

Universidade Federal de Lavras
Departamento de Estatística
Prof. Daniel Furtado Ferreira

3ª Aula Prática

Medidas de Posição

- 1) Os dados apresentados a seguir referem-se ao tempo que um determinado computador usando o sistema operacional Linux levou para apresentar a primeira falha grave, em anos, obtidos em uma amostra de $n = 30$ computadores realizada na região de Lavras, MG. Os resultados do tempo de falhas em anos são dados por:

8,13	8,23	8,60	8,80	8,97	9,05	9,12	9,30	9,35	9,78
9,80	9,86	9,90	9,95	10,00	10,11	10,13	10,15	10,16	10,23
10,31	10,33	10,40	10,46	10,50	11,14	11,29	11,46	12,05	12,14

- a) Obter a média aritmética \bar{X} .
- b) Calcular a mediana m_d .
- c) Se o tempo de falha de cada computador for multiplicada pela constante 1,086, qual será o valor médio amostral da variável transformada?
- d) Agrupar os dados em uma tabela de distribuição de frequências e estimar a média, a mediana e a moda. Comparar os resultados obtidos a partir dos dados originais com estes, dos dados agrupados.
- e) Obter o valor que deixa 50% dos valores abaixo do mesmo, usando a distribuição de frequência obtida em 1d. Comparar o resultado obtido com a mediana. Qual é a sua conclusão?
- 2) Os dados a seguir referem-se ao número de empresas falidas por ano observadas em uma amostra de $n = 85$ anos obtida em Lavras, MG.

Empresas falidas	Frequências
0	36
1	19
2	16
3	7
4	4
5	2
6	1

Determinar:

- a) A média.
- b) A mediana e a moda.
- c) Qual dessas medidas você considera melhor para representar o número de empresas falidas/ano? Por quê?

Resolução

1) As medidas de posição e as demais quantidades solicitadas a respeito dos dados dos tempos de falhas dos computadores são:

a) A média aritmética é dada por:

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_{30}}{30} = \frac{8,13 + 8,23 + \dots + 12,14}{30} \\ &= \frac{299,7}{30} = 9,99 \text{ anos}\end{aligned}$$

b) Como n é par, a mediana é dada por:

$$\begin{aligned}m_d &= \frac{x_{(\frac{n}{2})} + x_{(\frac{n+2}{2})}}{2} = \frac{x_{(\frac{30}{2})} + x_{(\frac{32}{2})}}{2} = \frac{x_{(15)} + x_{(16)}}{2} \\ &= \frac{10,00 + 10,11}{2} = 10,0550\end{aligned}$$

c) Utilizando as propriedades da média, a nova média \bar{X}^* é dada por:

$$\begin{aligned}\bar{X}^* &= k\bar{X} = 1,086 \times 9,99 = 10,84914 \\ &\cong 2,70\end{aligned}$$

d) Calcular as medidas de posição: média, mediana e moda.

i) Para agrupar os dados deve-se obter:

O número de classe é dado por $k = \sqrt{n} = \sqrt{30} \approx 5$ e amplitude total por $A = x_{(30)} - x_{(1)} = 12,14 - 8,13 = 4,01$. Assim, a amplitude de classe é dada por $c = A/(k-1) = 4,01/4 \approx 1,00$ e o limite inferior da primeira classe por $LI_1 = x_{(1)} - c/2 = 8,13 - 1,00/2 = 7,63$. Os demais limites de classe são obtidos somando-se $c = 1,00$ aos limites anteriormente obtidos. A distribuição de frequências obtida desta forma é:

Classes de tempo	\bar{X}_i	F_i	Fr_i	$Fp_i(\%)$
7,63 ─ 8,63	8,13	3	0,10	10,00
8,63 ─ 9,63	9,13	6	0,20	20,00
9,63 ─ 10,63	10,13	16	0,53	53,33
10,63 ─ 11,63	11,13	3	0,10	10,00
11,63 ─ 12,63	12,13	2	0,07	6,67

ii) A média aritmética é dada por:

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^k F_i \bar{X}_i}{n} = \frac{8,13 \times 3 + 9,13 \times 6 + 10,13 \times 16 + 11,13 \times 3 + 12,13 \times 2}{30} = \frac{298,9}{30} \\ &= 9,9633\end{aligned}$$

A diferença encontrada para a média dos dados não agrupados pode ser atribuída ao agrupamento. Toda forma de representar os dados de uma maneira mais simplificada conduz a algum tipo de perda de precisão. Mas o que deve ficar claro é que apesar de menos precisa, a estimativa obtida a partir dos dados agrupados é uma “estimativa confiável” da média

populacional, tanto quanto a estimativa dos dados originais. A perda de precisão é, em geral, pequena e pode ser considerada desprezível.

- iii) A mediana é obtida da seguinte maneira. A classe mediana é aquela que contém a posição número $n/2 = 30/2 = 15$. Portanto, a classe mediana é a terceira, pois as frequências acumuladas das duas primeiras classes somam apenas 9, que é inferior a 15. Logo,

$$\begin{aligned} m_d &= LI_{m_d} + \frac{\frac{n}{2} - F_A}{F_{m_d}} c_{m_d} = 9,63 + \frac{15 - 9}{16} \times 1,00 \\ &= 10,0050 \end{aligned}$$

A mesma observação feita para a diferença das estimativas da média vale para a mediana.

- iv) Para obter a moda, é necessário determinar a classe de maior frequência, ou seja, a classe modal. A classe modal neste exercício é a terceira. A diferença das frequências da classe modal e classe anterior é $\Delta_1 = 16 - 6 = 10$ e a diferença das frequências da classe modal e classe posterior é $\Delta_2 = 16 - 3 = 13$. Assim, tem-se

$$\begin{aligned} m_o &= LI_{m_o} + \frac{\Delta_1}{\Delta_1 + \Delta_2} c_{m_o} = 9,63 + \frac{10}{10 + 13} \times 1,00 \\ &= 10,0648. \end{aligned}$$

As três medidas, média, mediana e moda, estão muito próximas e isso é um indicativo que a distribuição dos dados deve ser aproximadamente simétrica.

- 2) Para a variável número de empresas falidas por ano tem-se:

- a) A média aritmética:

$$\begin{aligned} \bar{X} &= \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 \dots + x_{85}}{85} = \frac{0 + 0 + \dots + 5 + 6}{85} \\ &= \frac{0 \times 36 + 1 \times 19 + \dots + 6 \times 1}{85} = \frac{104}{85} \\ &= 1,2235 \text{ empresas falidas/ano} \end{aligned}$$

- b) Como $n = 85$ é ímpar, a mediana é obtida por

$$\begin{aligned} m_d &= x_{(\frac{n+1}{2})} = x_{(\frac{86}{2})} = x_{(43)} \\ &= 1, \end{aligned}$$

pois da posição 37 até a posição 45, na amostra ordenada, está o valor 1.

A moda é o valor mais frequente, pois os dados são quantitativos discretos. Assim, a moda é dada por $m_o = 0$, que é o valor que repete mais vezes, ou seja, possui frequência 36 que é a maior de todas.

- c) A distribuição é assimétrica à direita, portanto, a média não é uma boa medida para representar estes dados, uma vez que é influenciada por valores extremos. Assim, pode-se utilizar tanto a mediana, quanto a moda para isso.