Iniciação à Estatística

Prof. Fernando de Souza Bastos
fernando.bastos@ufv.br

Departamento de Estatística
Universidade Federal de Viçosa
Campus UFV - Viçosa

Sumário

Resumir dados por meio de tabelas de frequências, gráficos e diagramas já nos dá uma boa ideia do comportamento de uma variável. Mas o que você acha de irmos além? **E se pudéssemos usar um ou alguns números mágicos para resumir tudo isso?**

Resumir dados por meio de tabelas de frequências, gráficos e diagramas já nos dá uma boa ideia do comportamento de uma variável. Mas o que você acha de irmos além? **E se pudéssemos usar um ou alguns números mágicos para resumir tudo isso?**

Esses números mágicos são chamados de **medidas de posição central**: a **média**, a **mediana** e a **moda**. São como resumos rápidos que ajudam a entender a essência dos dados!

Moda

Moda: é o valor que mais aparece nos dados, o mais "popular"!
Mas, atenção! Podemos ter situações: **amodais** (sem moda), **unimodais** (uma moda), **bimodais** (duas modas) e até **multimodais** (várias modas).
Vamos aos exemplos!

1. Para $X = \{8, 10, 5, 10, 15, 14\}$, qual é a moda? (Pense na popularidade!)

Moda

Moda: é o valor que mais aparece nos dados, o mais "popular"!
Mas, atenção! Podemos ter situações: **amodais** (sem moda), **unimodais** (uma moda), **bimodais** (duas modas) e até **multimodais** (várias modas).
Vamos aos exemplos!

1. Para $X = \{8, 10, 5, 10, 15, 14\}$, qual é a moda? (Pense na popularidade!)
2. Para $X = \{8, 10, 8, 7, 15, 7\}$, temos mais de uma moda! Descubra quais são!

Mediana

Imagine que você está organizando uma fila de números. A **mediana** é o valor bem no meio dessa fila!

- Se temos uma quantidade ímpar de números, a mediana é o número que fica no centro.
- Se temos uma quantidade par, fazemos a média dos dois números centrais.

Exemplo: Se as observações são $3, 4, 7, 8, 8$, a mediana é 7. Adicionando o valor 9, a mediana será $(7 + 8)/2 = 7, 5$.

Média Aritmética

Agora, a queridinha de todos: a **média aritmética**! Ela é simplesmente a soma de todas as observações dividida pelo número total de observações. É como dividir igualmente o valor total entre todos.

Média Aritmética - Um Passo Além!

Se temos n valores de X, a média pode ser calculada assim:

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

Se alguns valores se repetem, podemos agrupar assim:

$$\bar{x} = \frac{n_1 x_1 + \cdots + n_k x_k}{n_1 + \cdots + n_k} = \frac{\sum_{i=1}^{k} n_i x_i}{n} \tag{2}$$

Dica: A média é ótima para dados estáveis, mas... e se tivermos um valor muito extremo?

Média, Mediana e Moda em Ação!

Figura: Visualização Geométrica da Moda, Média e Mediana

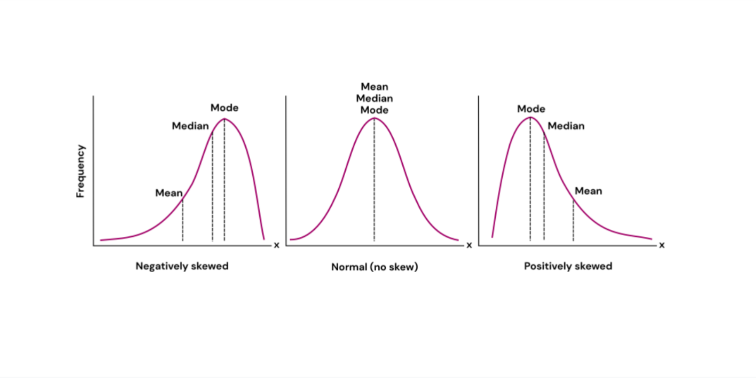
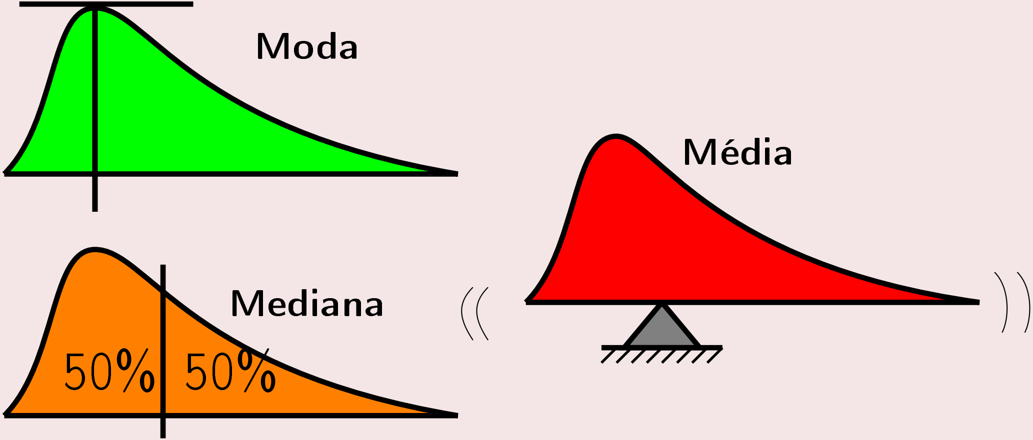


Figura: Visualização Geométrica da moda, média e mediana de uma função densidade de probabilidade arbitrária



Propriedade 1:

Sejam X e Y variáveis aleatórias e k uma constante. Se $X = Y + k$, então $\bar{X} = \bar{Y} + k$

Propriedade 1:

Sejam X e Y variáveis aleatórias e k uma constante. Se $X = Y + k$, então $\bar{X} = \bar{Y} + k$

Demonstração

$$\begin{aligned}
\bar{X} &= \frac{\sum\limits_{i=1}^{n} x_i}{n} \\
&= \frac{x_1 + \cdots + x_n}{n} \\
&= \frac{(y_1 + k) + \cdots + (y_n + k)}{n}
\end{aligned}$$

Propriedade 1:

Sejam X e Y variáveis aleatórias e k uma constante. Se $X = Y + k$, então $\bar{X} = \bar{Y} + k$

Demonstração

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + \cdots + x_n}{n} = \frac{(y_1 + k) + \cdots + (y_n + k)}{n}$$

$$= \frac{(y_1 + \cdots + y_n) + (k + \cdots + k)}{n} = \frac{(y_1 + \cdots + y_n) + nk}{n} = \frac{\sum_{i=1}^{n} y_i}{n} + k = \bar{Y} + k$$

Propriedade 2:

Sejam X e Z variáveis aleatórias e k uma constante. Se $X = kZ$, então $\bar{X} = k\bar{Z}$

Propriedade 2:

Sejam X e Z variáveis aleatórias e k uma constante. Se $X = kZ$, então $\bar{X} = k\bar{Z}$

Demonstração

$$\begin{aligned}
\bar{X} &= \frac{\displaystyle\sum_{i=1}^{n} x_i}{n} \\
&= \frac{x_1 + \cdots + x_n}{n} \\
&= \frac{(kz_1) + \cdots + (kz_n)}{n}
\end{aligned}$$

Propriedade 2:

Sejam X e Z variáveis aleatórias e k uma constante. Se $X = kZ$, então $\bar{X} = k\bar{Z}$

Demonstração

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + \cdots + x_n}{n} = \frac{(kz_1) + \cdots + (kz_n)}{n}$$

$$= \frac{k(z_1 + \cdots + z_n)}{n} = \frac{k\sum_{i=1}^{n} z_i}{n} = k\bar{Z}$$

Propriedade 3:

Seja X uma variável aleatória qualquer. Considere $e_i = x_i - \bar{x}$ o i-ésimo desvio. Então $\sum_{i=1}^{n} e_i = 0$.

Propriedade 3:

Seja X uma variável aleatória qualquer. Considere $e_i = x_i - \bar{x}$ o i-ésimo desvio. Então $\sum_{i=1}^{n} e_i = 0$.

Demonstração

$$\begin{aligned}\sum_{i=1}^{n} e_i &= \sum_{i=1}^{n}(x_i - \bar{x}) \\ &= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} \\ &= \sum_{i=1}^{n} x_i - n\bar{x}\end{aligned}$$

Propriedade 3:

Seja X uma variável aleatória qualquer. Considere $e_i = x_i - \bar{x}$ o i-ésimo desvio. Então $\sum_{i=1}^{n} e_i = 0$.

Demonstração

$$
\begin{aligned}
\sum_{i=1}^{n} e_i &= \sum_{i=1}^{n} (x_i - \bar{x}) \\
&= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} \\
&= \sum_{i=1}^{n} x_i - n\bar{x}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} x_i - n\frac{\sum_{i=1}^{n} x_i}{n} \\
&= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i \\
&= 0
\end{aligned}
$$

Propriedade 1:

Sejam X e Y variáveis aleatórias e k uma constante. Se $X = Y + k$, então $Md(X) = Md(Y) + k$

Propriedade 1:

Sejam X e Y variáveis aleatórias e k uma constante. Se $X = Y + k$, então $Md(X) = Md(Y) + k$

Propriedade 2:

Sejam X e Z variáveis aleatórias e k uma constante. Se $X = kZ$, então $Md(X) = kMd(Z)$

Propriedade 1:

Sejam X e Y variáveis aleatórias e k uma constante. Se $X = Y + k$, então $Mo(X) = Mo(Y) + k$

Propriedade 1:

Sejam X e Y variáveis aleatórias e k uma constante. Se $X = Y + k$, então $Mo(X) = Mo(Y) + k$

Propriedade 2:

Sejam X e Z variáveis aleatórias e k uma constante. Se $X = kZ$, então $Mo(X) = kMo(Z)$

Observação Importante

Lembre-se: **a mediana é mais robusta que a média** em situações com valores muito extremos (outliers). Por isso, sempre avalie qual medida de posição faz mais sentido para o conjunto de dados que você está analisando!

Referências I

- Ferreira, Eric Batista e Marcelo Silva de Oliveira (2020). *Introdução à Estatística com R*. Editora Universidade Federal de Alfenas. URL: https://www.unifal-mg.edu.br/bibliotecas/wp-content/uploads/sites/125/2021/12/32-EBR_Unifal.pdf.
- Morettin, P.A. e W.O Bussab (2009). *Estatística básica*. 6ª ed. São Paulo: Editora Saraiva.
- Regazzi, Adair José, Carlos Henrique Osório Silva, Gerson Rodrigues dos Santos, Paulo César Emiliano e Eduardo Campana Barbosa (s.d.). *Roteiro de Aulas (EST 105)*. Disponível no PVANet - Moodle.