Probabilidade e estatística - Aula 12

Estatística descritiva

Dr. Giannini Italino Alves Vieira

Universidade Federal do Ceará - Campus de Crateús

2024

Estatística descritiva

Motivação

- A estatística descritiva é um importante e fundamental aspecto que consiste em resumir os dados, objetivando facilitar a interpretação e análise dos mesmos.
- Veremos, na aula de hoje, métodos para resumir dados. Tais métodos capturam características importantes dos dados, tais como tendência central e variabilidade. Essas características são frequentemente importantes para a tomada de decisão.
- Na próxima aula veremos alguns recursos para podermos fazer apresentações de dados. Resumos e apresentações de dados, bem constituídos, são fundamentais para fazermos uma boa análise estatística, uma vez que, a partir destes, podemos extrair características importantes dos dados e também ter noção sobre o tipo de modelo que é mais apropriado na solução do problema.

Estatística descritiva

Motivação

- A estatística descritiva é um importante e fundamental aspecto que consiste em resumir os dados, objetivando facilitar a interpretação e análise dos mesmos.
- Veremos, na aula de hoje, métodos para resumir dados. Tais métodos capturam características importantes dos dados, tais como tendência central e variabilidade. Essas características são frequentemente importantes para a tomada de decisão.
- Na próxima aula veremos alguns recursos para podermos fazer apresentações de dados. Resumos e apresentações de dados, bem constituídos, são fundamentais para fazermos uma boa análise estatística, uma vez que, a partir destes, podemos extrair características importantes dos dados e também ter noção sobre o tipo de modelo que é mais apropriado na solução do problema.

Estatística descritiva

Motivação

- A estatística descritiva é um importante e fundamental aspecto que consiste em resumir os dados, objetivando facilitar a interpretação e análise dos mesmos.
- Veremos, na aula de hoje, métodos para resumir dados. Tais métodos capturam características importantes dos dados, tais como tendência central e variabilidade. Essas características são frequentemente importantes para a tomada de decisão.
- Na próxima aula veremos alguns recursos para podermos fazer apresentações de dados. Resumos e apresentações de dados, bem constituídos, são fundamentais para fazermos uma boa análise estatística, uma vez que, a partir destes, podemos extrair características importantes dos dados e também ter noção sobre o tipo de modelo que é mais apropriado na solução do problema.

Em todas as medidas de resumo que veremos a seguir o termo *amostra* se refere às observações coletadas a partir de uma população de interesse, isto é, a partir do conjunto de todos os elementos que estão sob investigação.

Média da amostra

Se as n observações em uma amostra forem denotadas por $x_1, x_2, \ldots, x_n$ então a média amostral é definida por

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Intuitivamente, a média da amostra é uma medida de tendência central dos dados.
- Veremos, mais na frente, que a média da amostra, isto é $\bar{x}$, é uma boa estimativa para a média da população.

Estatística descritiva

Em todas as medidas de resumo que veremos a seguir o termo *amostra* se refere às observações coletadas a partir de uma população de interesse, isto é, a partir do conjunto de todos os elementos que estão sob investigação.

Média da amostra

Se as n observações em uma amostra forem denotadas por $x_1, x_2, \ldots, x_n$ então a média amostral é definida por

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Intuitivamente, a média da amostra é uma medida de tendência central dos dados.
- Veremos, mais na frente, que a média da amostra, isto é $\bar{x}$, é uma boa estimativa para a média da população.

Estatística descritiva

Em todas as medidas de resumo que veremos a seguir o termo *amostra* se refere às observações coletadas a partir de uma população de interesse, isto é, a partir do conjunto de todos os elementos que estão sob investigação.

Média da amostra

Se as n observações em uma amostra forem denotadas por $x_1, x_2, \ldots, x_n$ então a média amostral é definida por

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Intuitivamente, a média da amostra é uma medida de tendência central dos dados.
- Veremos, mais na frente, que a média da amostra, isto é $\bar{x}$, é uma boa estimativa para a média da população.

Exemplo

Exemplo: Prevenir a propagação de trinca de fadiga em estruturas de aviões é um importante elemento da segurança em aeronaves. Um estudo de engenharia para investigar a trinca de fadiga em $n = 9$ asas carregadas ciclicamente reportou os seguintes comprimentos (em mm) de trinca:

$$2.13, 2.96, 3.02, 1.82, 1.15, 1.37, 2.04, 2.47, 2.60$$

Calcule a média da amostra.

- Sol. Temos que as 9 observações da nossa amostra são $x_1 = 2.13, x_2 = 2.96, x_3 = 3.02, x_4 = 1.82, x_5 = 1.15, x_6 = 1.37, x_7 = 2.04, x_8 = 2.47, x_9 = 2.60$. Logo, a média da amostra é

$$\bar{x} = \frac{2.13 + 2.96 + 3.02 + 1.82 + 1.15 + 1.37 + 2.04 + 2.47 + 2.60}{9} \approx 2.173.$$

Exemplo

Exemplo: Prevenir a propagação de trinca de fadiga em estruturas de aviões é um importante elemento da segurança em aeronaves. Um estudo de engenharia para investigar a trinca de fadiga em $n = 9$ asas carregadas ciclicamente reportou os seguintes comprimentos (em mm) de trinca:

$$2.13, 2.96, 3.02, 1.82, 1.15, 1.37, 2.04, 2.47, 2.60$$

Calcule a média da amostra.

- Sol. Temos que as 9 observações da nossa amostra são $x_1 = 2.13, x_2 = 2.96, x_3 = 3.02, x_4 = 1.82, x_5 = 1.15, x_6 = 1.37, x_7 = 2.04, x_8 = 2.47, x_9 = 2.60$. Logo, a média da amostra é

$$\bar{x} = \frac{2.13 + 2.96 + 3.02 + 1.82 + 1.15 + 1.37 + 2.04 + 2.47 + 2.60}{9} \approx 2.173.$$

Exemplo

Exemplo: Prevenir a propagação de trinca de fadiga em estruturas de aviões é um importante elemento da segurança em aeronaves. Um estudo de engenharia para investigar a trinca de fadiga em $n = 9$ asas carregadas ciclicamente reportou os seguintes comprimentos (em mm) de trinca:

$$2.13, 2.96, 3.02, 1.82, 1.15, 1.37, 2.04, 2.47, 2.60$$

Calcule a média da amostra.

- Sol. Temos que as 9 observações da nossa amostra são $x_1 = 2.13, x_2 = 2.96, x_3 = 3.02, x_4 = 1.82, x_5 = 1.15, x_6 = 1.37, x_7 = 2.04, x_8 = 2.47, x_9 = 2.60$. Logo, a média da amostra é

$$\bar{x} = \frac{2.13 + 2.96 + 3.02 + 1.82 + 1.15 + 1.37 + 2.04 + 2.47 + 2.60}{9} \approx 2.173.$$

Variância da amostra

- Embora a média seja uma medida interessante, ela não captura todas as informações a respeito da amostra.
- Uma outra medida importante, que captura a variabilidade dos dados, é a variância da amostra.

Variância da amostra

Se $x_1, x_2, \ldots, x_n$ for uma amostra de n observações, então a variância da amostra é definida por

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

O desvio-padrão da amostra, denotado por s, é a raiz quadrada positiva da variância da amostra.

- Note que as unidades de medida da variância são o quadrado das unidades de medida originais da variável, enquanto que o desvio-padrão tem mesma unidade de medida dos dados.

Variância da amostra

- Embora a média seja uma medida interessante, ela não captura todas as informações a respeito da amostra.
- Uma outra medida importante, que captura a variabilidade dos dados, é a variância da amostra.

Variância da amostra

Se $x_1, x_2, \ldots, x_n$ for uma amostra de n observações, então a variância da amostra é definida por

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

O desvio-padrão da amostra, denotado por s, é a raiz quadrada positiva da variância da amostra.

- Note que as unidades de medida da variância são o quadrado das unidades de medida originais da variável, enquanto que o desvio-padrão tem mesma unidade de medida dos dados.

Variância da amostra

- Embora a média seja uma medida interessante, ela não captura todas as informações a respeito da amostra.
- Uma outra medida importante, que captura a variabilidade dos dados, é a variância da amostra.

Variância da amostra

Se $x_1, x_2, \ldots, x_n$ for uma amostra de n observações, então a variância da amostra é definida por

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

O desvio-padrão da amostra, denotado por s, é a raiz quadrada positiva da variância da amostra.

- Note que as unidades de medida da variância são o quadrado das unidades de medida originais da variável, enquanto que o desvio-padrão tem mesma unidade de medida dos dados.

- Embora a média seja uma medida interessante, ela não captura todas as informações a respeito da amostra.
- Uma outra medida importante, que captura a variabilidade dos dados, é a variância da amostra.

Variância da amostra

Se $x_1, x_2, \ldots, x_n$ for uma amostra de n observações, então a variância da amostra é definida por

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

O desvio-padrão da amostra, denotado por s, é a raiz quadrada positiva da variância da amostra.

- Note que as unidades de medida da variância são o quadrado das unidades de medida originais da variável, enquanto que o desvio-padrão tem mesma unidade de medida dos dados.

Variância da amostra

- Embora a média seja uma medida interessante, ela não captura todas as informações a respeito da amostra.
- Uma outra medida importante, que captura a variabilidade dos dados, é a variância da amostra.

Variância da amostra

Se $x_1, x_2, \ldots, x_n$ for uma amostra de n observações, então a variância da amostra é definida por

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

O desvio-padrão da amostra, denotado por s, é a raiz quadrada positiva da variância da amostra.

- Note que as unidades de medida da variância são o quadrado das unidades de medida originais da variável, enquanto que o desvio-padrão tem mesma unidade de medida dos dados.

Variância da amostra

- Embora a média seja uma medida interessante, ela não captura todas as informações a respeito da amostra.
- Uma outra medida importante, que captura a variabilidade dos dados, é a variância da amostra.

Variância da amostra

Se $x_1, x_2, \ldots, x_n$ for uma amostra de n observações, então a variância da amostra é definida por

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

O desvio-padrão da amostra, denotado por s, é a raiz quadrada positiva da variância da amostra.

- Note que as unidades de medida da variância são o quadrado das unidades de medida originais da variável, enquanto que o desvio-padrão tem mesma unidade de medida dos dados.

Alternativa ao cálculo da variância da amostra

- Note que o cálculo de s^2 necessita do cálculo de $\bar{x}$ e de n diferenças $(x_i - \bar{x})^2$. Se $(x_i - \bar{x})$ não for inteiro, então calcular a variância da amostra pode resultar em bastante trabalho. Uma maneira alternativa de se obter a variância da amostra, que em alguns casos pode simplificar os cálculos, é a seguinte:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

ou seja,

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2\bar{x}\frac{n}{n}\sum_{i=1}^{n}x_i + n\bar{x}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right)$$

ou ainda

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right)$$

Variância da amostra

Alternativa ao cálculo da variância da amostra

- Note que o cálculo de s^2 necessita do cálculo de $\bar{x}$ e de n diferenças $(x_i - \bar{x})^2$. Se $(x_i - \bar{x})$ não for inteiro, então calcular a variância da amostra pode resultar em bastante trabalho. Uma maneira alternativa de se obter a variância da amostra, que em alguns casos pode simplificar os cálculos, é a seguinte:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

ou seja,

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - 2\bar{x}\frac{n}{n}\sum_{i=1}^{n} x_i + n\bar{x}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right)$$

ou ainda

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$$

Variância da amostra

Alternativa ao cálculo da variância da amostra

- Note que o cálculo de s^2 necessita do cálculo de $\bar{x}$ e de n diferenças $(x_i - \bar{x})^2$. Se $(x_i - \bar{x})$ não for inteiro, então calcular a variância da amostra pode resultar em bastante trabalho. Uma maneira alternativa de se obter a variância da amostra, que em alguns casos pode simplificar os cálculos, é a seguinte:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

ou seja,

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2\bar{x}\frac{n}{n}\sum_{i=1}^{n}x_i + n\bar{x}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right)$$

ou ainda

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right)$$

Variância da amostra

Alternativa ao cálculo da variância da amostra

- Note que o cálculo de s^2 necessita do cálculo de $\bar{x}$ e de n diferenças $(x_i - \bar{x})^2$. Se $(x_i - \bar{x})$ não for inteiro, então calcular a variância da amostra pode resultar em bastante trabalho. Uma maneira alternativa de se obter a variância da amostra, que em alguns casos pode simplificar os cálculos, é a seguinte:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

ou seja,

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2\bar{x}\frac{n}{n}\sum_{i=1}^{n}x_i + n\bar{x}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right)$$

ou ainda

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right)$$

Alternativa ao cálculo da variância da amostra

- Note que o cálculo de s^2 necessita do cálculo de $\bar{x}$ e de n diferenças $(x_i - \bar{x})^2$. Se $(x_i - \bar{x})$ não for inteiro, então calcular a variância da amostra pode resultar em bastante trabalho. Uma maneira alternativa de se obter a variância da amostra, que em alguns casos pode simplificar os cálculos, é a seguinte:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

ou seja,

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2\bar{x}\frac{n}{n}\sum_{i=1}^{n}x_i + n\bar{x}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right)$$

ou ainda

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right)$$

Alternativa ao cálculo da variância da amostra

- Note que o cálculo de s^2 necessita do cálculo de $\bar{x}$ e de n diferenças $(x_i - \bar{x})^2$. Se $(x_i - \bar{x})$ não for inteiro, então calcular a variância da amostra pode resultar em bastante trabalho. Uma maneira alternativa de se obter a variância da amostra, que em alguns casos pode simplificar os cálculos, é a seguinte:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

ou seja,

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2\bar{x}\frac{n}{n}\sum_{i=1}^{n}x_i + n\bar{x}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right)$$

ou ainda

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right)$$

Exemplo

Exemplo: Considere novamente o exemplo anterior, ou seja, os dados referentes a trinca de fadiga. Temos que os dados são os seguintes:

$$2.13, 2.96, 3.02, 1.82, 1.15, 1.37, 2.04, 2.47, 2.60$$

Suponha que nosso interesse seja em obter a variância dessa amostra.

- Sol. No exercício anterior, obtivemos que a média da amostra é aproximadamente 2.173. Logo, temos que a variância da amostra é dada por

$$s^2 = \frac{1}{9-1}\left(\sum_{i=1}^{9} x_i^2 - 9(2.173)^2\right)$$

ou seja,

$$s^2 = \frac{2.13^2 + 2.96^2 + 3.02^2 + \ldots + 2.60^2 - 9(2.173)^2}{8} \approx 0.43035$$

Exemplo

Exemplo: Considere novamente o exemplo anterior, ou seja, os dados referentes a trinca de fadiga. Temos que os dados são os seguintes:

$$2.13, 2.96, 3.02, 1.82, 1.15, 1.37, 2.04, 2.47, 2.60$$

Suponha que nosso interesse seja em obter a variância dessa amostra.

- Sol. No exercício anterior, obtivemos que a média da amostra é aproximadamente 2.173. Logo, temos que a variância da amostra é dada por

$$s^2 = \frac{1}{9-1}\left(\sum_{i=1}^{9} x_i^2 - 9(2.173)^2\right)$$

ou seja,

$$s^2 = \frac{2.13^2 + 2.96^2 + 3.02^2 + \ldots + 2.60^2 - 9(2.173)^2}{8} \approx 0.43035$$

Exemplo

Exemplo: Considere novamente o exemplo anterior, ou seja, os dados referentes a trinca de fadiga. Temos que os dados são os seguintes:

$$2.13, 2.96, 3.02, 1.82, 1.15, 1.37, 2.04, 2.47, 2.60$$

Suponha que nosso interesse seja em obter a variância dessa amostra.

- Sol. No exercício anterior, obtivemos que a média da amostra é aproximadamente 2.173. Logo, temos que a variância da amostra é dada por

$$s^2 = \frac{1}{9-1}\left(\sum_{i=1}^{9} x_i^2 - 9(2.173)^2\right)$$

ou seja,

$$s^2 = \frac{2.13^2 + 2.96^2 + 3.02^2 + \ldots + 2.60^2 - 9(2.173)^2}{8} \approx 0.43035$$

Exemplo

Exemplo: Considere novamente o exemplo anterior, ou seja, os dados referentes a trinca de fadiga. Temos que os dados são os seguintes:

$$2.13, 2.96, 3.02, 1.82, 1.15, 1.37, 2.04, 2.47, 2.60$$

Suponha que nosso interesse seja em obter a variância dessa amostra.

- Sol. No exercício anterior, obtivemos que a média da amostra é aproximadamente 2.173. Logo, temos que a variância da amostra é dada por

  $$s^2 = \frac{1}{9-1}\left(\sum_{i=1}^{9} x_i^2 - 9(2.173)^2\right)$$

  ou seja,

  $$s^2 = \frac{2.13^2 + 2.96^2 + 3.02^2 + \ldots + 2.60^2 - 9(2.173)^2}{8} \approx 0.43035$$

Mediana

- A mediana é uma outra medida de tendência central, que divide a amostra em duas partes iguais, metade abaixo da mediana e metade acima.

Mediana

A fim de entender como se calcula a mediana, considere que $x_1, x_2, \ldots, x_n$ é uma amostra de n observações e que

$$x_{(1)}, x_{(2)}, \ldots, x_{(n)}$$

é a ordenação da amostra em ordem crescente, ou seja, $x_{(1)}$ é a menor observação, $x_{(2)}$ é a segunda e assim por diante. Então a mediana é dada por

$$Md = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ é impar,} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

Mediana

- A mediana é uma outra medida de tendência central, que divide a amostra em duas partes iguais, metade abaixo da mediana e metade acima.

Mediana

A fim de entender como se calcula a mediana, considere que $x_1, x_2, \ldots, x_n$ é uma amostra de n observações e que

$$x_{(1)}, x_{(2)}, \ldots, x_{(n)}$$

é a ordenação da amostra em ordem crescente, ou seja, $x_{(1)}$ é a menor observação, $x_{(2)}$ é a segunda e assim por diante. Então a mediana é dada por

$$Md = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ é impar,} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

Exemplo

Exemplo: Considere novamente os dados referentes a trinca de fadiga, fornecidos no primeiro exemplo. Temos que os dados, ordenados, são os seguintes: $x_{(1)} = 1.15, x_{(2)} = 1.37, x_{(3)} = 1.82, x_{(4)} = 2.04, x_{(5)} = 2.13, x_{(6)} = 2.47, x_{(7)} = 2.60, x_{(8)} = 2.96, x_{(9)} = 3.02.$

Como $n = 9$ (impar), então temos que

$$Md = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{10}{2}\right)} = x_{(5)} = 2.13.$$

Outro exemplo

Reconsidere novamente os dados acima. Suponha que uma outra medida referente a trinca de fadiga foi coletada e o valor dessa observação foi $x_{(10)} = 3.03$.

Note que a mediana desse novo conjunto de dados é

$$Md = \frac{x_{\left(\frac{10}{2}\right)} + x_{\left(\frac{10}{2}+1\right)}}{2} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{2.13 + 2.47}{2} = 2.3.$$

Exemplo

Exemplo: Considere novamente os dados referentes a trinca de fadiga, fornecidos no primeiro exemplo. Temos que os dados, ordenados, são os seguintes:
$x_{(1)} = 1.15, x_{(2)} = 1.37, x_{(3)} = 1.82, x_{(4)} = 2.04, x_{(5)} = 2.13, x_{(6)} = 2.47, x_{(7)} = 2.60, x_{(8)} = 2.96, x_{(9)} = 3.02.$

Como $n = 9$ (impar), então temos que

$$Md = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{10}{2}\right)} = x_{(5)} = 2.13.$$

Outro exemplo

Reconsidere novamente os dados acima. Suponha que uma outra medida referente a trinca de fadiga foi coletada e o valor dessa observação foi $x_{(10)} = 3.03$.

Note que a mediana desse novo conjunto de dados é

$$Md = \frac{x_{\left(\frac{10}{2}\right)} + x_{\left(\frac{10}{2}+1\right)}}{2} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{2.13 + 2.47}{2} = 2.3.$$

Exemplo

Exemplo: Considere novamente os dados referentes a trinca de fadiga, fornecidos no primeiro exemplo. Temos que os dados, ordenados, são os seguintes:
$x_{(1)} = 1.15, x_{(2)} = 1.37, x_{(3)} = 1.82, x_{(4)} = 2.04, x_{(5)} = 2.13, x_{(6)} = 2.47, x_{(7)} = 2.60, x_{(8)} = 2.96, x_{(9)} = 3.02.$

Como $n = 9$ (impar), então temos que

$$Md = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{10}{2}\right)} = x_{(5)} = 2.13.$$

Outro exemplo

Reconsidere novamente os dados acima. Suponha que uma outra medida referente a trinca de fadiga foi coletada e o valor dessa observação foi $x_{(10)} = 3.03$.

Note que a mediana desse novo conjunto de dados é

$$Md = \frac{x_{\left(\frac{10}{2}\right)} + x_{\left(\frac{10}{2}+1\right)}}{2} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{2.13 + 2.47}{2} = 2.3.$$

Exemplo

Exemplo: Considere novamente os dados referentes a trinca de fadiga, fornecidos no primeiro exemplo. Temos que os dados, ordenados, são os seguintes: $x_{(1)} = 1.15, x_{(2)} = 1.37, x_{(3)} = 1.82, x_{(4)} = 2.04, x_{(5)} = 2.13, x_{(6)} = 2.47, x_{(7)} = 2.60, x_{(8)} = 2.96, x_{(9)} = 3.02$.

Como $n = 9$ (impar), então temos que

$$Md = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{10}{2}\right)} = x_{(5)} = 2.13.$$

Outro exemplo

Reconsidere novamente os dados acima. Suponha que uma outra medida referente a trinca de fadiga foi coletada e o valor dessa observação foi $x_{(10)} = 3.03$.

Note que a mediana desse novo conjunto de dados é

$$Md = \frac{x_{\left(\frac{10}{2}\right)} + x_{\left(\frac{10}{2}+1\right)}}{2} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{2.13 + 2.47}{2} = 2.3.$$

Exemplo

Exemplo: Considere novamente os dados referentes a trinca de fadiga, fornecidos no primeiro exemplo. Temos que os dados, ordenados, são os seguintes: $x_{(1)} = 1.15, x_{(2)} = 1.37, x_{(3)} = 1.82, x_{(4)} = 2.04, x_{(5)} = 2.13, x_{(6)} = 2.47, x_{(7)} = 2.60, x_{(8)} = 2.96, x_{(9)} = 3.02$.

Como $n = 9$ (impar), então temos que

$$Md = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{10}{2}\right)} = x_{(5)} = 2.13.$$

Outro exemplo

Reconsidere novamente os dados acima. Suponha que uma outra medida referente a trinca de fadiga foi coletada e o valor dessa observação foi $x_{(10)} = 3.03$.

Note que a mediana desse novo conjunto de dados é

$$Md = \frac{x_{\left(\frac{10}{2}\right)} + x_{\left(\frac{10}{2}+1\right)}}{2} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{2.13 + 2.47}{2} = 2.3.$$

Amplitude

- Além da variância e do desvio-padrão de uma amostra, a amplitude da amostra é também uma outra medida útil de variabilidade.

Amplitude

Se as n observações de uma amostra forem denotadas por $x_1, x_2, \ldots, x_n$, então a amplitude da amostra é definida por

$$\text{ap} = \max(x_i) - \min(x_i).$$

Ou seja, a amplitude é simplesmente a diferença entre a maior e a menor observação da amostra.

- Por exemplo, temos que a amplitude da amostra do primeiro exemplo é ap$= 3.02 - 1.15 = 1.87$

Amplitude

- Além da variância e do desvio-padrão de uma amostra, a amplitude da amostra é também uma outra medida útil de variabilidade.

Amplitude

Se as n observações de uma amostra forem denotadas por $x_1, x_2, \ldots, x_n$, então a amplitude da amostra é definida por

$$\text{ap} = \max(x_i) - \min(x_i).$$

Ou seja, a amplitude é simplesmente a diferença entre a maior e a menor observação da amostra.

- Por exemplo, temos que a amplitude da amostra do primeiro exemplo é ap$= 3.02 - 1.15 = 1.87$

Amplitude

- Note que a amplitude é uma medida bem simples de ser calculada. Contudo, ela não leva em consideração as informações contidas nos dados entre o maior e o menor valor.
- Por exemplo, considere os dois conjuntos de dados a seguir: $1, 3, 5, 8, 9$ e $1, 5, 5, 5, 9$. Note que a amplitude dos dois conjuntos de dados é a mesma, isto é, ap$= 8$. Contudo, o desvio-padrão da primeira amostra é $s_1 \approx 3.35$ enquanto que o desvio-padrão da segunda amostra é $s_2 \approx 2.83$.

Moda

Uma outra medida de resumo bem simples e útil é a moda da amostra. A moda da amostra é o valor da observação que ocorre com mais frequência.

Por exemplo, no segundo conjunto de dados do exemplo acima temos que 5 é a moda da amostra, uma vez que apareceu com maior frequência no conjunto de dados. Alguns conjuntos de dados podem ter mais de uma moda. Neste caso, dizemos que o conjunto é multimodal.
Pode ocorrer também de não existir moda na amostra, como por exemplo, no primeiro conjunto de dados do exemplo acima.

Amplitude

- Note que a amplitude é uma medida bem simples de ser calculada. Contudo, ela não leva em consideração as informações contidas nos dados entre o maior e o menor valor.
- Por exemplo, considere os dois conjuntos de dados a seguir: $1, 3, 5, 8, 9$ e $1, 5, 5, 5, 9$. Note que a amplitude dos dois conjuntos de dados é a mesma, isto é, ap$= 8$. Contudo, o desvio-padrão da primeira amostra é $s_1 \approx 3.35$ enquanto que o desvio-padrão da segunda amostra é $s_2 \approx 2.83$.

Moda

Uma outra medida de resumo bem simples e útil é a moda da amostra. A moda da amostra é o valor da observação que ocorre com mais frequência.

Por exemplo, no segundo conjunto de dados do exemplo acima temos que 5 é a moda da amostra, uma vez que apareceu com maior frequência no conjunto de dados. Alguns conjuntos de dados podem ter mais de uma moda. Neste caso, dizemos que o conjunto é multimodal.
Pode ocorrer também de não existir moda na amostra, como por exemplo, no primeiro conjunto de dados do exemplo acima.

Amplitude

- Note que a amplitude é uma medida bem simples de ser calculada. Contudo, ela não leva em consideração as informações contidas nos dados entre o maior e o menor valor.
- Por exemplo, considere os dois conjuntos de dados a seguir: $1, 3, 5, 8, 9$ e $1, 5, 5, 5, 9$. Note que a amplitude dos dois conjuntos de dados é a mesma, isto é, ap$= 8$. Contudo, o desvio-padrão da primeira amostra é $s_1 \approx 3.35$ enquanto que o desvio-padrão da segunda amostra é $s_2 \approx 2.83$.

Moda

Uma outra medida de resumo bem simples e útil é a moda da amostra. A moda da amostra é o valor da observação que ocorre com mais frequência.

Por exemplo, no segundo conjunto de dados do exemplo acima temos que 5 é a moda da amostra, uma vez que apareceu com maior frequência no conjunto de dados. Alguns conjuntos de dados podem ter mais de uma moda. Neste caso, dizemos que o conjunto é multimodal.
Pode ocorrer também de não existir moda na amostra, como por exemplo, no primeiro conjunto de dados do exemplo acima.

Amplitude

- Note que a amplitude é uma medida bem simples de ser calculada. Contudo, ela não leva em consideração as informações contidas nos dados entre o maior e o menor valor.
- Por exemplo, considere os dois conjuntos de dados a seguir: $1, 3, 5, 8, 9$ e $1, 5, 5, 5, 9$. Note que a amplitude dos dois conjuntos de dados é a mesma, isto é, ap$= 8$. Contudo, o desvio-padrão da primeira amostra é $s_1 \approx 3.35$ enquanto que o desvio-padrão da segunda amostra é $s_2 \approx 2.83$.

Moda

Uma outra medida de resumo bem simples e útil é a moda da amostra. A moda da amostra é o valor da observação que ocorre com mais frequência.

Por exemplo, no segundo conjunto de dados do exemplo acima temos que 5 é a moda da amostra, uma vez que apareceu com maior frequência no conjunto de dados. Alguns conjuntos de dados podem ter mais de uma moda. Neste caso, dizemos que o conjunto é multimodal.
Pode ocorrer também de não existir moda na amostra, como por exemplo, no primeiro conjunto de dados do exemplo acima.

Amplitude

- Note que a amplitude é uma medida bem simples de ser calculada. Contudo, ela não leva em consideração as informações contidas nos dados entre o maior e o menor valor.
- Por exemplo, considere os dois conjuntos de dados a seguir: $1, 3, 5, 8, 9$ e $1, 5, 5, 5, 9$. Note que a amplitude dos dois conjuntos de dados é a mesma, isto é, ap$= 8$. Contudo, o desvio-padrão da primeira amostra é $s_1 \approx 3.35$ enquanto que o desvio-padrão da segunda amostra é $s_2 \approx 2.83$.

Moda

Uma outra medida de resumo bem simples e útil é a moda da amostra. A moda da amostra é o valor da observação que ocorre com mais frequência.

Por exemplo, no segundo conjunto de dados do exemplo acima temos que 5 é a moda da amostra, uma vez que apareceu com maior frequência no conjunto de dados. Alguns conjuntos de dados podem ter mais de uma moda. Neste caso, dizemos que o conjunto é multimodal.

Pode ocorrer também de não existir moda na amostra, como por exemplo, no primeiro conjunto de dados do exemplo acima.

Amplitude

- Note que a amplitude é uma medida bem simples de ser calculada. Contudo, ela não leva em consideração as informações contidas nos dados entre o maior e o menor valor.
- Por exemplo, considere os dois conjuntos de dados a seguir: $1, 3, 5, 8, 9$ e $1, 5, 5, 5, 9$. Note que a amplitude dos dois conjuntos de dados é a mesma, isto é, ap$= 8$. Contudo, o desvio-padrão da primeira amostra é $s_1 \approx 3.35$ enquanto que o desvio-padrão da segunda amostra é $s_2 \approx 2.83$.

Moda

Uma outra medida de resumo bem simples e útil é a moda da amostra. A moda da amostra é o valor da observação que ocorre com mais frequência.

Por exemplo, no segundo conjunto de dados do exemplo acima temos que 5 é a moda da amostra, uma vez que apareceu com maior frequência no conjunto de dados. Alguns conjuntos de dados podem ter mais de uma moda. Neste caso, dizemos que o conjunto é multimodal.

Pode ocorrer também de não existir moda na amostra, como por exemplo, no primeiro conjunto de dados do exemplo acima.

Amplitude

- Note que a amplitude é uma medida bem simples de ser calculada. Contudo, ela não leva em consideração as informações contidas nos dados entre o maior e o menor valor.
- Por exemplo, considere os dois conjuntos de dados a seguir: $1, 3, 5, 8, 9$ e $1, 5, 5, 5, 9$. Note que a amplitude dos dois conjuntos de dados é a mesma, isto é, ap$= 8$. Contudo, o desvio-padrão da primeira amostra é $s_1 \approx 3.35$ enquanto que o desvio-padrão da segunda amostra é $s_2 \approx 2.83$.

Moda

Uma outra medida de resumo bem simples e útil é a moda da amostra. A moda da amostra é o valor da observação que ocorre com mais frequência.

Por exemplo, no segundo conjunto de dados do exemplo acima temos que 5 é a moda da amostra, uma vez que apareceu com maior frequência no conjunto de dados. Alguns conjuntos de dados podem ter mais de uma moda. Neste caso, dizemos que o conjunto é multimodal.

Pode ocorrer também de não existir moda na amostra, como por exemplo, no primeiro conjunto de dados do exemplo acima.

Amplitude

- Note que a amplitude é uma medida bem simples de ser calculada. Contudo, ela não leva em consideração as informações contidas nos dados entre o maior e o menor valor.
- Por exemplo, considere os dois conjuntos de dados a seguir: $1, 3, 5, 8, 9$ e $1, 5, 5, 5, 9$. Note que a amplitude dos dois conjuntos de dados é a mesma, isto é, ap$= 8$. Contudo, o desvio-padrão da primeira amostra é $s_1 \approx 3.35$ enquanto que o desvio-padrão da segunda amostra é $s_2 \approx 2.83$.

Moda

Uma outra medida de resumo bem simples e útil é a moda da amostra. A moda da amostra é o valor da observação que ocorre com mais frequência.

Por exemplo, no segundo conjunto de dados do exemplo acima temos que 5 é a moda da amostra, uma vez que apareceu com maior frequência no conjunto de dados. Alguns conjuntos de dados podem ter mais de uma moda. Neste caso, dizemos que o conjunto é multimodal.

Pode ocorrer também de não existir moda na amostra, como por exemplo, no primeiro conjunto de dados do exemplo acima.

Amplitude

- Note que a amplitude é uma medida bem simples de ser calculada. Contudo, ela não leva em consideração as informações contidas nos dados entre o maior e o menor valor.
- Por exemplo, considere os dois conjuntos de dados a seguir: $1, 3, 5, 8, 9$ e $1, 5, 5, 5, 9$. Note que a amplitude dos dois conjuntos de dados é a mesma, isto é, ap$= 8$. Contudo, o desvio-padrão da primeira amostra é $s_1 \approx 3.35$ enquanto que o desvio-padrão da segunda amostra é $s_2 \approx 2.83$.

Moda

Uma outra medida de resumo bem simples e útil é a moda da amostra. A moda da amostra é o valor da observação que ocorre com mais frequência.

Por exemplo, no segundo conjunto de dados do exemplo acima temos que 5 é a moda da amostra, uma vez que apareceu com maior frequência no conjunto de dados. Alguns conjuntos de dados podem ter mais de uma moda. Neste caso, dizemos que o conjunto é multimodal.
Pode ocorrer também de não existir moda na amostra, como por exemplo, no primeiro conjunto de dados do exemplo acima.

Quartis

- Note que tanto a média como a variância são afetadas por valores extremos da amostra.
- Além disso, somente com essas medidas não tem como termos ideia da simetria ou assimetria da distribuição dos dados.
- Vimos que a mediana divide a amostra ordenada em duas partes iguais. Veremos que, por meio dos quartis, podemos dividir a amostra ordenada em quatro partes iguais. Os pontos de divisão são chamados de quartis.

Quartis

Seja $x_1, x_2, \ldots, x_n$ uma amostra e $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ a ordenação da amostra. Então

(i) O primeiro quartil, denotado por q_1, é um valor que tem aproximadamente 25% da amostra abaixo dele e aproximadamente 75% acima dele.

(ii) O segundo quartil, denotado por q_2, é um valor que tem aproximadamente 50% da amostra abaixo dele. Ou seja, q_2 é a mediana.

(iii) O terceiro quartil, denotado por q_3, é um valor que tem aproximadamente 75% da amostra abaixo dele.

Quartis

- Note que tanto a média como a variância são afetadas por valores extremos da amostra.
- Além disso, somente com essas medidas não tem como termos ideia da simetria ou assimetria da distribuição dos dados.
- Vimos que a mediana divide a amostra ordenada em duas partes iguais. Veremos que, por meio dos quartis, podemos dividir a amostra ordenada em quatro partes iguais. Os pontos de divisão são chamados de quartis.

Quartis

Seja $x_1, x_2, \ldots, x_n$ uma amostra e $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ a ordenação da amostra. Então

(i) O primeiro quartil, denotado por q_1, é um valor que tem aproximadamente 25% da amostra abaixo dele e aproximadamente 75% acima dele.

(ii) O segundo quartil, denotado por q_2, é um valor que tem aproximadamente 50% da amostra abaixo dele. Ou seja, q_2 é a mediana.

(iii) O terceiro quartil, denotado por q_3, é um valor que tem aproximadamente 75% da amostra abaixo dele.

Quartis

- Note que tanto a média como a variância são afetadas por valores extremos da amostra.
- Além disso, somente com essas medidas não tem como termos ideia da simetria ou assimetria da distribuição dos dados.
- Vimos que a mediana divide a amostra ordenada em duas partes iguais. Veremos que, por meio dos quartis, podemos dividir a amostra ordenada em quatro partes iguais. Os pontos de divisão são chamados de quartis.

Quartis

Seja $x_1, x_2, \ldots, x_n$ uma amostra e $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ a ordenação da amostra. Então

(i) O primeiro quartil, denotado por q_1, é um valor que tem aproximadamente 25% da amostra abaixo dele e aproximadamente 75% acima dele.

(ii) O segundo quartil, denotado por q_2, é um valor que tem aproximadamente 50% da amostra abaixo dele. Ou seja, q_2 é a mediana.

(iii) O terceiro quartil, denotado por q_3, é um valor que tem aproximadamente 75% da amostra abaixo dele.

Quartis

- Note que tanto a média como a variância são afetadas por valores extremos da amostra.
- Além disso, somente com essas medidas não tem como termos ideia da simetria ou assimetria da distribuição dos dados.
- Vimos que a mediana divide a amostra ordenada em duas partes iguais. Veremos que, por meio dos quartis, podemos dividir a amostra ordenada em quatro partes iguais. Os pontos de divisão são chamados de quartis.

Quartis

Seja $x_1, x_2, \ldots, x_n$ uma amostra e $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ a ordenação da amostra. Então

(i) O primeiro quartil, denotado por q_1, é um valor que tem aproximadamente 25% da amostra abaixo dele e aproximadamente 75% acima dele.

(ii) O segundo quartil, denotado por q_2, é um valor que tem aproximadamente 50% da amostra abaixo dele. Ou seja, q_2 é a mediana.

(iii) O terceiro quartil, denotado por q_3, é um valor que tem aproximadamente 75% da amostra abaixo dele.

Quartis

- Note que tanto a média como a variância são afetadas por valores extremos da amostra.
- Além disso, somente com essas medidas não tem como termos ideia da simetria ou assimetria da distribuição dos dados.
- Vimos que a mediana divide a amostra ordenada em duas partes iguais. Veremos que, por meio dos quartis, podemos dividir a amostra ordenada em quatro partes iguais. Os pontos de divisão são chamados de quartis.

Quartis

Seja $x_1, x_2, \ldots, x_n$ uma amostra e $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ a ordenação da amostra. Então

(i) O primeiro quartil, denotado por q_1, é um valor que tem aproximadamente 25% da amostra abaixo dele e aproximadamente 75% acima dele.

(ii) O segundo quartil, denotado por q_2, é um valor que tem aproximadamente 50% da amostra abaixo dele. Ou seja, q_2 é a mediana.

(iii) O terceiro quartil, denotado por q_3, é um valor que tem aproximadamente 75% da amostra abaixo dele.

Quartis

- Note que tanto a média como a variância são afetadas por valores extremos da amostra.
- Além disso, somente com essas medidas não tem como termos ideia da simetria ou assimetria da distribuição dos dados.
- Vimos que a mediana divide a amostra ordenada em duas partes iguais. Veremos que, por meio dos quartis, podemos dividir a amostra ordenada em quatro partes iguais. Os pontos de divisão são chamados de quartis.

Quartis

Seja $x_1, x_2, \ldots, x_n$ uma amostra e $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ a ordenação da amostra. Então

(i) O primeiro quartil, denotado por q_1, é um valor que tem aproximadamente 25% da amostra abaixo dele e aproximadamente 75% acima dele.

(ii) O segundo quartil, denotado por q_2, é um valor que tem aproximadamente 50% da amostra abaixo dele. Ou seja, q_2 é a mediana.

(iii) O terceiro quartil, denotado por q_3, é um valor que tem aproximadamente 75% da amostra abaixo dele.

Quartis

- Note que tanto a média como a variância são afetadas por valores extremos da amostra.
- Além disso, somente com essas medidas não tem como termos ideia da simetria ou assimetria da distribuição dos dados.
- Vimos que a mediana divide a amostra ordenada em duas partes iguais. Veremos que, por meio dos quartis, podemos dividir a amostra ordenada em quatro partes iguais. Os pontos de divisão são chamados de quartis.

Quartis

Seja $x_1, x_2, \ldots, x_n$ uma amostra e $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ a ordenação da amostra. Então

(i) O primeiro quartil, denotado por q_1, é um valor que tem aproximadamente 25% da amostra abaixo dele e aproximadamente 75% acima dele.

(ii) O segundo quartil, denotado por q_2, é um valor que tem aproximadamente 50% da amostra abaixo dele. Ou seja, q_2 é a mediana.

(iii) O terceiro quartil, denotado por q_3, é um valor que tem aproximadamente 75% da amostra abaixo dele.

Exemplo

Exemplo: Considere novamente os dados referentes a trinca de fadiga, fornecidos no primeiro exemplo. Vimos que a mediana dos dados é $x_{(5)} = 2.13$. Logo, $q_2 = x_{(5)} = 2.13$.

$$1.15, 1.37, 1.82, 2.04, 2.13, 2.47, 2.60, 2.96, 3.02.$$

Uma opção possível para obter q_1 e q_3 é calcular a mediana dos primeiros quatro valores e dos últimos quarto valores, respectivamente. Ou seja

$$q_1 = 1.82.$$

e

$$q_3 = 2.6.$$

Exemplo

Exemplo: Considere novamente os dados referentes a trinca de fadiga, fornecidos no primeiro exemplo. Vimos que a mediana dos dados é $x_{(5)} = 2.13$. Logo, $q_2 = x_{(5)} = 2.13$.

$$1.15, 1.37, 1.82, 2.04, 2.13, 2.47, 2.60, 2.96, 3.02.$$

Uma opção possível para obter q_1 e q_3 é calcular a mediana dos primeiros quatro valores e dos últimos quarto valores, respectivamente. Ou seja

$$q_1 = 1.82.$$

e

$$q_3 = 2.6.$$

Exemplo

Exemplo: Considere novamente os dados referentes a trinca de fadiga, fornecidos no primeiro exemplo. Vimos que a mediana dos dados é $x_{(5)} = 2.13$. Logo, $q_2 = x_{(5)} = 2.13$.

$$1.15, 1.37, 1.82, 2.04, 2.13, 2.47, 2.60, 2.96, 3.02.$$

Uma opção possível para obter q_1 e q_3 é calcular a mediana dos primeiros quatro valores e dos últimos quarto valores, respectivamente. Ou seja

$$q_1 = 1.82.$$

e

$$q_3 = 2.6.$$

Exemplo

Exemplo: Considere novamente os dados referentes a trinca de fadiga, fornecidos no primeiro exemplo. Vimos que a mediana dos dados é $x_{(5)} = 2.13$. Logo, $q_2 = x_{(5)} = 2.13$.

$$1.15, 1.37, 1.82, 2.04, 2.13, 2.47, 2.60, 2.96, 3.02.$$

Uma opção possível para obter q_1 e q_3 é calcular a mediana dos primeiros quatro valores e dos últimos quarto valores, respectivamente. Ou seja

$$q_1 = 1.82.$$

e

$$q_3 = 2.6.$$

Quartis

Os cinco valores $x_{(1)}$, q_1, q_2, q_3 e $x_{(n)}$ são importantes para se ter uma boa ideia a respeito da simetria da distribuição dos dados. Para termos uma distribuição simétrica, ou aproximadamente simétrica, devemos ter:

(i) $q_2 - x_{(1)} \approx x_{(n)} - q_2$;

(ii) $q_2 - q_1 \approx q_3 - q_2$;

(iii) $q_1 - x_{(1)} \approx x_{(n)} - q_3$;

(iv) distâncias entre a mediana e q_1, q_3 menores do que as distâncias entre os extremos e q_1, q_3.

A figura abaixo ilustra esses fatos para a distribuição normal.

Quartis

Os cinco valores $x_{(1)}$, q_1, q_2, q_3 e $x_{(n)}$ são importantes para se ter uma boa ideia a respeito da simetria da distribuição dos dados. Para termos uma distribuição simétrica, ou aproximadamente simétrica, devemos ter:

(i) $q_2 - x_{(1)} \approx x_{(n)} - q_2$;

(ii) $q_2 - q_1 \approx q_3 - q_2$;

(iii) $q_1 - x_{(1)} \approx x_{(n)} - q_3$;

(iv) distâncias entre a mediana e q_1, q_3 menores do que as distâncias entre os extremos e q_1, q_3.

A figura abaixo ilustra esses fatos para a distribuição normal.

Quartis

Os cinco valores $x_{(1)}$, q_1, q_2, q_3 e $x_{(n)}$ são importantes para se ter uma boa ideia a respeito da simetria da distribuição dos dados. Para termos uma distribuição simétrica, ou aproximadamente simétrica, devemos ter:

(i) $q_2 - x_{(1)} \approx x_{(n)} - q_2$;

(ii) $q_2 - q_1 \approx q_3 - q_2$;

(iii) $q_1 - x_{(1)} \approx x_{(n)} - q_3$;

(iv) distâncias entre a mediana e q_1, q_3 menores do que as distâncias entre os extremos e q_1, q_3.

A figura abaixo ilustra esses fatos para a distribuição normal.

Quartis

Os cinco valores $x_{(1)}$, q_1, q_2, q_3 e $x_{(n)}$ são importantes para se ter uma boa ideia a respeito da simetria da distribuição dos dados. Para termos uma distribuição simétrica, ou aproximadamente simétrica, devemos ter:

(i) $q_2 - x_{(1)} \approx x_{(n)} - q_2$;

(ii) $q_2 - q_1 \approx q_3 - q_2$;

(iii) $q_1 - x_{(1)} \approx x_{(n)} - q_3$;

(iv) distâncias entre a mediana e q_1, q_3 menores do que as distâncias entre os extremos e q_1, q_3.

A figura abaixo ilustra esses fatos para a distribuição normal.

Quartis

Os cinco valores $x_{(1)}$, q_1, q_2, q_3 e $x_{(n)}$ são importantes para se ter uma boa ideia a respeito da simetria da distribuição dos dados. Para termos uma distribuição simétrica, ou aproximadamente simétrica, devemos ter:

(i) $q_2 - x_{(1)} \approx x_{(n)} - q_2$;

(ii) $q_2 - q_1 \approx q_3 - q_2$;

(iii) $q_1 - x_{(1)} \approx x_{(n)} - q_3$;

(iv) distâncias entre a mediana e q_1, q_3 menores do que as distâncias entre os extremos e q_1, q_3.

A figura abaixo ilustra esses fatos para a distribuição normal.

Quartis

Os cinco valores $x_{(1)}$, q_1, q_2, q_3 e $x_{(n)}$ são importantes para se ter uma boa ideia a respeito da simetria da distribuição dos dados. Para termos uma distribuição simétrica, ou aproximadamente simétrica, devemos ter:

(i) $q_2 - x_{(1)} \approx x_{(n)} - q_2$;

(ii) $q_2 - q_1 \approx q_3 - q_2$;

(iii) $q_1 - x_{(1)} \approx x_{(n)} - q_3$;

(iv) distâncias entre a mediana e q_1, q_3 menores do que as distâncias entre os extremos e q_1, q_3.

A figura abaixo ilustra esses fatos para a distribuição normal.

Quartis

Os cinco valores $x_{(1)}$, q_1, q_2, q_3 e $x_{(n)}$ são importantes para se ter uma boa ideia a respeito da simetria da distribuição dos dados. Para termos uma distribuição simétrica, ou aproximadamente simétrica, devemos ter:

(i) $q_2 - x_{(1)} \approx x_{(n)} - q_2$;

(ii) $q_2 - q_1 \approx q_3 - q_2$;

(iii) $q_1 - x_{(1)} \approx x_{(n)} - q_3$;

(iv) distâncias entre a mediana e q_1, q_3 menores do que as distâncias entre os extremos e q_1, q_3.

A figura abaixo ilustra esses fatos para a distribuição normal.

Quartis - Distribuição normal

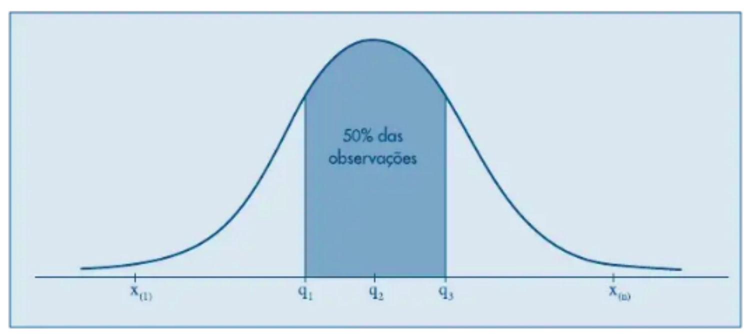

Figure 1: MORETTIN, Pedro Alberto; BUSSAB, WILTON OLIVEIRA. Estatística básica. Saraiva Educação SA, 2017.