



FUNDAMENTOS DA COMPUTAÇÃO

AULA 2



Prof. Ricardo Alexandre Deckmann Zanardini



CONVERSA INICIAL

Olá! Seja muito bem-vindo à segunda aula de Fundamentos da Computação. Nesta aula estudaremos importantes temas relacionados à estatística. A estatística é um ramo da matemática que tem por objetivo coletar, analisar, interpretar e apresentar informações numéricas ou dados por meio de métodos estatísticos que são formas organizadas de tratamento de dados.

Quando estamos trabalhando com estas informações, temos dois importantes conjuntos: a população e a amostra. A população é o conjunto de todos elementos a serem observados enquanto que a amostra é um subconjunto da população.

Em muitos casos, é comum precisarmos de valores para termos uma noção do comportamento destes dados. Sendo assim, é comum utilizarmos valores que descrevem o centro destes dados. Estes valores são chamados de medidas de tendência central ou também chamados de medidas de posição. Os mais utilizados são a média, a moda e a mediana. Temos também as medidas de dispersão, dentre elas a variância e o desvio padrão. Por meio das medidas de dispersão, podemos saber se os dados estão próximos ou não da média. Podemos também saber qual é a amplitude da faixa em torno da média que contém a maior parte dos dados.

TEMA 1 – MÉDIA

A média \bar{X} é um número que representa a tendência central de um conjunto de dados e quando os dados não estão agrupados é calculada pela soma dos valores de um conjunto de dados x_i dividida pelo total de elementos:

$$\bar{X} = \frac{\sum x_i}{n}$$

Exemplo:

Os dados a seguir se referem aos preços de um determinado disco rígido externo de 1TB encontrado em diferentes estabelecimentos comerciais.

R\$ 439,90, R\$ 478,00, R\$ 399,90, R\$ 512,90, R\$ 419,90,

R\$ 439,90, R\$ 439,90, R\$ 489,00, R\$ 459,00, R\$ 499,90



Calcule o preço médio do disco rígido pesquisado.

Para calcularmos a média, precisamos somar os valores e dividir o resultado pelo total de elementos:

$$\bar{X} = \frac{\sum x_i}{n}$$

$$\bar{X} = \frac{439,9 + 478 + 399,9 + 512,9 + 419,9 + 439,9 + 439,9 + 489 + 459 + 499,9}{10}$$

$$\bar{X} = \frac{4578,3}{10}$$

$$\bar{X} = 457,83$$

Assim, o preço médio corresponde a R\$ 457,83. É um valor que dá uma noção dos preços praticados.

Quando um conjunto de dados possui valores repetidos, é comum que seja feita uma representação simplificada dos dados por meio de uma tabela onde ao lado de cada dado é informado o número de repetições (frequência) deste dado. A esta forma de organização, dá-se o nome de distribuição de frequência.

Neste caso, a média é calculada multiplicando cada dado pela respectiva frequência, somando estes resultados e dividindo pelo número total de dados que corresponde à soma das frequências:

$$\bar{X} = \frac{\sum x_i \cdot f_i}{n}$$

Esta média é conhecida como média ponderada, pois leva em consideração diferentes frequências, podendo também serem chamadas de pesos, para cada dado.

Exemplo:

Uma empresa de comércio eletrônico está fazendo um estudo em relação aos valores dos fretes de suas mercadorias. Na tabela a seguir são apresentados os valores dos fretes e as respectivas quantidades de envios feitos a cada dia.



Frete (x_i)	Quantidade (f_i)
R\$ 12,90	5
R\$ 17,50	12
R\$ 19,80	22
R\$ 19,90	45
R\$ 21,10	39
R\$ 23,40	23
R\$ 27,70	31
R\$ 32,30	18
R\$ 39,90	7
R\$ 45,10	11

A partir destas informações, qual é o valor médio do frete das mercadorias da empresa?

Inicialmente, vamos multiplicar o valor de cada frete pelas respectivas quantidades. Depois, precisamos somar os resultados:

Frete (x_i)	Quantidade (f_i)	$x_i \cdot f_i$
R\$ 12,90	5	R\$ 64,50
R\$ 17,50	12	R\$ 210,00
R\$ 19,80	22	R\$ 435,60



R\$ 19,90	45	R\$ 895,50
R\$ 21,10	39	R\$ 822,90
R\$ 23,40	23	R\$ 538,20
R\$ 27,70	31	R\$ 858,70
R\$ 32,30	18	R\$ 581,40
R\$ 39,90	7	R\$ 279,30
R\$ 45,10	11	R\$ 496,10
Total:	213	R\$ 5182,20

Basta agora substituímos os resultados obtidos na fórmula da média e efetuarmos os cálculos necessários:

$$\bar{X} = \frac{\sum x_i \cdot f_i}{n}$$

$$\bar{X} = \frac{5182,2}{213}$$

$$\bar{X} = 24,33$$

Portanto, o valor médio do frete das mercadorias da empresa é igual a R\$ 24,33.

Quando os dados estão organizados em uma distribuição de **frequência por classe, ou seja, quando temos intervalos contendo os dados e a respectiva quantidade de dados em cada um destes intervalos**, sabemos quantos dados temos em cada classe, mas não sabemos quais são os valores destes dados. Assim, precisamos calcular o ponto médio de cada intervalo para que tenhamos um valor que representa os dados de cada intervalo. **A média, neste caso, é o somatório da multiplicação de cada ponto médio pela respectiva frequência dividido pela quantidade total de dados:**

$$\bar{X} = \frac{\sum p_{m_i} \cdot f_i}{n}$$

Exemplo:

Uma empresa de tecnologia desenvolveu um aplicativo de comunicação entre os funcionários de uma indústria e está monitorando o tempo diário em horas de uso deste aplicativo. A seguir são apresentados os resultados diários obtidos.

Tempo	Quantidade
0 - 1	12
1 - 2	21
2 - 3	39
3 - 4	15
4 - 5	22

Qual o tempo médio de permanência de cada funcionário no aplicativo?

Vamos calcular o ponto médio de cada intervalo somando os limites de cada intervalo e dividindo o resultado por 2. Por exemplo, o ponto médio do primeiro intervalo é igual a 0+1 dividido por 2, o que resulta em 1/2 que é igual a 0,5. O ponto médio do segundo intervalo é dado por 1+2 dividido por 2, ou seja, 3/2 = 1,5. Para obtermos os demais pontos médios, seguimos a mesma ideia. Em seguida, multiplicamos os pontos médios pelas respectivas frequências e somamos os resultados:

Tempo	Quantidade	p_{m_i}	$p_{m_i} \cdot f_i$
0 - 1	12	0,5	6
1 - 2	21	1,5	31,5



2 f 3	39	2,5	97,5
3 f 4	15	3,5	52,5
4 f 5	22	4,5	99
Total	109	---	286,5

A média corresponde a:

$$\bar{X} = \frac{\sum p_{m_i} \cdot f_i}{n}$$

$$\bar{X} = \frac{286,5}{109}$$

$$\bar{X} = 2,63$$

A média referente ao tempo diário de uso do aplicativo é de 2,63 horas.

TEMA 2 – MODA

Outra importante medida de posição é a moda, indicada por Mo. A moda está associada ao dado que aparece o maior número de vezes em um determinado conjunto. Podemos ter um ou mais valores relacionados à moda. Caso não existam dados que se repetem, não há moda e neste caso o conjunto é chamado de amodal.

Exemplo:

Os dados a seguir se referem aos preços de um determinado disco rígido externo de 1TB encontrado em diferentes estabelecimentos comerciais.

R\$ 439,90, R\$ 478,00, R\$ 399,90, R\$ 512,90, R\$ 419,90,

R\$ 439,90, R\$ 439,90, R\$ 489,00, R\$ 459,00, R\$ 499,90

Obtenha a moda do preço do disco rígido pesquisado.



Para facilitar o processo de identificação do dado que aparece o maior número de vezes, vamos colocar estes dados em ordem (rol):

R\$ 399,90, R\$ 419,90, **R\$ 439,90, R\$ 439,90, R\$ 439,90,**

R\$ 459,00, R\$ 478,00, R\$ 489,00, R\$ 499,90, R\$ 512,90

Como o valor que aparece o maior número de vezes é R\$ 439,90, temos que $M_o = \text{R\$ } 439,90$.

Exemplo:

Uma empresa de comércio eletrônico está fazendo um estudo em relação aos valores dos fretes de suas mercadorias. Na tabela a seguir são apresentados os valores dos fretes e as respectivas quantidades de envios feitos a cada dia.

Frete (x_i)	Quantidade (f_i)
R\$ 12,90	5
R\$ 17,50	12
R\$ 19,80	22
R\$ 19,90	45
R\$ 21,10	39
R\$ 23,40	23
R\$ 27,70	31
R\$ 32,30	18
R\$ 39,90	7
R\$ 45,10	11



A partir destas informações, qual é a moda do frete das mercadorias da empresa?

Para identificarmos a moda, precisamos verificar qual dado possui a maior frequência.

Frete (x_i)	Quantidade (f_i)
R\$ 12,90	5
R\$ 17,50	12
R\$ 19,80	22
R\$ 19,90	45
R\$ 21,10	39
R\$ 23,40	23
R\$ 27,70	31
R\$ 32,30	18
R\$ 39,90	7
R\$ 45,10	11

Neste caso, o valor R\$ 19,90 tem a maior frequência. Portanto, $M_o = \text{R\$ } 19,90$.

Quando os dados estão em uma distribuição de frequência por classe, a moda é obtida por meio da fórmula



$$Mo = L_i + \frac{f_{post} \cdot A}{f_{ant} + f_{post}}$$

Esta fórmula leva em consideração o limite inferior da classe que contém a moda (L_i), a frequência da classe posterior à que contém a moda (f_{post}), a amplitude do intervalo da classe que contém a moda (A) e a frequência da classe anterior à classe que contém a moda (f_{ant}).

Exemplo:

Uma empresa de tecnologia desenvolveu um aplicativo de comunicação entre os funcionários de uma indústria e está monitorando o tempo diário em horas de uso deste aplicativo. A seguir são apresentados os resultados diários obtidos.

Tempo	Quantidade
0 - 1	12
1 - 2	21
2 - 3	39
3 - 4	15
4 - 5	22

Qual é a moda relacionada ao tempo de permanência de cada funcionário no aplicativo?

Precisamos inicialmente identificar qual é a classe que contém a moda. Para isto, precisamos considerar a maior frequência. No intervalo 2 - 3, temos a maior frequência que corresponde a 39.

Agora que sabemos qual é a classe que contém a moda, vamos utilizar a fórmula

$$Mo = L_i + \frac{f_{post} \cdot A}{f_{ant} + f_{post}}$$

onde:

$$L_i = 2$$

$$f_{ant} = 21$$

$$f_{post} = 15$$

$$A = 1$$

Substituindo os valores na fórmula da moda, temos

$$Mo = L_i + \frac{f_{post} \cdot A}{f_{ant} + f_{post}}$$

$$Mo = 2 + \frac{15 \cdot 1}{21 + 15}$$

$$Mo = 2 + \frac{15}{36}$$

$$Mo = 2 + 0,416667$$

$$Mo = 2,42$$

A moda referente a este problema corresponde a 2,42 horas.

TEMA 3 – MEDIANA

A mediana (M_d) está associada o elemento que ocupa a posição central de um rol, ou seja, o elemento que está no centro de um conjunto de dados ordenados. Se o total de dados é um número ímpar, temos um único elemento ocupando a posição central. Quando temos um número par de dados, dois elementos ocupam a posição central do rol e a mediana é obtida a partir da média destes dois elementos.

Sabemos que aproximadamente metade dos dados está abaixo da mediana e aproximadamente metade dos dados está acima da mediana. Logo, a mediana divide um conjunto ordenado de dados em duas partes iguais.

Exemplo:

Os dados a seguir se referem aos preços de um determinado disco rígido externo de 1TB encontrado em diferentes estabelecimentos comerciais.

R\$ 439,90, R\$ 478,00, R\$ 399,90, R\$ 512,90, R\$ 419,90,

R\$ 439,90, R\$ 439,90, R\$ 489,00, R\$ 459,00, R\$ 499,90

Obtenha a mediana do preço do disco rígido pesquisado.



Organizando os dados, temos:

R\$ 399,90, R\$ 419,90, R\$ 439,90, R\$ 439,90, R\$ **439,90**,

R\$ **459,00**, R\$ 478,00, R\$ 489,00, R\$ 499,90, R\$ 512,90

Para identificarmos qual é a posição central, precisamos agora dividir o total de elementos por dois

$$\frac{10}{2} = 5$$

Como o número de elementos é par e os valores R\$ 439,90 e 459,00 estão na posição central (5° posição e 6° posição do rol), a mediana é dada por

$$Md = \frac{439,90 + 459,00}{2}$$

$$Md = \frac{898,90}{2}$$

$$Md = 449,45$$

Exemplo:

Uma empresa de comércio eletrônico está fazendo um estudo em relação aos valores dos fretes de suas mercadorias. Na tabela a seguir são apresentados os valores dos fretes e as respectivas quantidades de envios feitos a cada dia.

Frete (x_i)	Quantidade (f_i)
R\$ 12,90	5
R\$ 17,50	12
R\$ 19,80	22



R\$ 19,90	45
R\$ 21,10	39
R\$ 23,40	23
R\$ 27,70	31
R\$ 32,30	18
R\$ 39,90	7
R\$ 45,10	11

A partir destas informações, qual é a mediana do frete das mercadorias da empresa?

Para o cálculo da mediana referente a uma distribuição de frequência, precisamos obter os valores associados à frequência acumulada (f_a), ou seja à soma das frequências anteriores a cada uma das frequências.

Frete (x_i)	Quantidade (f_i)	f_a
R\$ 12,90	5	5
R\$ 17,50	12	17
R\$ 19,80	22	39
R\$ 19,90	45	84
R\$ 21,10	39	123
R\$ 23,40	23	146
R\$ 27,70	31	177



R\$ 32,30	18	192
R\$ 39,90	7	202
R\$ 45,10	11	213

Para obtermos a mediana dos dados, vamos dividir o total de elementos por 2, ou seja, $213/2=106,5$. Arredondando para o próximo inteiro, temos que a 107ª posição é ocupada pelo elemento central. Na coluna da frequência acumulada, precisamos identificar qual é o valor relacionado à posição de número 107. Como o número 107 está entre o número 84 e o número 123, na linha do número 123, temos o valor R\$ 21,10. Logo, $Md = R\$ 21,10$.

Quando os dados organizados em distribuição de frequência por classe, a fórmula destinada ao cálculo da mediana é:

$$Md = L_i + \frac{(n/2 - \sum f_{ant})}{f_{Md}} \cdot A$$

onde

L_i é o limite inferior da classe que contém a mediana

n é o número de elementos

$\sum f_{ant}$ é o somatório da frequência da classe anterior à classe que contém a mediana

f_{md} é a frequência da classe que contém a mediana

A é a amplitude do intervalo da classe que contém a mediana

Exemplo:

Uma empresa de tecnologia desenvolveu um aplicativo de comunicação entre os funcionários de uma indústria e está monitorando o tempo diário em horas de uso deste aplicativo. A seguir são apresentados os resultados diários obtidos.

Tempo	Quantidade
-------	------------



0 ┆ 1	12
1 ┆ 2	21
2 ┆ 3	39
3 ┆ 4	15
4 ┆ 5	22

Qual é a mediana referente ao tempo de permanência de cada funcionário no aplicativo?

O primeiro passo é adicionarmos uma coluna na tabela contendo a frequência acumulada.

Tempo	Quantidade	f _a
0 ┆ 1	12	12
1 ┆ 2	21	33
2 ┆ 3	39	72
3 ┆ 4	15	87
4 ┆ 5	22	109

Precisamos dividir o total de elementos por 2, ou seja,

$$\frac{109}{2} = 54,5$$

Como temos um número ímpar de elementos, vamos arredondar o resultado para o próximo inteiro. Assim, a mediana ocupa a 55ª posição. Como 55 está entre 33 e 72, a mediana está no intervalo 2 ┆ 3.

Logo:

$$L_i = 2$$

$$n = 109$$



$$\Sigma f_{ant} = 33$$

$$f_{md} = 39$$

$$A = 1$$

Substituindo os valores na fórmula

$$Md = L_i + \frac{(n/2 - \Sigma f_{ant})}{f_{Md}} \cdot A$$

temos

$$Md = 2 + \frac{(109/2 - 33)}{39} \cdot 1$$

$$Md = 2 + \frac{(54,5 - 33)}{39}$$

$$Md = 2 + \frac{21,5}{39}$$

$$Md = 2 + 0,551282$$

$$Md = 2,55$$

A mediana corresponde a 2,55 horas.

TEMA 4 – VARIÂNCIA

A variância, representada por σ^2 , é a média dos quadrados dos desvios de cada dado em relação à média. Para dados não agrupados, a variância populacional é dada por

$$\sigma^2 = \frac{\Sigma(x_i - \bar{X})^2}{n}$$

Quando temos uma amostra, a variância dos dados não agrupados é dada por

$$\sigma^2 = \frac{\Sigma(x_i - \bar{X})^2}{n - 1}$$

Exemplo:

Os dados a seguir se referem aos preços de um determinado disco rígido externo de 1TB encontrado em diferentes estabelecimentos comerciais.



R\$ 439,90, R\$ 478,00, R\$ 399,90, R\$ 512,90, R\$ 419,90,

R\$ 439,90, R\$ 439,90, R\$ 489,00, R\$ 459,00, R\$ 499,90

Calcule a variância referente ao preço do disco rígido pesquisado.

Primeiro, precisamos calcular a média dos dados:

$$\bar{X} = \frac{\sum x_i}{n}$$

$$\bar{X} = \frac{439,9 + 478 + 399,9 + 512,9 + 419,9 + 439,9 + 439,9 + 489 + 459 + 499,9}{10}$$

$$\bar{X} = \frac{4578,3}{10}$$

$$\bar{X} = 457,83$$

Como temos o preço médio que é igual a R\$ 457,83, podemos agora calcular $(x_i - \bar{X})^2$ e somar os resultados obtidos.

x_i	$x_i - \bar{X}$	$(x_i - \bar{X})^2$
R\$ 399,90	-R\$ 57,93	R\$ 3.355,88
R\$ 419,90	-R\$ 37,93	R\$ 1.438,68
R\$ 439,90	-R\$ 17,93	R\$ 321,48
R\$ 439,90	-R\$ 17,93	R\$ 321,48
R\$ 439,90	-R\$ 17,93	R\$ 321,48
R\$ 459,00	R\$ 1,17	R\$ 1,37
R\$ 478,00	R\$ 20,17	R\$ 406,83
R\$ 489,00	R\$ 31,17	R\$ 971,57



R\$ 499,90	R\$ 42,07	R\$ 1.769,88
R\$ 512,90	R\$ 55,07	R\$ 3.032,70
Total:	---	R\$ 11.941,38

Estamos considerando alguns preços de disco rígido e não todos os preços existentes. Por este motivo, temos uma amostra e neste caso a variância é calculada por meio da fórmula

$$\sigma^2 = \frac{\sum (x_i - \bar{X})^2}{n - 1}$$

o que resulta em

$$\sigma^2 = \frac{\sum (x_i - \bar{X})^2}{n - 1}$$

$$\sigma^2 = \frac{11941,38}{10 - 1}$$

$$\sigma^2 = \frac{11941,38}{9}$$

$$\sigma^2 = 1326,82$$

A variância corresponde a 1326,82, valor que será utilizado para o cálculo do desvio padrão, tema que ainda nesta aula.

Para dados agrupados em uma distribuição de frequência, a variância populacional é dada por

$$\sigma^2 = \frac{\sum (x_i - \bar{X})^2 \cdot f_i}{n}$$

A variância amostral é dada por

$$\sigma^2 = \frac{\sum (x_i - \bar{X})^2 \cdot f_i}{n - 1}$$

Exemplo:



Uma empresa de comércio eletrônico está fazendo um estudo em relação aos valores dos fretes de suas mercadorias. Na tabela a seguir são apresentados os valores dos fretes e as respectivas quantidades de envios feitos a cada dia.

Frete (x_i)	Quantidade (f_i)
R\$ 12,90	5
R\$ 17,50	12
R\$ 19,80	22
R\$ 19,90	45
R\$ 21,10	39
R\$ 23,40	23
R\$ 27,70	31
R\$ 32,30	18
R\$ 39,90	7
R\$ 45,10	11

A partir destas informações, qual é a variância referente ao frete das mercadorias da empresa?

Para o cálculo da variância, o primeiro passo é calcularmos o valor médio dos fretes:



Frete (x_i)	Quantidade (f_i)	$x_i \cdot f_i$
R\$ 12,90	5	R\$ 64,50
R\$ 17,50	12	R\$ 210,00
R\$ 19,80	22	R\$ 435,60
R\$ 19,90	45	R\$ 895,50
R\$ 21,10	39	R\$ 822,90
R\$ 23,40	23	R\$ 538,20
R\$ 27,70	31	R\$ 858,70
R\$ 32,30	18	R\$ 581,40
R\$ 39,90	7	R\$ 279,30
R\$ 45,10	11	R\$ 496,10
Total:	213	R\$ 5182,20

$$\bar{X} = \frac{\sum x_i \cdot f_i}{n}$$

$$\bar{X} = \frac{5182,2}{213}$$

$$\bar{X} = 24,33$$

A média, neste caso, é igual a R\$ 24,33.

Podemos agora calcular a variância:

Frete (x_i)	Quantidade (f_i)	$(x_i - \bar{X})^2$	$(x_i - \bar{X})^2 \cdot f_i$
--------------------	-------------------------	---------------------	-------------------------------



R\$ 12,90	5	R\$ 130,64	R\$ 653,22
R\$ 17,50	12	R\$ 46,65	R\$ 559,79
R\$ 19,80	22	R\$ 20,52	R\$ 451,46
R\$ 19,90	45	R\$ 19,62	R\$ 883,12
R\$ 21,10	39	R\$ 10,43	R\$ 406,88
R\$ 23,40	23	R\$ 0,86	R\$ 19,89
R\$ 27,70	31	R\$ 11,36	R\$ 352,06
R\$ 32,30	18	R\$ 63,52	R\$ 1143,38
R\$ 39,90	7	R\$ 242,42	R\$ 1696,97
R\$ 45,10	11	R\$ 431,39	R\$ 4745,32
Total:	213	---	R\$ 10912,10

Como temos uma população, a variância é dada por:

$$\sigma^2 = \frac{\sum (x_i - \bar{X})^2 \cdot f_i}{n}$$

$$\sigma^2 = \frac{10912,10}{213}$$

$$\sigma^2 = 51,23$$

Para dados agrupados em uma distribuição de frequência por classe, a variância populacional corresponde a

$$\sigma^2 = \frac{\sum (p_{m_i} - \bar{X})^2 \cdot f_i}{n}$$

e a variância amostral corresponde a



$$\sigma^2 = \frac{\sum (p_{m_i} - \bar{X})^2 \cdot f_i}{n - 1}$$

Exemplo:

Uma empresa de tecnologia desenvolveu um aplicativo de comunicação entre os funcionários de uma indústria e está monitorando o tempo diário em horas de uso deste aplicativo. A seguir são apresentados os resultados diários obtidos.

Tempo	Quantidade
0 ∟ 1	12
1 ∟ 2	21
2 ∟ 3	39
3 ∟ 4	15
4 ∟ 5	22

Qual a variância associada ao tempo de permanência de cada funcionário no aplicativo?

Para o cálculo da média, temos:

Tempo	Quantidade	p_{m_i}	$p_{m_i} \cdot f_i$
0 ∟ 1	12	0,5	6
1 ∟ 2	21	1,5	31,5
2 ∟ 3	39	2,5	97,5
3 ∟ 4	15	3,5	52,5
4 ∟ 5	22	4,5	99



Total	109	---	286,5
-------	-----	-----	-------

$$\bar{X} = \frac{\sum p_{m_i} \cdot f_i}{n}$$

$$\bar{X} = \frac{286,5}{109}$$

$$\bar{X} = 2,63$$

A média referente ao tempo diário de uso do aplicativo é de 2,63 horas.
Vamos agora calcular a variância populacional:

Tempo	Quantidade	p_{m_i}	$(p_{m_i} - \bar{X})^2$	$(p_{m_i} - \bar{X})^2 \cdot f_i$
0 1	12	0,5	4,5369	54,4428
1 2	21	1,5	1,2769	26,8149
2 3	39	2,5	0,0169	0,6591
3 4	15	3,5	0,7569	11,3535
4 5	22	4,5	3,4969	76,9318
Total	109	---	---	170,2021

A variância é dada por:

$$\sigma^2 = \frac{\sum (p_{m_i} - \bar{X})^2 \cdot f_i}{n}$$

$$\sigma^2 = \frac{170,2021}{109}$$

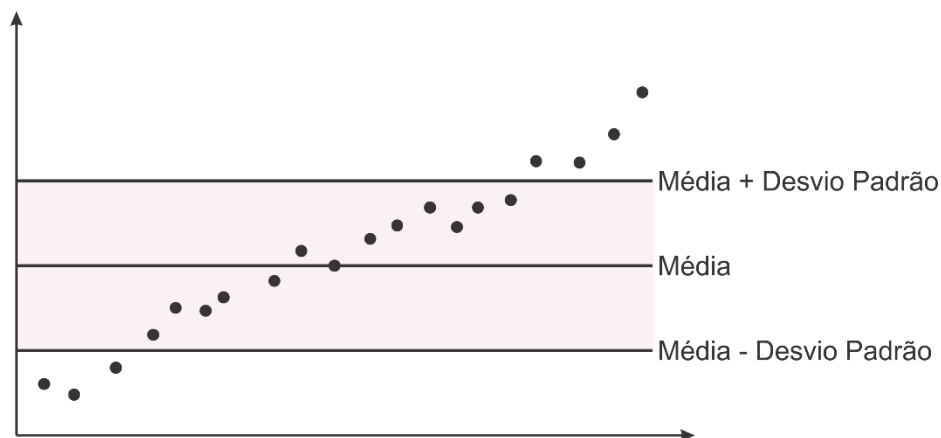
$$\sigma^2 = 1,56$$

A variância é um importante elemento para o cálculo do desvio padrão.



TEMA 5 – DESVIO PADRÃO

A raiz quadrada da variância é o desvio padrão. Esta medida de dispersão indica uma faixa em relação à média que concentra a maior parte dos dados, algo em torno de 65% a 80%. Quanto maior o desvio padrão, maior a dispersão dos dados e quanto menor o desvio padrão, mais próximos os dados estão da média.



Podemos calcular a variância e em seguida calcular a respectiva raiz quadrada. Assim, o desvio padrão da população é calculado por meio da fórmula

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{X})^2}{n}}$$

e o desvio padrão da amostra é calculado por meio da fórmula

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{X})^2}{n - 1}}$$

Exemplo:

Os dados a seguir se referem aos preços de um determinado disco rígido externo de 1TB encontrado em diferentes estabelecimentos comerciais.

R\$ 439,90, R\$ 478,00, R\$ 399,90, R\$ 512,90, R\$ 419,90,

R\$ 439,90, R\$ 439,90, R\$ 489,00, R\$ 459,00, R\$ 499,90

Calcule o desvio padrão referente ao preço do disco rígido pesquisado.



A média dos dados corresponde a:

$$\bar{X} = \frac{\sum x_i}{n}$$

$$\bar{X} = \frac{439,9 + 478 + 399,9 + 512,9 + 419,9 + 439,9 + 439,9 + 489 + 459 + 499,9}{10}$$

$$\bar{X} = \frac{4578,3}{10}$$

$$\bar{X} = 457,83$$

Vamos calcular $(x_i - \bar{X})^2$ e em seguida somar os resultados obtidos.

x_i	$x_i - \bar{X}$	$(x_i - \bar{X})^2$
R\$ 399,90	-R\$ 57,93	R\$ 3.355,88
R\$ 419,90	-R\$ 37,93	R\$ 1.438,68
R\$ 439,90	-R\$ 17,93	R\$ 321,48
R\$ 439,90	-R\$ 17,93	R\$ 321,48
R\$ 439,90	-R\$ 17,93	R\$ 321,48
R\$ 459,00	R\$ 1,17	R\$ 1,37
R\$ 478,00	R\$ 20,17	R\$ 406,83
R\$ 489,00	R\$ 31,17	R\$ 971,57
R\$ 499,90	R\$ 42,07	R\$ 1.769,88
R\$ 512,90	R\$ 55,07	R\$ 3.032,70
Total:	---	R\$ 11.941,38

Como temos uma amostra, o desvio padrão é dado por:



$$\sigma = \sqrt{\frac{\sum (x_i - \bar{X})^2}{n - 1}}$$

$$\sigma = \sqrt{\frac{11941,38}{10 - 1}}$$

$$\sigma = \sqrt{\frac{11941,38}{9}}$$

$$\sigma = \sqrt{1326,82}$$

$$\sigma = 36,43$$

Temos então uma faixa em torno da média, com amplitude igual a 36,43, onde podemos encontrar a maior parte dos dados.

Para dados agrupados em uma distribuição de frequência, o desvio padrão populacional é dado por

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{X})^2 \cdot f_i}{n}}$$

e o desvio padrão amostral é dado por

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{X})^2 \cdot f_i}{n - 1}}$$

Exemplo:

Uma empresa de comércio eletrônico está fazendo um estudo em relação aos valores dos fretes de suas mercadorias. Na tabela a seguir são apresentados os valores dos fretes e as respectivas quantidades de envios feitos a cada dia.

Frete (x_i)	Quantidade (f_i)
R\$ 12,90	5
R\$ 17,50	12



R\$ 19,80	22
R\$ 19,90	45
R\$ 21,10	39
R\$ 23,40	23
R\$ 27,70	31
R\$ 32,30	18
R\$ 39,90	7
R\$ 45,10	11

A partir destas informações, qual é o desvio padrão referente ao frete das mercadorias da empresa?

O valor médio dos fretes é calculado como mostramos a seguir:

Frete (x_i)	Quantidade (f_i)	$x_i \cdot f_i$
R\$ 12,90	5	R\$ 64,50
R\$ 17,50	12	R\$ 210,00
R\$ 19,80	22	R\$ 435,60
R\$ 19,90	45	R\$ 895,50
R\$ 21,10	39	R\$ 822,90
R\$ 23,40	23	R\$ 538,20
R\$ 27,70	31	R\$ 858,70



R\$ 32,30	18	R\$ 581,40
R\$ 39,90	7	R\$ 279,30
R\$ 45,10	11	R\$ 496,10
Total:	213	R\$ 5182,20

$$\bar{X} = \frac{\sum x_i \cdot f_i}{n}$$

$$\bar{X} = \frac{5182,2}{213}$$

$$\bar{X} = 24,33$$

Logo, a média é igual a R\$ 24,33.

Calculando o desvio padrão populacional, temos:

Frete (x_i)	Quantidade (f_i)	$(x_i - \bar{X})^2$	$(x_i - \bar{X})^2 \cdot f_i$
R\$ 12,90	5	R\$ 130,64	R\$ 653,22
R\$ 17,50	12	R\$ 46,65	R\$ 559,79
R\$ 19,80	22	R\$ 20,52	R\$ 451,46
R\$ 19,90	45	R\$ 19,62	R\$ 883,12
R\$ 21,10	39	R\$ 10,43	R\$ 406,88
R\$ 23,40	23	R\$ 0,86	R\$ 19,89
R\$ 27,70	31	R\$ 11,36	R\$ 352,06
R\$ 32,30	18	R\$ 63,52	R\$ 1143,38



R\$ 39,90	7	R\$ 242,42	R\$ 1696,97
R\$ 45,10	11	R\$ 431,39	R\$ 4745,32
Total:	213	---	R\$ 10912,10

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{X})^2 \cdot f_i}{n}}$$

$$\sigma = \sqrt{\frac{10912,10}{213}}$$

$$\sigma = \sqrt{51,23}$$

$$\sigma = 7,16$$

O desvio padrão corresponde a 7,16. Podemos concluir que os valores dos frestes estão, em sua maioria, afastados em até R\$ 7,16 da média, para mais ou para menos.

Quando os dados estão agrupados em uma distribuição de frequência por classe, o desvio padrão populacional é obtido por meio da fórmula

$$\sigma = \sqrt{\frac{\sum (p_{m_i} - \bar{X})^2 \cdot f_i}{n}}$$

O desvio padrão amostral é dado por

$$\sigma = \sqrt{\frac{\sum (p_{m_i} - \bar{X})^2 \cdot f_i}{n - 1}}$$

Exemplo:

Uma empresa de tecnologia desenvolveu um aplicativo de comunicação entre os funcionários de uma indústria e está monitorando o tempo diário em horas de uso deste aplicativo. A seguir são apresentados os resultados diários obtidos.



Tempo	Quantidade
0 ┆ 1	12
1 ┆ 2	21
2 ┆ 3	39
3 ┆ 4	15
4 ┆ 5	22

Qual o desvio padrão associado ao tempo de permanência de cada funcionário no aplicativo?

Para o cálculo da média, temos:

Tempo	Quantidade	p_{mi}	$p_{mi} \cdot f_i$
0 ┆ 1	12	0,5	6
1 ┆ 2	21	1,5	31,5
2 ┆ 3	39	2,5	97,5
3 ┆ 4	15	3,5	52,5
4 ┆ 5	22	4,5	99
Total	109	---	286,5

$$\bar{X} = \frac{\sum p_{m_i} \cdot f_i}{n}$$

$$\bar{X} = \frac{286,5}{109}$$

$$\bar{X} = 2,63$$

A média referente ao tempo diário de uso do aplicativo é de 2,63 horas. O



desvio padrão é calculado como mostrado a seguir:

Tempo	Quantidade	p_{mi}	$(p_{m_i} - \bar{X})^2$	$(p_{m_i} - \bar{X})^2 \cdot f_i$
0 1	12	0,5	4,5369	54,4428
1 2	21	1,5	1,2769	26,8149
2 3	39	2,5	0,0169	0,6591
3 4	15	3,5	0,7569	11,3535
4 5	22	4,5	3,4969	76,9318
Total	109	---	---	170,2021

$$\sigma = \sqrt{\frac{\sum (p_{m_i} - \bar{X})^2 \cdot f_i}{n}}$$

$$\sigma = \sqrt{\frac{170,2021}{109}}$$

$$\sigma = \sqrt{1,561487}$$

$$\sigma = 1,25$$

Como o desvio padrão corresponde a 1,25, os dados estão em sua maioria concentrados em uma faixa com 1,25 horas acima da média e 1,25 horas abaixo da média.



FINALIZANDO

Estamos chegando ao final da nossa segunda aula de Fundamentos Computação. Nesta aula aprendemos a calcular a média, moda e mediana de dados não agrupados. Também aprendemos a calcular a média, a moda e a mediana de dados agrupados em distribuição de frequência e de dados agrupados em distribuição de frequência por classe. Aprendemos também a calcular a variância e o desvio padrão de dados agrupados e de dados não agrupados.

REFERÊNCIAS

- CASTANHEIRA, N. P.; Estatística aplicada a todos os níveis. Curitiba: InterSaberes, 2012.
- DAGHLIAN, Jacob. Lógica e álgebra de boole. 4. ed. São Paulo: Atlas, 1995.
- DASGUPTA, Sanjoy; PAPADIMITRIOU, Christos; VAZIRANI, Umesh. Algoritmos. Porto Alegre: AMGH, 2010.
- GUPTA, B. C.; GUTTMAN, I. Estatística e probabilidade: com aplicações para engenheiros e cientistas. Rio de Janeiro: LTC, 2017.
- LARSON, R.; FARBER, B.; Estatística aplicada. 6ª ed. São Paulo: Pearson, 2015.
- ROCHA, S. Estatística geral e aplicada: para cursos de engenharia. 2ª ed. São Paulo: Atlas, 2015.