PROBABILIDADE E ESTATÍSTICA

AULA 3

Profª Aline Purcote Quinsler

CONVERSA INICIAL

Na Aula 2, estudamos as medidas de posição, chegando a um único valor que apresenta uma ideia de todo o conjunto, mas essas medidas não descrevem detalhadamente o comportamento dos dados. Assim, podemos utilizar as medidas de dispersão para complementar nossas análises e tomar decisões mais assertivas.

Segundo Castanheira (2010, p. 78), as chamadas *medidas de dispersão* são medidas utilizadas para verificar o quanto os valores encontrados em uma pesquisa estão dispersos ou afastados em relação à média.

As medidas de dispersão verificam o grau de variação existente entre os dados, ou seja, se os valores apresentados estão dispersos ou afastados uns dos outros. Por exemplo, considere o valor de um equipamento eletrônico nos últimos cinco meses:

Mês	Valor
1	500
2	1.500
3	1.800
4	2.200
5	2.500

Se calcularmos a média dos últimos cinco meses, teremos um valor médio de R$ 1.700,00, mas, analisando os valores apresentados na tabela, percebemos que há valores diferentes, abaixo e acima da média calculada, ou seja, existe uma dispersão. Mas qual é essa variação? Para responder a essa questão será necessário utilizar as medidas de dispersão, pois só as medidas de posição não são conclusivas.

Nesta aula, estudaremos as medidas de dispersão e como calcular e interpretar os resultados obtidos, além das medidas de assimetria e curtose.

TEMA 1 – MEDIDAS DE DISPERSÃO

A análise realizada pelas medidas de posição pode ser complementada com a utilização das medidas de dispersão. Segundo Castanheira (2010), as medidas de dispersão servem para verificar com que confiança as medidas de

posição resumem as informações fornecidas pelos dados obtidos em uma pesquisa.

As medidas de dispersão indicam se os dados estão afastados da região central, medindo o grau de variação existente entre os valores, e servem também para medir a representatividade da média.

Considere uma pesquisa que represente o preço de dois produtos (A e B) em diferentes pontos de venda:

A: 20, 20, 20

B: 15, 10, 20, 25, 30

Ao calcular a média de preço, obtemos o valor igual a R$ 20,00 para os dois produtos, mas, analisando os valores, temos que no produto A não há variação entre os preços; já no produto B, temos preços diferentes, ou seja, a média é de R$ 20,00, e encontramos o produto por R$10,00 e R$30,00. Logo, para o mesmo produto, encontramos diferenças entre os preços. Assim, os valores apresentam dispersão em torno da média.

Dentre as medidas de dispersão, podemos citar a **amplitude total**, o **desvio médio**, a **variância** e o **desvio padrão**.

A amplitude total é considerada a medida de dispersão mais simples, e é calculada pela diferença entre o maior e o menor valores de uma série de dados:

A = maior – menor

Se o resultado encontrado para a amplitude for um número elevado, significa que os valores da série estão afastados uns dos outros. Caso o valor encontrado seja baixo, os valores da série estão próximos uns dos outros. Dessa forma, quanto maior a amplitude, maior a dispersão dos valores.

Exemplo 1

Considere os valores 40, 45, 48, 62 e 70. Calcule a amplitude total.

Para encontrar a amplitude, precisamos do maior e do menor valor para depois realizar a diferença:

Maior valor = 70

Menor valor = 40

Amplitude = 70 – 40 = 30

Exemplo 2

Qual é a amplitude do preço pago por um equipamento eletrônico nos últimos cinco meses?

Mês	Valor
1	500
2	1.500
3	1.800
4	2.200
5	2.500

Maior valor = 2.500

Menor valor = 500

Amplitude = 2.500 – 500 = 2.000

Segundo Castanheira (2010), para o caso de os dados estarem agrupados em classes, o cálculo da amplitude total pode ser realizado de duas formas:

1. pelos pontos médios das classes. Nesse caso, a amplitude total é igual ao ponto médio da última classe, menos o ponto médio da primeira classe;
2. pelos limites das classes. Nesse caso, a amplitude total é igual ao limite superior da última classe, menos o limite inferior da primeira classe.

Exemplo 3

Qual é a amplitude da seguinte distribuição?

Estaturas (cm)	Frequência
$150 \vdash 154$	4
$154 \vdash 158$	9
$158 \vdash 162$	11
$162 \vdash 166$	8
$166 \vdash 170$	5
$170 \vdash 174$	3
Total	40

Calcule a amplitude considerando as duas formas citadas anteriormente:

1. pelo ponto médio das classes. Nesse caso, para calcular a amplitude total, precisamos encontrar o ponto médio da última classe e o ponto médio da primeira classe para depois realizar a diferença. Lembre-se de que o ponto médio é calculado pela fórmula:

$$Pm = \frac{Ls + Li}{2}$$

Ponto médio da última classe:

$$Pm = \frac{174 + 170}{2} = \frac{344}{2} = 172$$

Ponto médio da primeira classe:

$$Pm = \frac{154 + 150}{2} = \frac{304}{2} = 152$$

Amplitude = 172 – 152 = 20 cm

2. pelos limites das classes. Para calcular a amplitude total, precisamos encontrar o limite superior da última classe e o limite inferior da primeira classe para depois realizar a diferença:

Limite superior da última classe = 174
Limite inferior da primeira classe = 150
Amplitude = 174 – 150 = 24 cm

A amplitude total apresenta algumas restrições, pois considera apenas os valores extremos da série, desprezando os valores intermediários. Segundo Martins (2010, p. 52), a utilização da amplitude total como medida de dispersão é limitada, pois, sendo uma medida que depende apenas dos valores extremos, não capta possíveis variações entre seus limites.

TEMA 2 – DESVIO MÉDIO

O **desvio médio** é uma medida de dispersão que analisa a média dos desvios em torno da média de cada um dos valores da série e é calculado pela

média dos valores absolutos dos desvios. Representa a média das distâncias entre cada elemento da amostra e seu valor médio.

Chamamos *Dm* o desvio médio e o calculamos pela fórmula:

$$Dm = \frac{\sum |x - \bar{x}|.f}{N}$$

onde $|x - \bar{x}|$ é o módulo de cada desvio em relação à média e *N* é igual à soma das frequências. O módulo (| |) utilizado no cálculo do desvio médio possui a função de tornar o número positivo, assim, se a diferença entre o dado e a média resultar em um número positivo, ao se retirar o módulo ele continua positivo, e se for negativo, vira positivo.

Como o desvio médio verifica o afastamento em relação à média, o primeiro passo é calcular a média. Depois, aplicamos a fórmula para encontrar o desvio médio.

Exemplo 1

Suponha os seguintes dados que representem a quantidade de anos de vida útil de um equipamento eletrônico e determine o desvio médio desse conjunto de dados:

3 7 8 10 11

Para calcular o desvio médio, calculamos primeiramente a média. Lembre-se de que para calcular a média em dados não agrupados somamos todos os valores e dividimos pelo número de observações:

$$\bar{X} = \frac{\sum X}{N}$$

$$\bar{X} = \frac{3+7+8+10+11}{5} = \frac{39}{5} = 7{,}8$$

O segundo passo é aplicar a fórmula do desvio médio:

$$Dm = \frac{\sum |x - \bar{x}|.f}{N}$$

Primeiro, calculamos o desvio de cada valor em relação à média, ou seja, cada valor menos a média, que é 7,8. Os valores encontrados,

multiplicamos pela frequência, que é o número de vezes que o valor aparece. Por exemplo, se considerarmos o primeiro valor, que é 3, temos |3 – 7,8|.1, ou seja, o número 3 menos a média, que é 7,8 vezes 1, pois o número 3 aparece apenas uma vez $Dm = \frac{|3-7,8|1+|7-7,8|1+|8-7,8|1+|10-7,8|1+|11-7,8|1}{5}$ e, depois, dividimos por 5, [illegible] ntidade de dados apresentados:

Resolvendo a subtração dentro de cada módulo, temos:

$$Dm = \frac{|-4,8|1+|-0,8|1+|0,2|1+|2,2|1+|3,2|1}{5}$$

Agora, precisamos retirar os valores do módulo, lembrando que se o número for positivo ele continua positivo e o número negativo torna-se positivo, assim:

$$Dm = \frac{4,8.1+0,8.1+0,2.1+2,2.1+3,2.1}{5}$$

Multiplicamos os valores pela frequência, somamos e dividimos por 5:

$$Dm = \frac{4,8+0,8+0,2+2,2+3,2}{5}$$

$$Dm = \frac{11,2}{5} = 2,24$$

Esse resultado indica que, em média, a vida útil desse equipamento eletrônico se desvia em 2,24 anos da média, que é de 7,8 anos.

Exemplo 2

Em um determinado dia foi registrado o número de veículos negociados por uma amostra de 10 vendedores de uma agência de automóveis, como mostra a tabela a seguir. Calcule o desvio médio. (Adaptado de Shiguti; Shiguti, 2006.)

veículos negociados (x_i)	número de vendedores (F_i)
1	1
2	3
3	5
4	1

O primeiro passo é calcular a média. Lembre-se de que, nesse exemplo, temos uma distribuição de frequência e que a média é calculada pela fórmula:

$$\overline{X} = \frac{\sum X.f}{N}$$

veículos negociados (x_i)	número de vendedores (F_i)	x_i F_i
1	1	**1**
2	3	**6**
3	5	**15**
4	1	**4**
TOTAL	10	**26**

$$\overline{X} = \frac{26}{10} = 2{,}6$$

Agora, calculamos o desvio em relação à média. Para facilitar, incluimos uma nova coluna na tabela, identificando o cálculo |x – média|, assim para o primeiro valor da tabela, temos: |1 – 2,6| = |-1,6| = 1,6. Seguimos esse mesmo processo para os demais valores da tabela:

veículos negociados (x_i)	número de vendedores (F_i)	**|xi-média|**
1	1	**1,60**
2	3	**0,60**
3	5	**0,40**
4	1	**1,40**
TOTAL	10	

Encontrados os valores dos desvios, devemos multiplicá-los pelas suas respectivas frequências, incluindo mais uma coluna chamada |x – média|*f. Para o primeiro valor, temos: 1,60 * 1 = 1,60. Seguimos esse processo para os demais valores da tabela e, depois, somamos todos os valores encontrados:

veículos negociados (x_i)	número de vendedores (F_i)	|xi-média|	|xi-média|*Fi
1	1	1,60	1,60
2	3	0,60	1,80
3	5	0,40	2,00
4	1	1,40	1,40
TOTAL	10	4,00	6,80

Para finalizar, aplicamos a fórmula do desvio médio:

$$Dm = \frac{\sum |x - \bar{x}|.f}{N}$$

$$Dm = \frac{6,80}{10} = 0,68$$

A quantidade de veículos negociados por cada vendedor possui um desvio médio de 0,68 em torno dos 2,6 veículos comercializados em média.

Para dados agrupados em classes ou intervalos, substituímos o *X* na fórmula do desvio médio pelo ponto médio de cada classe (*Pm*).

$$Dm = \frac{\sum |x - \bar{x}|.f}{N} \Longrightarrow Dm = \frac{\sum |Pm - \bar{x}|.f}{N}$$

Dessa forma, para calcular o desvio médio em uma distribuição de frequência por classe, temos os seguintes passos:

1. calcular o ponto médio de cada classe;
2. calcular a média;
3. calcular o desvio em relação à média: $\sum |Pm - \bar{x}|.f$
4. calcular o desvio médio.

Exemplo 3

A tabela a seguir representa as notas obtidas por um grupo de 58 alunos matriculados em uma determinada disciplina. Calcule o desvio médio. (Adaptado de Shiguti; Shiguti, 2006.)

O primeiro passo é calcular o ponto médio de cada classe. Lembre-se de que, para calcular o ponto médio, utilizamos a fórmula:

$$Pm = \frac{Ls + Li}{2}$$

Considerando a primeira classe, temos:

$$Pm = \frac{45 + 35}{2} = 40$$

Seguindo o mesmo processo para as demais classes, obtemos:

ESCORES			ALUNOS (F_i)	*Pm*
35	|-	45	5	40
45	|-	55	12	50
55	|-	65	18	60
65	|-	75	14	70
75	|-	85	6	80
85	|-	95	3	90
TOTAL			58	-

No próximo passo, calculamos a média da distribuição de frequência utilizando a fórmula:

$$\bar{X} = \frac{\sum Pm.f}{N}$$

ESCORES			ALUNOS (F_i)	*Pm*	*Pm.f*
35	|-	45	5	40	200
45	|-	55	12	50	600
55	|-	65	18	60	1.080
65	|-	75	14	70	980
75	|-	85	6	80	480
85	|-	95	3	90	270
TOTAL			58	-	3.610

$$\bar{X} = \frac{3610}{58} = 62{,}24$$

Encontrada a média, precisamos calcular os desvios em relação a esse valor:

$$\sum |Pm - \bar{x}|.f$$

Para primeira classe, temos:

|40 – 62,24|*5

| -22,24|*5 = 22,24*5 = 111

Seguindo esse cálculo para as demais classes, e após somarmos os valores obtidos, temos:

Notas			ALUNOS F_i	*Pm*	$\lvert Pm - \bar{x} \rvert$	$\lvert Pm - \bar{x} \rvert f$
35	|-	45	5	**40**	**22**	**111**
45	|-	55	12	**50**	**12**	**147**
55	|-	65	18	**60**	**2**	**40**
65	|-	75	14	**70**	**8**	**109**
75	|-	85	6	**80**	**18**	**107**
85	|-	95	3	**90**	**28**	**83**
TOTAL			58	-	-	**597**

Por fim, aplicamos a fórmula do desvio médio:

$$Dm = \frac{\sum |Pm - \bar{x}|.f}{N}$$

$$Dm = \frac{597}{58} = 10{,}29$$

A nota de cada aluno possui uma distância de 10,29 pontos do desempenho médio, que foi de 62,24 pontos.

TEMA 3 – VARIÂNCIA E DESVIO PADRÃO

A dispersão dos dados também pode ser calculada considerando os quadrados dos desvios médios. Segundo Castanheira, à média aritmética dos quadrados dos desvios damos o nome de variância, que pode ser calculada de duas formas: considerando uma população ou uma amostra.

População:

$$S^2 = \frac{\sum (x - \bar{x})^2 . f}{N}$$

No cálculo da variância de uma amostra, o denominador deverá ser (N – 1), ou seja:

$$S^2 = \frac{\sum (x - \bar{x})^2 . f}{N - 1}$$

onde *x* representa os dados, $\bar{x}$ é a média do conjunto de dados, *f* é a frequência com que o dado aparece e *N* é o número de observações. Como a

variância utiliza o quadrado dos desvios médios, o primeiro passo é calcular a média para depois aplicar as fórmulas indicadas.

Ao analisar o resultado da variância, temos que, quanto maior for o seu valor, mais distante da média estarão os dados, e quanto menor, mais próximos os valores estarão da média, ou seja, se os desvios forem baixos, teremos pouca dispersão, e se forem altos, teremos elevada dispersão.

Segundo Martins (2010), para melhor interpretar a dispersão de uma variável, calcula-se a raiz quadrada da variância, obtendo-se o desvio padrão. O desvio padrão também será calculado para uma população ou uma amostra:

- população:

$$S = \sqrt{\frac{\sum (x - \bar{x})^2 . f}{N}}$$

- amostra:

$$S = \sqrt{\frac{\sum (x - \bar{x})^2 . f}{N - 1}}$$

Podemos utilizar as fórmulas anteriores ou calcular a variância e, depois, tirar a raiz quadrada, assim:

$$S = \sqrt{S^2}$$

Exemplo 1

Suponha o conjunto de tempo de serviço de 5 funcionários e determine a variância e o desvio padrão desse conjunto de dados, considerando uma amostra.

3 7 8 10 11

O primeiro passo é calcular a média. Lembre-se de que, para dados não agrupados, somamos os dados e dividimos pela quantidade de observações:

$$\bar{X} = \frac{\sum X}{N}$$

$$\bar{X} = \frac{3 + 7 + 8 + 10 + 11}{5} = \frac{39}{5} = 7{,}8$$

Depois de encontrada a média, calculamos a variância, verificando que o enunciado solicita a variância considerando uma amostra. Assim, utilizamos a seguinte fórmula:

$$S^2 = \frac{\sum (x - \bar{x})^2 . f}{N - 1}$$

$$S^2 = \frac{(3 - 7{,}8)^2 . 1 + (7 - 7{,}8)^2 . 1 + (8 - 7{,}8)^2 . 1 + (10 - 7{,}8)^2 . 1 + (11 - 7{,}8)^2 . 1}{5 - 1}$$

$$S^2 = \frac{(-4{,}8)^2 . 1 + (-0{,}8)^2 . 1 + (0{,}2)^2 . 1 + (2{,}2)^2 . 1 + (3{,}2)^2 . 1}{4}$$

$$S^2 = \frac{23{,}04 . 1 + 0{,}64 . 1 + 0{,}04 . 1 + 4{,}84 . 1 + 10{,}24 . 1}{4}$$

$$S^2 = \frac{23{,}04 + 0{,}64 + 0{,}04 + 4{,}84 + 10{,}24}{4}$$

$$S^2 = \frac{38{,}80}{4} = 9{,}7$$

Para finalizar, calculamos o desvio padrão tirando a raiz quadrada da variância.

$$S = \sqrt{S^2} = \sqrt{9{,}7} = 3{,}11$$

Exemplo 2

Considere os seguintes dados e calcule a variância e o desvio padrão considerando uma população.

40 45 48 52 54 62 70

Calcule a média desse conjunto de dados:

$$\bar{X} = \frac{40 + 45 + 48 + 52 + 54 + 62 + 70}{7} = \frac{371}{7} = 53$$

Depois de encontrada a média, calculamos a variância, verificando que o enunciado solicita a variância considerando uma população. Assim, utilizamos a seguinte fórmula:

$$S^2 = \frac{\sum (x - \bar{x})^2 . f}{N}$$

$$S^2 = \frac{(40-53)^2.1 + (45-53)^2.1 + (48-53)^2.1 + (52-53)^2.1 + (54-53)^2.1 + (62-53)^2.1 + (70-53)^2.1}{7}$$

$$S^2 = \frac{169+64+25+1+1+81+289}{7}$$

$$S^2 = \frac{630}{7} = 90$$

Para finalizar, calculamos o desvio padrão tirando a raiz quadrada da variância:

$$S = \sqrt{S^2} = \sqrt{90} = 9{,}4868$$

Exemplo 3

Em um determinado dia foi registrado o número de veículos negociados por uma amostra de 10 vendedores de uma agência de automóveis, como mostra a tabela a seguir. Calcule a variância e o desvio padrão. (Adaptado de Shiguti; Shiguti, 2006.)

veículos negociados (x_i)	número de vendedores (F_i)
1	1
2	3
3	5
4	1

Nesse exemplo, temos uma distribuição de frequência e precisamos calcular a variância. Logo, o primeiro passo é o cálculo da média. Lembre-se de que a média em uma distribuição de frequência é calculada pela fórmula:

$$\overline{X} = \frac{\sum X.f}{N}$$

veículos negociados (x_i)	número de vendedores (F_i)	$x_i F_i$
1	1	**1**
2	3	**6**
3	5	**15**
4	1	**4**
TOTAL	10	**26**

$$\overline{X} = \frac{26}{10} = 2{,}6$$

Após o cálculo da média, calculamos o quadrado dos desvios em relação à média e multiplicamos o valor encontrado por sua respectiva frequência. Para o primeiro valor, temos:

$$(1 - 2{,}6)^2 . 1 = (-1{,}6)^2 . 1 = 2{,}56 . 1 = 2{,}56$$

Seguindo esse cálculo, para os demais valores da distribuição, temos:

veículos negociados (x_i)	número de vendedores (F_i)	**(xi-média)^2**	**(xi-média)^2*Fi**
1	1	**2,56**	**2,56**
2	3	**0,36**	**1,08**
3	5	**0,16**	**0,80**
4	1	**1,96**	**1,96**
TOTAL	10		**6,40**

Somamos o valor encontrado em $(x - \bar{x})^2.f$ e aplicamos a fórmula da variância para uma amostra, encontrando o seguinte valor:

$$S^2 = \frac{\sum (x - \bar{x})^2 . f}{N - 1}$$

$$S^2 = \frac{6{,}4}{10 - 1} = \frac{6{,}4}{9} = 0{,}71$$

Tiramos a raiz quadrada da variância para encontrar o desvio padrão:

$$S = \sqrt{S^2} = \sqrt{0{,}71} = 0{,}84$$

Para uma distribuição de frequência por classe ou intervalos, substituímos na fórmula da variância o valor de *x* pelo ponto médio (*Pm*) de cada classe. Dessa forma, o primeiro passo será o cálculo do ponto médio, para depois calcular a média e a variância.

- população:

$$S^2 = \frac{\sum (Pm - \bar{x})^2 . f}{N}$$

- amostra:

$$S^2 = \frac{\sum (Pm - \bar{x})^2 . f}{N - 1}$$

Exemplo 4

A tabela a seguir representa as notas obtidas por um grupo de 58 alunos matriculados em uma determinada disciplina. Calcule a variância e o desvio padrão da amostra. (Adaptado de Shiguti; Shiguti, 2006.)

ESCORES			ALUNOS F_i
35	|-	45	5
45	|-	55	12
55	|-	65	18
65	|-	75	14
75	|-	85	6
85	|-	95	3

Nesse exemplo, temos uma distribuição de frequência por classe. Iniciamos calculando o ponto médio (*Pm*) e a média da distribuição:

ESCORES			ALUNOS (F_i)	*Pm*	*Pm.f*
35	|-	45	5	40	200
45	|-	55	12	50	600
55	|-	65	18	60	1.080
65	|-	75	14	70	980
75	|-	85	6	80	480
85	|-	95	3	90	270
TOTAL			58	-	3.610

$$\overline{X} = \frac{\sum Pm.f}{N}$$

$$\overline{X} = \frac{3610}{58} = 62{,}24$$

Agora, calculamos os desvios: (*Pm*– $\bar{x}$)². O resultado, multiplicamos pela frequência. Para a primeira classe, temos:

$$(40 - 62{,}24)^2 = (-22{,}24)^2 = 495$$
$$495 \,.\, 5 = 2.473$$

Seguindo o mesmo processo para as demais classes e somando os valores obtidos, temos:

ESCORES	ALUNOS F_i	Pm	(Pm-média)²	(Pm-média)².f
35 |- 45	5	40	495	2.473
45 |- 55	12	50	150	1.798
55 |- 65	18	60	5	90
65 |- 75	14	70	60	843
75 |- 85	6	80	315	1.893
85 |- 95	3	90	771	2.312
TOTAL	58	-	-	9.409

Agora, calculamos a variância solicitada da amostra:

$$S^2 = \frac{\sum (Pm - \bar{x})^2 . f}{N - 1}$$

$$S^2 = \frac{9409}{58 - 1}$$

$$S^2 = \frac{9409}{57} = 165{,}1$$

Para calcular o desvio padrão, tiramos a raiz quadrada da variância:

$$S = \sqrt{S^2} = \sqrt{165{,}1} = 12{,}85$$

No desvio padrão, obtemos valores altos sempre que ocorrem mudanças consideráveis nos dados analisados e valores baixos quando os dados são mais estáveis. Segundo Martins (2010), quanto maior o desvio padrão, maiores a dispersão e a amplitude total da distribuição.

TEMA 4 – MEDIDAS DE ASSIMETRIA

De acordo com Castanheira (2010), a média corresponde ao centro de gravidade dos dados; a variância e o desvio padrão medem a variabilidade, mas a distribuição dos pontos sobre um eixo ainda tem outras características que podem ser medidas – uma delas é a assimetria. A assimetria complementa as medidas de posição e dispersão, pois proporciona uma descrição e a compreensão mais completa das distribuições de frequências, já que as distribuições também se diferenciam quanto à sua forma.

Definimos assimetria como o grau de afastamento de uma distribuição da unidade de simetria, pois indica o grau de deformação de uma curva de frequências. Quando uma distribuição é simétrica, temos a igualdade dos valores de média, mediana e moda, conforme figura abaixo:

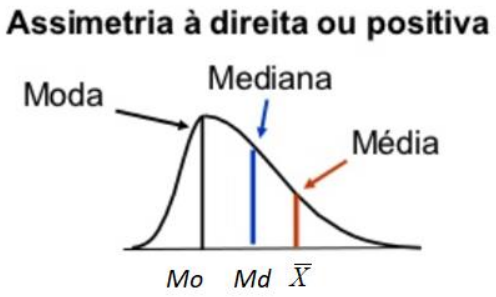
Figura 1 – Distribuição simétrica



Uma distribuição assimétrica pode ser assimétrica positiva, também chamada de *assimétrica à direita,* ou assimétrica negativa, também chamada de *assimétrica à esquerda*.

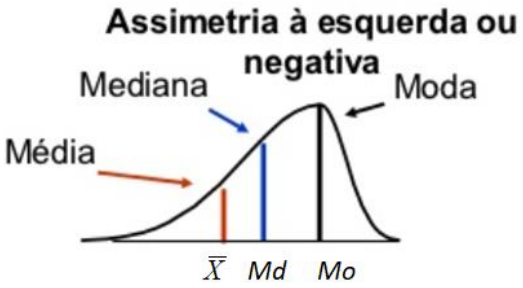
Em uma distribuição assimétrica positiva a média é maior que a mediana e a moda, ou seja, $\overline{X}$ >*Md* > *Mo,* conforme observamos na figura a seguir:

Figura 2 – Assimetria à direita ou positiva



Na distribuição assimétrica negativa, temos que a média é menor que a mediana e a moda, assim, $\overline{X} < Md < Mo$, conforme observamos na figura a seguir:

Figura 3 – Assimetria à esquerda ou negativa



Existem várias fórmulas para o cálculo do coeficiente de assimetria. Dentre eles, estudaremos o coeficiente de assimetria de Pearson. O 1º coeficiente de assimetria de Pearson é calculado por:

$$A_s = \frac{\overline{X} - Mo}{S}$$

Além do 1º coeficiente, podemos calcular o 2º coeficiente de Pearson aplicando a seguinte fórmula:

$$A_s = \frac{3.(\overline{X} - Md)}{S}$$

Analisando o valor do coeficiente, temos:

- $A_S = 0$, a distribuição é simétrica;
- $A_S > 0$, a distribuição é assimétrica positiva ou à direita;
- $A_S < 0$, a distribuição é assimétrica negativa ou à esquerda.

Exemplo 1

Uma empresa inspecionou 50 componentes eletrônicos para determinar o tempo de vida útil desse componente, obtendo a distribuição que vemos a seguir. Calcule o 1º coeficiente de assimetria de Pearson.

Tempo (horas)	Frequências
1200 |--- 1300	1
1300 |--- 1400	3
1400 |--- 1500	11
1500 |--- 1600	20
1600 |--- 1700	10
1700 |--- 1800	3
1800 |--- 1900	2

Para calcular o 1º coeficiente de Pearson, precisamos dos valores de média, moda e desvio padrão. Na Aula 2, calculamos a média e a moda obtendo os seguintes resultados:

- média:

Tempo (horas)	Frequências	PM	PM.f
1200 |--- 1300	1	1250	1250
1300 |--- 1400	3	1350	4050
1400 |--- 1500	11	1450	15950
1500 |--- 1600	20	1550	31000
1600 |--- 1700	10	1650	16500
1700 |--- 1800	3	1750	5250
1800 |--- 1900	2	1850	3700
	50		77700

$$\overline{X} = \frac{77700}{50} = 1554$$

- moda:

$$Mo = 1500 + \frac{10.100}{11+10}$$

- $$Mo = 1500 + \frac{1000}{21}$$

- $Mo = 1500 + 47{,}62 = 1547{,}62$

Para calcular o desvio padrão, seguimos os passos indicados no Tema 3 desta aula:

- desvio padrão:

Tempo (horas)	Frequências	*PM*	*(PM – Média)²*	*(PM – Média)².f*
1200 |--- 1300	1	1250	92416	92416
1300 |--- 1400	3	1350	41616	124848
1400 |--- 1500	11	1450	10816	118976
1500 |--- 1600	20	1550	16	320
1600 |--- 1700	10	1650	9216	92160
1700 |--- 1800	3	1750	38416	115248
1800 |--- 1900	2	1850	87616	175232
	50			719200

$$S^2 = \frac{719200}{49} = 14677{,}55$$

$$S^2 = \sqrt{14677{,}55} = 121{,}15$$

Agora, calculamos o 1º coeficiente de assimetria de Pearson aplicando os valores obtidos na fórmula:

$$A_s = \frac{\overline{X} - Mo}{S}$$

$$A_s = \frac{1554 - 1547{,}62}{121{,}15}$$

$$A_s = \frac{6{,}38}{121{,}15} = 0{,}0052662$$

Exemplo 2

Considere uma distribuição de frequência que apresente média igual a 88, mediana igual a 82 e desvio padrão igual a 40. Calcule o 2º coeficiente de Pearson.

Com os valores fornecidos no enunciado, calculamos o coeficiente aplicando a fórmula:

$$A_s = \frac{3.(\overline{X} - Md)}{S}$$

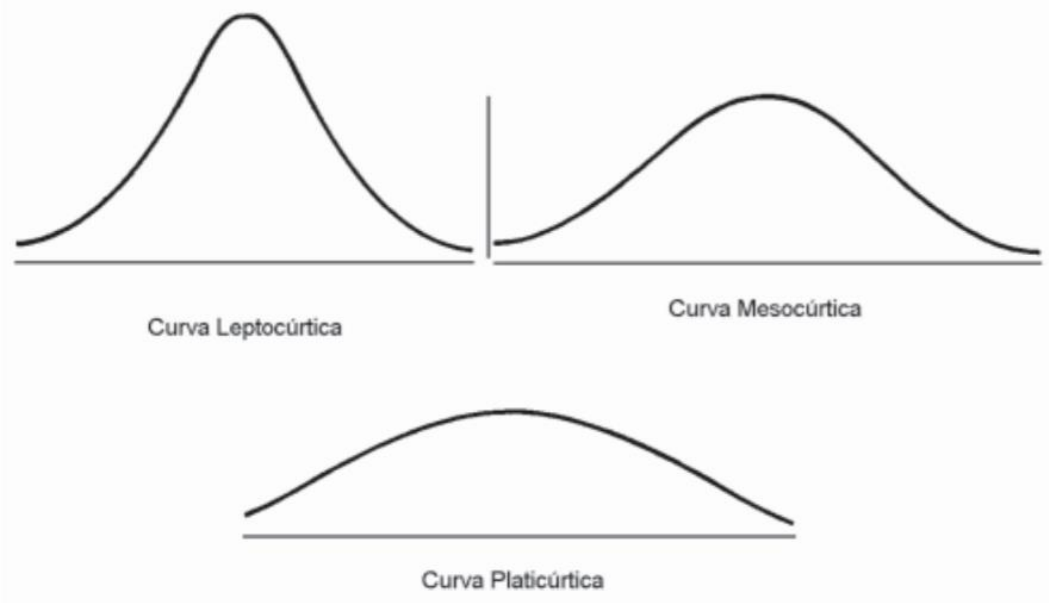
$$A_s = \frac{3.(88 - 82)}{40}$$

$$A_s = \frac{3.(6)}{40} = \frac{18}{40} = 0{,}45$$

TEMA 5 – MEDIDAS DE CURTOSE

Segundo Castanheira (2010), a curtose indica o quanto uma distribuição de frequências é mais achatada ou mais afilada do que uma curva padrão, a qual é denominada de curva normal. A **curva normal** ou **distribuição normal** será estudada na Aula 5.

Ao realizar a análise em relação ao achatamento, temos que a distribuição normal é chamada de *mesocúrtica*, em que os dados estão uniformemente distribuídos. As distribuições mais achatadas que a normal são as *platicúrticas*, em que os dados estão bem dispersos em relação à média. Às distribuições menos achatadas ou mais alongadas que a normal chamamos de *leptocúrticas*, em que os dados estão concentrados em torno da média. Analisaremos cada distribuição nos gráficos seguintes:



Para determinar a curtose, aplicamos a seguinte fórmula:

$$K = \frac{Q_3 - Q_1}{2(p_{90} - p_{10})}$$

onde:

K = coeficiente percentílico de curtose

Q_1= primeiro quartil

Q_3= terceiro quartil

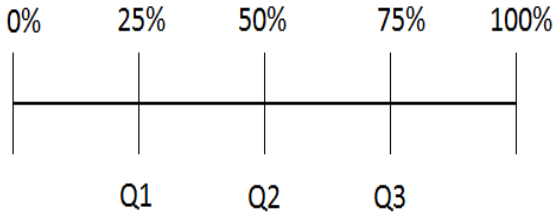
p_{10}= décimo percentil

p_{90}= nonagésimo percentil

Analisando o valor de K, temos a seguinte classificação:

- K = 0,263 – curva normal, distribuição mesocúrtica;
- K > 0,263 – curva mais achatada, distribuição platicúrtica;
- K < 0,263 – curva mais alongada, distribuição leptocúrtica.

Para o cálculo do coeficiente, precisamos encontrar o quartil e o percentil. O quartil divide uma distribuição em quatro partes iguais e é representado por Q_i, onde *i* representa a ordem do quartil. No diagrama a seguir, temos que o 1º quartil representa 25% dos dados, o 2º quartil representa 50% e o terceiro representa 75% dos dados. Isso ocorre porque dividimos 100% dos dados por 4, obtendo 25%. Assim, a cada quartil, acumulamos 25%.



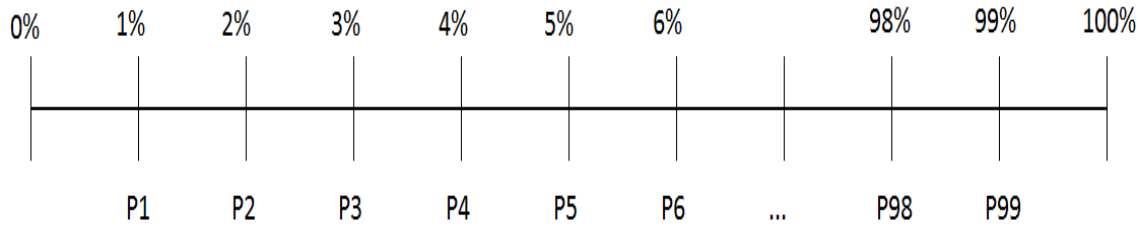
Para calcular o quartil em uma distribuição de frequência por classe e intervalos, seguimos alguns passos que são muito próximos ao cálculo realizado na mediana por classe. A diferença está no cálculo da posição que dividimos por 4, e em precisarmos indicar o quartil a ser calculado. Para calcular o quartil, aplicamos os passos seguintes:

1. encontramos o valor de *N* que é igual a soma das frequências;
2. calculamos a posição = $\frac{N}{4}.i$, onde i representa o quartil a ser calculado, assim i = 1, 2 ou 3;
3. calculamos a frequência acumulada (f_a);
4. identificamos na frequência acumulada a posição calculada no passo 2 (sempre devemos buscar um valor igual ou maior que a posição calculada).
5. calculamos o quartil, utilizando a fórmula:

$$Qi = Li + \frac{(\frac{N}{4}.i - \sum f_{ant})}{f_{Qi}}.A$$

Já o percentil permite dividir a distribuição em cem partes iguais e é representado por P_i, onde *i* representa a ordem do percentil (1, 2, 3,...., 99). No diagrama a seguir, verificamos que cada percentil corresponde a 1% do

conjunto de dados, pois dividimos 100% dos dados por 100, assim, a cada percentil, acumulamos 1%.



A estrutura de cálculo para o percentil é exatamente igual ao do quartil, porém, temos uma modificação no cálculo da posição, pois estamos falando de 100 partes iguais:

- posição:

  $\frac{N}{100}.i$, onde i representa o percentil a ser calculado, assim i = 1, 2,..., 99.

- fórmula:

$$Pi = Li + \frac{(\frac{N}{100}.i - \sum f_{ant})}{f_{Pi}}.A$$

Exemplo 1

Considere a seguinte distribuição que indica as faixas salariais, em salários mínimos, dos funcionários de uma determinada empresa. Calcule o coeficiente percentílico de curtose e interprete o resultado obtido indicando qual o tipo de curva de frequência temos nessa distribuição.

Salários	Nº de Funcionários
02 |--- 04	3
04 |--- 06	6
06 |--- 08	12
08 |--- 10	6
10 |---| 12	3

Para o cálculo, precisamos encontrar o quartil 1 e o 3, além do percentil 10 e do 90. Para calcular o quartil, seguimos os passos:

- quartil 1 (Q_1):

1. encontramos o valor de N que é igual à soma das frequências:

N = 30

Salários	Nº de Funcionários
02 |--- 04	3
04 |--- 06	6
06 |--- 08	12
08 |--- 10	6
10 |---| 12	3
Total	30

2. calculamos a posição = $\frac{N}{4}.i$, onde *i* representa o quartil a ser calculado, assim *i* = 1, 2 ou 3. Como queremos calcular o 1º quartil, vamos substituir *i* por 1:

$$\text{Posição} = \frac{N}{4}.i$$

$$\text{Posição} = \frac{30}{4}.1 = 7{,}5$$

3. calculamos frequência acumulada (f_a):

Salários	Nº de Funcionários	f_a
02 |--- 04	3	3
04 |--- 06	6	9
06 |--- 08	12	21
08 |--- 10	6	27
10 |---| 12	3	30
Total	30	

4. identificamos na frequência acumulada a posição calculada no passo 2. Sempre devemos buscar um valor igual ou maior que a posição calculada. Nesse caso, temos posição igual a 7,5 que identificamos na classe 2:

Salários	Nº de Funcionários	f_a
02 |--- 04	3	3
04 |--- 06	6	9
06 |--- 08	12	21
08 |--- 10	6	27
10 |---| 12	3	30
Total	30	

5. calculamos o quartil, utilizando a fórmula:

$$Qi = Li + \frac{(\frac{N}{4}.i - \sum f_{ant})}{f_{Di}}.A$$

Classe: 04 |--- 06

$L_i = 04$

A = 6 – 4 = 2

$$Q_1 = 4 + \frac{(7{,}5 - 3)}{6}.2$$

$$Q_1 = 4 + \frac{(4{,}5)}{6}.2$$

$$Q_1 = 4 + 1{,}5 = 5{,}5$$

Aplicando os mesmos passos encontramos o 3º quartil (Q3):

- quartil 3 (Q_3):

1. encontramos o valor de *N*, que é igual à soma das frequências:

N = 30

2. calculamos a posição = $\frac{N}{4}.i$, onde *i* representa o quartil a ser calculado, assim *i* = 1, 2 ou 3. Como queremos calcular o 3º quartil, substituímos *i* por 3:

$$\text{Posição} = \frac{N}{4}.i$$

$$\text{Posição} = \frac{30}{4}.3 = 22{,}5$$

3. calculamos a frequência acumulada (f_a).
4. identificamos na frequência acumulada a posição calculada no passo 2. Sempre devemos buscar um valor igual ou maior que a posição calculada. Nesse caso, temos posição igual a 22,5, que identificamos na classe 4:

Salários	Nº de Funcionários	f_a
02 |--- 04	3	3
04 |--- 06	6	9
06 |--- 08	12	21
08 |--- 10	6	27
10 |---| 12	3	30
Total	30	



5. calculamos o quartil, utilizando a fórmula:

$$L_i = 08$$

$$A = 6 - 4 = 2$$

$$Q_3 = 8 + \frac{(22{,}5 - 21)}{6}.2$$

$$Q_3 = 8 + \frac{(1{,}5)}{6}.2$$

$$Q_3 = 8 + 0{,}5 = 8{,}5$$

Agora, precisamos calcular o percentil 10 e o 90 aplicando os mesmos passos do quartil. A única modificação está no cálculo da posição e na aplicação da fórmula:

- percentil 10 (P_{10}):

Posição = $\frac{N}{100}.i$, onde i representa o percentil a ser calculado. No nosso exemplo, usaremos $i = 10$:

$$\text{Posição} = \frac{30}{100}.10 = 3$$

Salários	Nº de Funcionários	f_a
02 |--- 04	3	3
04 |--- 06	6	9
06 |--- 08	12	21
08 |--- 10	6	27
10 |---| 12	3	30
Total	30	

$$Pi = Li + \frac{(\frac{N}{100}.i - \sum f_{ant})}{f_{Pi}}.A$$

$$P_{10} = 2 + \frac{(3-0)}{3}.2$$

$$P_{10} = 2 + \frac{(3)}{3}.2 = 2 + 2 = 4$$

- percentil 90 (P_{90}):

Posição = $\frac{N}{100}.i$, onde *i* é igual a 90:

$$\text{Posição} = \frac{30}{100}.90 = 27$$

Salários	Nº de Funcionários	f_a
02 |--- 04	3	3
04 |--- 06	6	9
06 |--- 08	12	21
08 |--- 10	6	27
10 |---| 12	3	30
Total	30	

$$Pi = Li + \frac{(\frac{N}{100}.i - \sum f_{ant})}{f_{Pi}}.A$$

$$P_{90} = 8 + \frac{(27-21)}{6}.2$$

$$P_{90} = 8 + \frac{(6)}{6}.2 = 8 + 2 = 10$$

Com os valores do quartil e do percentil, aplicamos a fórmula para encontrar o coeficiente:

$$K = \frac{Q_3 - Q_1}{2(p_{90} - p_{10})}$$

$$K = \frac{8{,}5 - 5{,}5}{2(10 - 4)}$$

$$K = \frac{3}{2(6)} = \frac{3}{12} = 0{,}25$$

Obtemos um coeficiente igual a 0,25. Assim, temos uma curva leptocúrtica.

Exemplo 2

Considere uma distribuição que apresente as medidas a seguir. Calcule o coeficiente percentílico de curtose e interprete o resultado obtido indicando o tipo de curva de frequência que temos nessa distribuição.

Q_1 = 24,4 cm

Q_3 = 41,2 cm

P_{10} = 20,2 cm

P_{90} = 49,5 cm

$$K = \frac{Q_3 - Q_1}{2(p_{90} - p_{10})}$$

$$K = \frac{41{,}2 - 24{,}4}{2(49{,}5 - 20{,}2)}$$

$$K = \frac{16{,}8}{2(29{,}3)} = \frac{16{,}8}{58{,}6} = 0{,}2867$$

Obtivemos um coeficiente igual a 0,2867. Assim, temos uma distribuição platicúrtica.

FINALIZANDO

Nesta aula, apresentamos as principais medidas de **dispersão**, seus cálculos, aplicações e interpretações dos resultados obtidos. Estudamos também as medidas de assimetria e a curtose.

REFERÊNCIAS

CASTANHEIRA, N. P. **Estatística aplicada a todos os níveis.** Curitiba: Ibpex, 2010.

LARSON, R.; FARBER, B. **Estatística aplicada.** 2. ed. São Paulo: Pearson, 2004.

MARTINS, G. A. **Estatística geral e aplicada.** 3. ed. São Paulo: Atlas, 2010.

SHIGUTI, W. A.; SHIGUTI, V. S. C. **Apostila de estatística.** Brasília, 2006. Disponível em: <http://www.inf.ufsc.br/~paulo.s.borges/Download/Apostila5_INE5102_Quimica.pdf>. Acesso em: 6 mar. 2020.

WALPOLE, R. E. et al. **Probabilidade e estatística para engenharia e ciências.** São Paulo: Pearson, 2009.