

EST 105

INICIAÇÃO À ESTATÍSTICA

ESTATÍSTICA DESCRIPTIVA
Medidas de Dispersão - Resumo

Departamento de Estatística – UFV
Av. Peter Henry Rolfs, s/n
Campus Universitário
36570.977 – Viçosa, MG
<http://www.det.ufv.br/>



Medidas de dispersão

Exemplo introdutório: As temperaturas de 3 cidades amostradas em 5 dias e suas respectivas medidas de posição são apresentadas na Tabela 2 a seguir.

Tabela 2. Temperaturas em graus Celsius ($^{\circ}\text{C}$) em 3 cidades (A, B e C), avaliadas por 5 dias e suas respectivas médias (\bar{X}), medianas (Md_X) e modas (Mo_X).

Cidade	Dia					Medidas de Posição			
	1	2	3	4	5	\bar{X}	Md_X	Mo_X	
Temp. ($^{\circ}\text{C}$)	A	21,5	22,0	23,0	24,0	24,5	23,0	23,0	Amodal
	B	15,0	18,5	23,0	25,5	33,0	23,0	23,0	Amodal
	C	5,0	18,0	23,0	31,0	38,0	23,0	23,0	Amodal

Medidas de dispersão

i. Variância:

- A variância é a principal medida de dispersão. Muito utilizada em metodologias inferenciais tais como os Testes de Hipóteses .
- Esta medida avalia a dispersão do conjunto de valores em relação à média aritmética.

Se X_1, X_2, \dots, X_n são “n” valores quaisquer da variável X (valores indexados), então a variância de X , denotada por s_X^2 , é dada por:

$$s_X^2 = \frac{SQD_X}{n-1} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

SQD_X é a soma de quadrados dos desvios de X

Ou de forma mais simples e equivalente:

$$s_X^2 = \frac{SQD_X}{n-1} = \frac{\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}}{n-1}$$

Dados agrupados

Sejam n valores da variável X dos quais f_1 são iguais a X_1, f_2 são iguais a X_2, \dots, f_k são iguais a X_k . A variância de X , denotada por s_X^2 , é dada por:

$$s_X^2 = \frac{SQD_X}{\sum_{i=1}^k f_i - 1} = \frac{\sum_{i=1}^k f_i X_i^2 - \frac{(\sum_{i=1}^k f_i X_i)^2}{\sum_{i=1}^k f_i}}{\sum_{i=1}^k f_i - 1}$$

em que $\sum_{i=1}^k f_i = n$.

Propriedades da variância

□ **Propriedade 1:** A variância será sempre maior ou igual a zero, isto é, $s_X^2 \geq 0$;

□ **Propriedade 2:** Para $X_i = k$, sendo "k" uma constante qualquer, tem-se que

$$s_X^2 = 0;$$

□ **Propriedade 3:** Somando-se ou subtraindo-se uma constante "k" de cada

observação X_i , a variância não se altera, isto é, $s^2(X \pm k) = s_X^2$;

□ **Propriedade 4:** Multiplicando-se uma constante "k" de cada observação X_i , a

variância ficará multiplicada pelo quadrado da constante, isto é, $s^2(kX) = k^2 s_X^2$.

ii. Desvio-padrão:

O Desvio-padrão mede a dispersão dos dados em torno da média, assim como a variância, porém **apresenta a vantagem de possuir a mesma unidade de medida dos dados.**

O Desvio-padrão, denotado por s_X , é dado por: $s_X = +\sqrt{s_X^2}$

iii. Erro-padrão da Média:

- O Erro-padrão da média, denotado por $s(\bar{X})$, é uma medida utilizada para avaliar a precisão da média.
- Quanto menor o valor de $s(\bar{X})$ maior a precisão da média.

$$s(\bar{X}) = \frac{\sqrt{s_X^2}}{\sqrt{n}} = \frac{s_X}{\sqrt{n}}$$

Obs.: O erro-padrão da média é: (i) Inversamente proporcional ao tamanho da amostra; (ii) Diretamente proporcional à variância.

v. Coeficiente de variação:

O coeficiente de variação (CV) é uma medida de dispersão relativa.

$$CV_X(\%) = \frac{\sqrt{s_X^2}}{\bar{X}} \times 100\% = \frac{s_X}{\bar{X}} \times 100\%$$

- O CV é utilizado, principalmente, para comparar a variabilidade entre conjuntos que possuem médias diferentes.
- A grande vantagem desta medida é que ela é adimensional, permitindo comparações da variabilidade em conjuntos de dados com unidades de medidas diferentes.
- Quanto menor o CV, mais homogênea é a amostra.

iv. Amplitude total:

A amplitude total (A_T) é dada pela diferença entre o maior e o menor valor de um conjunto de dados, ou seja, $AT_X = X_{máx} - X_{min}$.

Exercício resolvido

Exercício 1: Duas amostras aleatórias de lobos Guará, obtidas em duas localidades distintas (Serra do Caraça e Serra do Brigadeiro), foram examinadas por um grupo de médicos veterinários da UFV, com o intuito de se avaliar o número de certo tipo de parasita por animal. Os resultados são apresentados na tabela a seguir:

Localidades										
Número de	Serra do Caraça (X)					Serra do Brigadeiro (Y)				
Parasitas	0	3	2	7	5	0	3	1	6	8
Lobos	5	3	3	3	6	10	3	5	2	1

Pede-se:

- Especifique a variável e a frequência nos dados acima.
- O número médio de parasitas por amostra.
- O número mediano de parasitas por amostra.

- d) O número modal de parasitas por amostra.**
- e) A variância por amostra.**
- f) O desvio-padrão por amostra.**
- g) A amplitude total por amostra.**
- h) Qual amostra tem a sua média estimada com maior precisão?**
- i) Qual amostra é mais homogênea?**