Estatística Básica

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

Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa
Campus UFV - Florestal

Sumário

Vimos que o resumo de dados por meio de tabelas de frequências, ramo-e-folhas e até mesmo através de gráficos fornece muito mais informações sobre o comportamento de uma variável do que a própria tabela original de dados.

Vimos que o resumo de dados por meio de tabelas de frequências, ramo-e-folhas e até mesmo através de gráficos fornece muito mais informações sobre o comportamento de uma variável do que a própria tabela original de dados.

Muitas vezes, queremos resumir ainda mais estes dados, apresentando um ou alguns valores que sejam representativos da série toda. **Quando usamos um só valor, obtemos uma redução drástica dos dados.** Usualmente, emprega-se uma das seguintes medidas de posição (ou localização) central: **média, mediana ou moda.**

Moda

A moda é definida como a realização mais frequente do conjunto de valores observados.

- No exemplo do número de filhos, $Mo = 2$.

Moda

A moda é definida como a realização mais frequente do conjunto de valores observados.

- No exemplo do número de filhos, $Mo = 2$.
- Em alguns casos, pode haver mais de uma moda, ou seja, a distribuição dos valores pode ser bimodal, trimodal etc.

Mediana

A mediana é a realização que ocupa a posição central da série de observações, quando estão ordenadas em ordem crescente. Assim, se as cinco observações de uma variável forem $3, 4, 7, 8$ e 8, a mediana é o valor 7, correspondendo à terceira observação. Quando o número de observações for par, usa-se como mediana a média aritmética das duas observações centrais. Acrescentando-se o valor 9 à série acima, a mediana será $(7 + 8)/2 = 7,5$.

Mediana

Consideremos, agora, as observações ordenadas em ordem crescente. Vamos denotar a menor observação por $x_{(1)}$, a segunda por $x_{(2)}$, e assim por diante, obtendo-se

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n-1)} \leq x_{(n)}. \tag{1}$$

As observações ordenadas como em (1) são chamadas **estatísticas de ordem**. Com esta notação, a mediana da variável X pode ser definida como:

$$md(X) = \begin{cases} X_{\left(\frac{n+1}{2}\right)}, & \text{se} \quad \text{n} \quad \text{é impar,} \\ \dfrac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se} \quad \text{n} \quad \text{é par.} \end{cases}$$

Média Aritmética

Finalmente, a média aritmética, conceito familiar ao leitor, é a soma das observações dividida pelo número de observações.

Média Aritmética

Se $x_1, \cdots, x_n$ são os n valores (distintos ou não) da variável X, a média aritmética, ou simplesmente média, de X pode ser escrita como:

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n} = \frac{\displaystyle\sum_{i=1}^{n} x_i}{n} \tag{2}$$

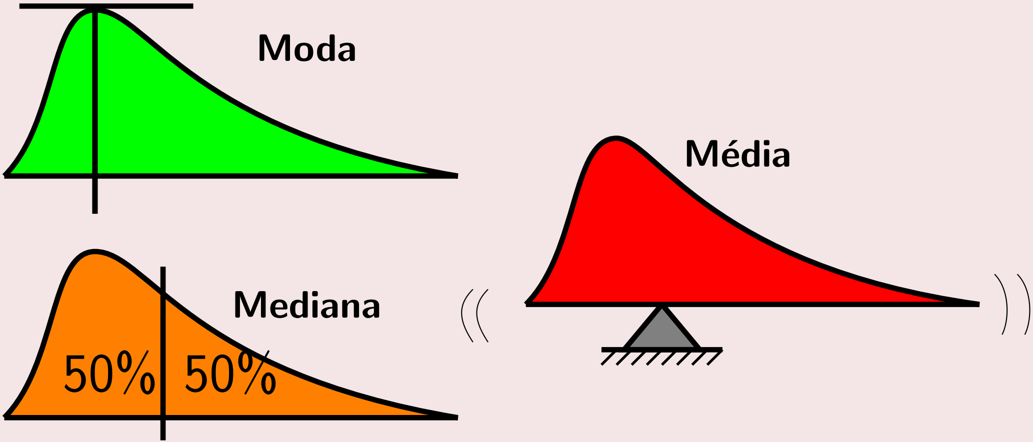
Agora, se tivermos n observações da variável X, das quais n_1 são iguais a x_1, n_2 são iguais a x_2, n_k iguais a x_k, então a média de X pode ser escrita como:

$$\bar{x} = \frac{n_1 x_1 + \cdots + n_k x_k}{n_1 + \cdots + n_k} = \frac{\displaystyle\sum_{i=1}^{k} n_i x_i}{n} \tag{3}$$

Observação importante:

A mediana é uma medida mais robusta que a média, quando submetida a mudanças nos valores observados ou a incorporação de mais observações no conjunto de dados original.

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

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

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

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

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 $Mo(X) = kMo(Z)$

Referências

E. B. Ferreira and M. S. d. Oliveira. *Introdução à Estatística com R*. Editora Universidade Federal de Alfenas, 2020. URL https://www.unifal-mg.edu.br/bibliotecas/system/files/imce/EBR_Unifal.pdf.

P. Morettin and W. Bussab. *Estatística básica*. Editora Saraiva, São Paulo, 6 edition, 2009.