seja $x_1, x_2, ..., x_n$ a ordenação de todas os dados em ordem crescente, ou seja, $x_1 \leq x_2 \leq ... \leq x_n$. Determinado a proporção q devemos determinar quantas das n observáveis corresponde a essa proporção, no caso nq.

A partir daí, temos duas possibilidades para determinar o quantil: supondo que $nq = k + m$ onde $k \in \mathbb{Z}$ e $0 \leq m < 1$,

- se $m = 0$, então o quantil P_q é tal que

$$P_q = \frac{x_j + x_{j+1}}{2};$$

- se $m > 0$, então o quantil P_q é tal que

$$P_q = x_{j+1}.$$

Considere a coleção de observáveis

X = 2 7 9 7 5 1 7 1.

Ordenando as observáveis em ordem crescente, temos

1 1 2 5 7 7 7 9.

Para determinar o quantil quando $q = 0,25$, então

$$0,25 \cdot 8 = 2$$

e, portanto,

$$P_{0,25} = \frac{1+2}{2} = \frac{3}{2} = 1,5.$$

Isso significa que 25% dos valores do conjunto X são menores que 1,5.

Tomando $q = 0,6$, temos então

$$0,6 \cdot 8 = 4,8$$

e, portanto,

$$P_{0,6} = 7.$$

Isso significa que 60% dos valores do conjunto X são menores que 7.

Esse processo de determinação de quantis pode ser aplicado para qualquer tipo de separatriz. Basta que, a partir do número de subconjuntos que se quer dividir a coleções de dados, determine a porcentagem que cada um dos conjuntos irá representar do total. Para determinar os valores de cada separatriz basta aplicar o mesmo procedimento descrito acima para cada um dos quantis necessários. Os quantis que determinam as separatrizes de ordem 4, isso é, os quartis são $P_{0,25}$, $P_{0,5}$ e $P_{0,75}$.

Importante

É indispensável considerar o volume de observáveis para realizar o cálculo das separatrizes de um conjunto. No exemplo 5, consideramos um conjunto com 12 observáveis referentes a cada um dos meses do ano, portanto dividir esse conjunto em decis (dez partes iguais) não faria muito sentido. Esse tipo de separatrizes são utilizadas em conjuntos com uma quantidade muito grande de dados.

As medidas de posição podem nos dar uma ideia da maneira como os valores de uma determinada variável assumem dentro de uma coleção de dados. Vimos que tais medidas sozinhas podem não ser suficientes para descrever o comportamento de um conjunto de valores. Neste capítulo, descrevemos os conceitos de média aritmética, média ponderada, mediana e moda. Trabalhamos também com as separatrizes que particionam nosso conjunto de dados em subconjunto de mesmo tamanho de acordo com a nossa necessidade. Apresentamos diversos exemplos comentados para fixar os conceitos apresentados. Nos próximos capítulos, trabalharemos com outras formas de descrever de maneira sintética um determinado conjunto de dados que possibilitam uma melhor interpretação dessa coleção.

Referências

CLIMA São Paulo: Temperatura, Tempo e Dados climatológicos São Paulo. Disponível em: https://pt.climate-data.org/america-do-sul/brasil/sao-paulo/sao-paulo-655/. Acesso em: 23 jun. 2021.

MEDALHAS Olímpicas por edição dos jogos. **COB**, [2021]. Disponível em: https://www.cob.org.br/pt/cob/time-brasil/brasil-nos-jogos/medalhas-olimpicas. Acesso em: 23 jun. 2021.

FONSECA, J. S. da; MARTINS, G. de A. **Curso de estatística**. 6. ed. São Paulo: Atlas, 2012.

MATTOS, V. L. D. de; AZAMBUJA, A. M. V. de; K, A. C. **Introdução à estatística**: aplicações em ciências exatas. Rio de Janeiro: LTC, 2017.

MOORE, D.; NOTZ, W.; FLIGNER, M. **A estatística básica e sua prática**. 7. ed. Rio de Janeiro: LTC, 2017.

MORETTIN, P.; BUSSAB, W. **Estatística básica**. 9. ed. São Paulo: Saraiva, 2017.

SPIEGEL, M.; STEPHENS, L. **Estatística**. Porto Alegre: Bookman, 2009.

SPIEGEL, M.; SCHILLIER, J.; SRINIVASAN, A. **Probabilidade e estatística**. 3. ed. Porto Alegre: Bookman, 2013.

TRIOLA, M. **Introdução à estatística**. 12. ed. Rio de Janeiro: LTC, 2017.

VIEIRA, S. **Fundamentos de Estatística**. 6. ed. São Paulo: Atlas, 2019.