ANÁLISE EXPLORATÓRIA DE DADOS

AULA 4

Profª Dayane Perez Bravo

CONVERSA INICIAL

Nesta aula, iremos aprender sobre as séries e os somatórios, além de iniciar uma discussão sobre as medidas de tendência central que permitem significados extras acerca dos dados medidos.

TEMA 1 – SÉRIES

Vejamos como interpretar e representar tipos de dados compostos por duas variáveis, o que chamamos de **série**. Entre os tipos, podemos citar a série temporal, em que uma das variáveis é o tempo, e a série geográfica, em que uma das variáveis é a localização geográfica.

1.1 Série temporal

A construção e a análise das **séries temporais** é uma das principais análises de dados existentes, visto que a maior parte dos dados envolve associação com tempo. Por exemplo, no estudo de comportamento do preço de uma ação, a relacionamos ao tempo, fazendo uma previsão de seu valor no futuro. Assim, a série temporal representa uma sequência de dados ordenados ao longo do tempo. Além da variável de tempo, denotamos uma variável conhecida como **variável de interesse**.

Considere, por exemplo, o preço de determinada ação no primeiro dia do pregão de cada mês para o ano de 2019. Esses dados estão disponíveis na Tabela 1.

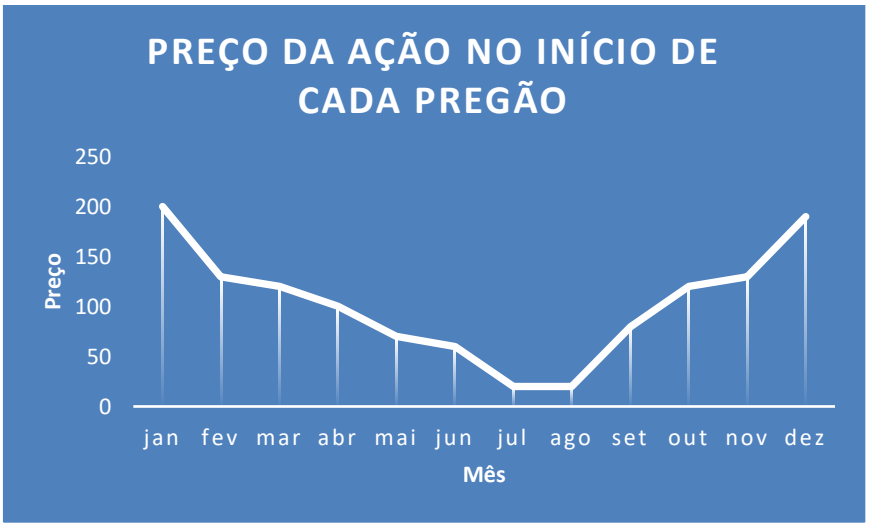
Tabela 1 – Preço de ação no primeiro dia do pregão de cada mês para o ano de 2019

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Preço	200	130	120	100	70	60	20	20	80	120	130	190

Fonte: Bravo, 2021.

Como o tempo é uma variável de tipo contínua, escolhemos por representar esses dados com o uso de um polígono de frequência, como o apresentado no Gráfico 1.

Gráfico 1 – Polígono de frequência para o preço da ação no início de cada pregão



Fonte: Bravo, 2021.

1.2 Série geográfica

A **série geográfica,** por sua vez, apresenta entre suas variáveis dados referentes a diferentes locais ou regiões, para um mesmo período de tempo. Como trata-se de uma série, existem duas variáveis, sendo a segunda de nosso interesse. Como exemplo, podemos observar os dados da Tabela 2.

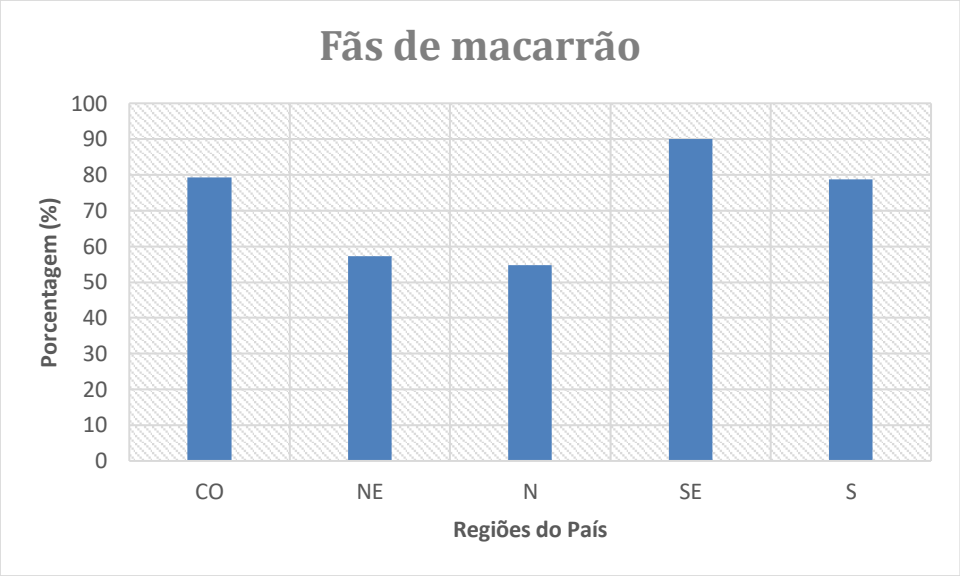
Tabela 2 – Série geográfica apresentando a porcentagem de entrevistados que preferem macarrão ao invés de arroz para as cinco regiões do país

Região	Fãs de macarrão
Centro-Oeste (CO)	79,2
Nordeste (NE)	57,2
Norte (N)	54,8
Sudeste (SE)	90,1
Sul (S)	78,7

Fonte: Bravo, 2021.

Na Tabela 2, é possível observar a quantidade dos entrevistados, em porcentagem, que preferem macarrão ao invés de arroz para as cinco regiões do país. Trata-se de uma série geográfica. Sua representação é feita de forma viável por meio de um gráfico de colunas, como mostra na Figura 2.

Gráfico 2 – Representação da série geográfica para a preferência entre macarrão ou arroz entre os entrevistados das diversas regiões do país



Fonte: Bravo, 2021.

TEMA 2 – SOMATÓRIO

Em estatística, diversas vezes somos convidados a trabalhar com o somatório dos dados. Vejamos como representar matematicamente uma série, com algumas propriedades essenciais para simplificar os resultados obtidos em cada somatório.

2.1 Índice

A representação de uma variável de interesse pode ser utilizada com um símbolo, por exemplo, x. Essa variável representa cada um dos valores que a variável de interesse pode admitir. Quando possuímos uma sequência, digamos:

$$9,14,6,28,11,$$

podemos utilizar um índice para indicar o ordenamento dos dados. Assim, para as cinco observações dadas, teríamos:

$$x_1 = 9, x_2 = 14, x_3 = 6, x_4 = 28, x_5 = 11.$$

Note que os valores do índice assumem valores inteiros, iniciando em 1.

Também podemos usar índices para construir tabelas de frequência. Por exemplo, no caso dos seguintes dados:

$$4,3,5,4,3,2,2,3,2,3,3,4,3,5,4,3$$

Aqui, vemos a ocorrência de quatro dados. Assim, podemos escrever:

$$x_1 = 2, x_2 = 3, x_3 = 4, x_4 = 5$$

A quantidade de vezes que cada dado aparece pode ser resumida pelo uso da função $f(x_k)$. Realizando o processo de contagem, verificamos o seguinte:

$$f(x_1) = 3, f(x_2) = 7, f(x_3) = 4, f(x_4) = 2$$

O leitor deve ter um pouco de cuidado para não confundir essa lógica com o uso do índice, que será um simplificador para a construção de medidas de tendência central. A Tabela 3 resume a frequência para os dados discutidos.

Tabela 3 – Tabela de frequência para os dados exemplificados a partir da notação indicial

k	x_k	$f(x_k)$
1	2	3
2	3	7
3	4	4
4	5	2

Fonte: Bravo, 2021.

2.2 Somatório

Vejamos os dados do exemplo:

$$9,14,6,28,11$$

Aqui, podemos estar interessados em encontrar seu somatório, isto é:

$$S = x_1 + x_2 + x_3 + x_4 + x_5$$

Note que, em termos de representação de uma grande quantidade de dados, precisamos de uma simplificação simbólica. Para isso, utilizamos o símbolo de somatório:

$$\sum_{i=1}^{n} x_i$$

Sabemos que i representa a variável de controle, $i = 1$ o primeiro termo a ser somado, e $i = n$ o último. Assim, para os dados do problema, há uma possível simplificação:

$$\sum_{i=1}^{5} x_i = 68$$

2.3 Propriedades do somatório

Para analisar as medidas de tendência central, é essencial conhecer algumas propriedades do somatório, com vistas a simplificar algumas expressões.

Suponha a multiplicação de todos os dados por uma constante. Somar os dados multiplicados é equivalente a somar os dados e multiplicar pela constante. Em termos simbólicos:

$$\sum_{i=1}^{n} c.x_i = c.\sum_{i=1}^{n} x_i \, , c \in \mathbb{R}$$

Continuamos com os seguintes dados:

$$9{,}14{,}6{,}28{,}11.$$

Se multiplicarmos cada um dos termos por uma constante, digamos 3, obtemos:

$$27, 42, 18, 84, 33$$

Seu somatório é 204. Note que esse é o triplo do resultado anterior, de forma que:

$$\sum_{i=1}^{5} 3.x_i = 3.\sum_{i=1}^{5} x_i$$

Em algumas expressões, seremos convidados a somar uma sequência de termos constantes. Nesse caso, em que $x_1 = x_2 = \cdots = x_n = c$, temos que:

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} c = n.c, c \in \mathbb{R}$$

Em outras situações, temos que simplificar a soma ou a diferença entre duas variáveis. Nesse caso:

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

Por exemplo, considere os dados contidos na tabela 4 para duas variáveis de interesse.

Tabela 4 – Dados obtidos na observação de duas variáveis de interesse

i	x_i	y_i
1	5	2
2	3	4
3	7	8
4	9	8

Fonte: Bravo, 2021.

Note que:

$$\sum_{i=1}^{4} (x_i + y_i) = (5 + 2) + (3 + 4) + (7 + 8) + (9 + 8) = 46$$

$$\sum_{i=1}^{4} x_i = 5 + 3 + 7 + 9 = 24$$

$$\sum_{i=1}^{4} y_i = 2 + 4 + 8 + 8 = 22$$

Ressaltamos que a propriedade discutida anteriormente é válida!

2.4 Exemplos

Vejamos como usar essas propriedades para simplificar o desenvolvimento de alguns somatórios. Por exemplo, considere os seguintes dados:

$$15,14,15,16,13,15$$

Nesse caso, podemos calcular alguns somatórios:

$$\sum_{i=1}^{6} x_i = 15 + 14 + 15 + 16 + 13 + 15 = 88$$

$$\sum_{i=1}^{6} |x_i - 15| = 0 + 1 + 0 + 1 + 2 + 0 = 4$$

$$\sum_{i=1}^{6} (x_1 - 15) = 0 - 1 + 0 + 1 - 2 + 0 = -2$$

$$\sum_{i=1}^{6} x_i^2 = 15^2 + 14^2 + 15^2 + 16^2 + 13^2 + 15^2 = 1.296$$

$$\sum_{i=1}^{6} (x_i - 15)^2 = 0^2 + (-1)^2 + 0^2 + 1^2 + (-2)^2 + 0^2 = 6$$

Considere, também, os dados descritos na Tabela 5.

Tabela 5 – Tabela de frequência para o exemplo dado

k	x_k	$f(x_k)$
1	3	2
2	4	5
3	5	8
4	7	4
5	9	1

Fonte: Bravo, 2021.

Note que, para esses dados, podemos, a título de exemplo, encontrar o valor de alguns somatórios:

$$\sum_{k=1}^{5} x_k = 3 + 4 + 5 + 7 + 9 = 28$$

$$\sum_{k=1}^{5} f(x_k) = 2 + 5 + 8 + 4 + 1 = 20$$

$$\sum_{k=1}^{5} x_k . f(x_k) = 3 \cdot 2 + 4 \cdot 5 + 5 \cdot 8 + 7 \cdot 4 + 9 \cdot 1 = 103$$

$$\sum_{k=1}^{5} x_k - \sum_{k=1}^{5} x_k . f(x_k) = 28 - 103 = -75$$

$$\sum_{k=1}^{5} (x_k - 8) \cdot f(x_k) = (-5) \cdot 2 + (-4) \cdot 5 + (-3) \cdot 8 + (-1) \cdot 4 + 1 \cdot 1 = -57$$

$$\sum_{k=1}^{5} (x_k - 8)^2 = (-5)^2 + (-4)^2 + (-3)^2 + (-1)^2 + 1^2 = 52$$

TEMA 3 – MEDIDAS DE TENDÊNCIA CENTRAL

As medidas de tendência central nos auxiliam com a descrição de um conjunto de dados. Vejamos como defini-las.

3.1 Mediana

Definimos a **mediana** como o valor que divide o conjunto de dados **ordenados** ao meio. Então, considere os dados referentes à quantidade de ganhadores da mega sena nos últimos 13 sorteios:

$$3\ 1\ 3\ 2\ 0\ 2\ 5\ 0\ 1\ 2\ 3\ 4\ 3\ 1$$

Para encontrarmos a mediana, devemos, inicialmente, ordenar os dados, obtendo:

$$0,0,1,1,1,2,2,3,3,3,3,4,5$$

Como são 13 dados, a mediana será o valor do dado que está na posição central, ou seja, na sétima posição. Assim, seu valor é 2, visto que $x_7 = 2$. Note que a mediana separa os seis primeiros dados dos seis últimos.

Nesse caso, a forma de encontrar a mediana é simples, visto que se trata de uma quantidade de dados ímpar. Suponha que não encontrássemos a observação, $x_{13} = 5$, obtendo 12 observações, assim distribuídas e ordenadas:

$$0,0,1,1,1,2,2,3,3,3,3,4$$

No caso de uma quantidade de dados pares, a mediana será dada pela média entre os dois valores centrais. Nesse caso, obtemos o valor central:

$$\frac{2+3}{2} = 2,5$$

3.2 Moda

A **moda,** em estatística, é definida como o valor que mais aparece no conjunto de dados. Em outras palavras, é aquele de maior frequência. Com essa definição, podemos ter mais de uma moda, caracterizando um conjunto de dados bimodais; ou não possuirmos moda nenhuma, caso os dados não se repitam.

Considere uma pesquisa de satisfação sobre o serviço oferecido pela internet de seu bairro por diversas operadoras. Os entrevistados foram solicitados a responder entre: E – Excelente, O – Ótimo, B – Bom, R – Regular, e P – Péssimo.

Assim, a operadora A teve as seguintes respostas:

$$P, R, B, B, O, O, O, O, E, E$$

A operadora B:

$$R, B, B, B, B, O, O, O, O, E$$

A operadora C:

$$P, P, R, R, B, B, O, O, E, E$$

Note que a moda para a operadora A é O, a moda para a operadora B é B e O (ou seja, é bimodal) e a operadora C não tem moda.

3.3 Média aritmética

A média aritmética é definida como a soma de todos os dados, valor dividido pelo número de dados. Em termos da notação de somatório, podemos escrever:

$$\text{média} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Em termos da notação da média, é comum registrarmos duas notações distintas, uma referente à média populacional e outra à média amostral. Assim, consideramos n como o tamanho da amostra e N como o tamanho da população:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Considere, por exemplo, uma amostra extraindo os dados de quantidade de acidentes registrados em uma determinada estrada:

$$3\ 1\ 2\ 0\ 2\ 5\ 0\ 1\ 2\ 2\ 4\ 3\ 1$$

Assim, $n = 13$ e

$$\bar{x} = \frac{\sum_{i=1}^{13} x_i}{13} = \frac{26}{13} = 2$$

3.4 Exemplos

Considere os dados obtidos em um experimento interessado em investigar a quantidade de faltas de um grupo de funcionários de determinada empresa:

$$0\ 5\ 6\ 0\ 2\ 0\ 8\ 0\ 15\ 0\ 0\ 0\ 0\ 8\ 0\ 0\ 6\ 0\ 5\ 5\ 4\ 0\ 0\ 5\ 0\ 0\ 6$$

Para encontrarmos a mediana dos dados, é preciso ordená-los:

$$0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 2\ 4\ 5\ 5\ 5\ 5\ 6\ 6\ 6\ 8\ 8\ 15$$

Como se trata de 27 dados, ou seja, $n = 27$, a mediana é dada por $x_{14} = 0$. A moda também é zero, pois que tem a maior frequência. A média é 2,77, obtida por:

$$\bar{x} = \frac{\sum_{i=1}^{27} x_i}{27} = \frac{75}{27} = 2{,}77$$

Se retirarmos os não faltantes do estudo, teremos um total de 12 dados (i.e., $n = 12$). Nesse caso, a mediana é a média entre os dois elementos centrais (visto que se trata de um número par de elementos), ou seja:

$$\frac{5 + 6}{2} = 5{,}5$$

A moda passa a ser 5 e a média, 6,25:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{12} = \frac{75}{12} = 6{,}25$$

TEMA 4 – MEDIDAS DE TENDÊNCIA CENTRAL COM TABELA DE FREQUÊNCIA

Com base na tabela de frequência, também podemos encontrar as medidas de tendência central discutidas na seção anterior. Vejamos como realizar esse procedimento.

4.1 Mediana

Considere os dados coletados e apresentados na Tabela 6.

Tabela 6 – Tabela de frequência e frequência acumulada para os dados do exemplo

k	x_k	$f(x_k)$	$F(x_k)$
1	11	2	2
2	12	5	7
3	13	8	15
4	14	6	21
5	15	5	26
6	16	4	30
$\sum$		30	

Fonte: Bravo, 2021.

Veja que a mediana, indicando o valor que separa os dados em dois grupos de mesmo tamanho, tem sua posição dada por:

$$i = \frac{1 + 30}{2} = 15{,}5$$

Assim, a tabela de frequência acumulada nos permite extrair, rapidamente, que a 15ª observação é de 13, enquanto a 16ª é de 14. Assim, a mediana é dada por 13,5, ou seja:

$$\frac{13 + 14}{2} = 13{,}5$$

4.2 Moda

Para os mesmos dados discutidos, a moda é facilmente extraída da Tabela 6, visto que se refere ao dado que mais aparece, ou seja, que possui a maior frequência. Assim, a moda é 13.

4.3 Média

Quando temos um conjunto de dados distribuídos em uma tabela de frequência, realizamos o cálculo da média de forma equivalente, mas com uma pequena alteração. Vejamos:

$$\bar{x} = \frac{\sum_{k=1}^{m} x_k . f(x_k)}{n}$$

Aqui, m representa a quantidade de observações semelhantes, que nesse caso é 6. Para calcularmos

$$\sum_{k=1}^{m} x_k . f(x_k)$$

podemos expandir a Tabela 6 na Tabela 7, realizando o somatório da nova coluna.

Tabela 7 – Tabela de frequência acumulada expandida para o cálculo da média dos dados do exemplo

k	x_k	$f(x_k)$	$F(x_k)$	$x_k.f(x_k)$
1	11	2	2	22
2	12	5	7	60
3	13	8	15	104
4	14	6	21	84
5	15	5	26	75
6	16	4	30	64
$\sum$		30		409

Fonte: Bravo, 2021.

Assim, verificamos que:

$$\bar{x} = \frac{\sum_{k=1}^{m} x_k.f(x_k)}{n} = \frac{409}{30} = 13{,}6$$

4.4 Características de mediana, moda e média

Vale a pena perceber algumas características da mediana, da moda e da média.

No caso da mediana, ela sempre existe e tem um único valor. Além disso, como divide os dados em dois grupos, mesmo existindo *outliers,* isto é, observações que não condizem com a maior parte dos dados, eles pouco interferem no significado geral extraído dos dados.

No caso da moda, que pode ou não existir, além de não ser única, *outliers* pouco interferem em seu resultado, visto que também há baixa frequência, sendo dificilmente escolhidos como moda.

Geralmente, dados bimodais podem indicar que as informações estão sendo coletadas de duas populações distintas, o que pode trazer indicativos estatísticos da análise.

No caso da média, os *outliers* interferem no resultado obtido. Veja, por exemplo, a média dos seguintes dados:

$$10, 11, 12, 13, 13, \bar{x} = 11{,}8$$

$$1{,}11{,}12{,}13{,}13, \bar{x} = 10$$

$$10{,}11{,}12{,}13{,}34, \bar{x} = 16$$

4.5 Exemplo

Considere os dados da Tabela 8, que traz a quantidade de moradores por casa em um determinado condomínio residencial.

Tabela 8 – Quantidade de moradores por casa em determinado condomínio residencial

Moradores	Frequência
1	10
2	19
3	5
4	2

Fonte: Bravo, 2021.

Se desejamos encontrar a média, a mediana e a moda, devemos verificar que temos um total de 36 observações, isto é., $n = 36$. Geramos a Tabela 9, com os dados de frequência acumulada e de $\sum_{k=1}^{m} x_k . f(x_k)$ para o cálculo da média.

Tabela 9 – Tabela de frequência acumulada para a quantidade de moradores por casa em determinado condomínio residencial

x_k	$f(x_k)$	$F(x_k)$	$x_k . f(x_k)$
1	10	10	10
2	19	29	38
3	5	34	15
4	2	36	8
$\sum$	36		71

Fonte: Bravo, 2021.

Aqui, notamos que a mediana é o elemento de posição:

$$i = \frac{1 + 36}{2} = 18{,}5$$

Assim, é o elemento 2. A moda é o elemento que mais aparece, que também é o 2. A média, por sua vez, é dada por:

$$\bar{x} = \frac{\sum_{k=1}^{m} x_k . f(x_k)}{36} = \frac{71}{36} = 1{,}97$$

TEMA 5 – MEDIDAS DE TENDÊNCIA CENTRAL PARA DADOS CONTÍNUOS

Para dados ordenados e organizados em classes, é necessário um cuidado para o cálculo das medidas de tendência central. Isso porque, ao tratar os dados com base nas classes, não temos informações sobre os dados individuais, como eles são coletados, visto que já passaram por algum tipo de tratamento.

5.1 Mediana

Para discutir esse cenário, considere a nota de 25 discentes matriculados em uma disciplina. Os dados já foram separados por classes (Tabela 10).

Tabela 10 – Nota de 25 discentes matriculados em uma disciplina

k	Classe	$f(x_k)$	$F(x_k)$
1	4,0 ⊢ 5,0	1	1
2	5,0 ⊢ 6,0	4	5
3	6,0 ⊢ 7,0	11	16
4	7,0 ⊢ 8,0	7	23
5	8,0 ⊢ 9,0	2	25
∑		25	

Fonte: Bravo, 2021.

Como a mediana representa o valor que divide os dados em dois grupos, inicialmente devemos verificar em que classe isso ocorre. Para isso, verificamos, na frequência acumulada, onde está o elemento de posição $\frac{n}{2}$. Nesse caso, é o elemento de posição 12,5. Note que esse resultado se encontra na terceira classe. Caso $n/2$ resultasse exatamente em 16, a mediana seria o limite superior da terceira classe, ou seja, 7. Entretanto, quando o resultado não é exato, o que é o caso, devemos encontrar a mediana a partir da seguinte aproximação:

$$\text{mediana} = x_{k(\text{inf})} + \frac{\left[\frac{n}{2} - F(x_{k-1})\right].\left(x_{k(sup)} - x_{k(inf)}\right)}{F(x_k) - F(x_{k-1})}$$

Nesse caso, note que $x_{k(inf)}$ representa o limite inferior da classe k, $x_{k(sup)}$ representa o limite superior da classe k, $F(x_{k-1})$ representa a frequência

acumulada da classe anterior à mediana, e $F(x_k)$ representa a frequência acumulada da classe que contém a mediana.

Para os dados do problema, pode-se verificar que:

$$k = 3$$
$$x_{k(inf)} = 6{,}0$$
$$x_{k(sup)} = 7{,}0$$
$$F(x_{k-1}) = 5$$
$$F(x_k) = 16$$

Assim:

$$\text{mediana} = 6{,}0 + \frac{[12{,}5 - 5].(7{,}0 - 6{,}0)}{16 - 5} = 6{,}68$$

5.2 Média

Para dados separados em classes, também precisamos **aproximar** a média amostral a partir de uma versão alternativa da versão original. Nesse caso, fazemos:

$$\bar{x} = \frac{\sum_{k=1}^{m} \bar{x}_k . f(x_k)}{n}$$

Assim, podemos calcular a média expandindo a Tabela 10 na Tabela 11.

Tabela 11 – Nota de 25 discentes matriculados em uma disciplina

k	*Classe*	$f(x_k)$	$F(x_k)$	$\bar{x}_k$	$\bar{x}_k . f(x_k)$
1	4,0 ⊢ 5,0	1	1	4,5	4,5
2	5,0 ⊢ 6,0	4	5	5,5	22,0
3	6,0 ⊢ 7,0	11	16	6,5	71,5
4	7,0 ⊢ 8,0	7	23	7,5	52,5
5	8,0 ⊢ 9,0	2	25	8,5	17,0
∑		25			167,5

Fonte: Bravo, 2021.

Assim:

$$\bar{x} = \frac{\sum_{k=1}^{m} \bar{x}_k . f(x_k)}{n} = \frac{167{,}5}{25} = 6{,}7$$

Caso haja indefinição quanto o limite inferior ou superior de uma classe, não seremos capazes de calcular a média.

5.3 Moda

Lembre-se que os dados descritos em classe são gerados a partir de um conjunto de dados que não podem ser escritos, de forma viável, individualmente. Isso porque, geralmente, tais dados não resultariam em nenhuma interpretação significativa, visto que a frequência individual é praticamente unitária.

Nesse caso, com base nos dados da tabela de frequência, podemos encontrar a **classe modal**. No caso do exemplo, a terceira classe é a modal, visto que apresenta a maior frequência. Para encontrar a moda, existem, pelo menos, dois métodos distintos. O primeiro é encontrar a média entre o limite inferior e o limite superior da classe modal, isto é:

$$\frac{6{,}0 + 7{,}0}{2} = 6{,}5$$

O segundo método utiliza a fórmula de Pearson, em que:

$$\text{moda} = 3.\,\text{mediana} - 2.\,\text{média}$$

Assim:

$$\text{moda} = 3.6{,}68 - 2.6{,}7 = 6{,}64$$

5.4 Exemplo

Considere os dados referentes aos valores armazenados por 65 famílias em poupança, descritos na Tabela 12 e separados em classes.

Tabela 12 – Valores armazenados em poupança para 65 famílias pesquisadas

k	*Valor da Poupança ($)*	$f(x_k)$
1	0,00 ⊢ 500,00	18
2	500,00 ⊢ 1.000,00	13
3	1.000,00 ⊢ 1.500,00	10
4	1.500,00 ⊢ 2.000,00	9
5	2.000,00 ⊢ 2.500,00	7
6	2.500,00 ⊢ 3.000,00	5
7	3.000,00 ⊢ 3.500,00	3
$\sum$		65

Fonte: Bravo, 2021.

Para os cálculos de moda, média e mediana, iremos expandir a Tabela 12 na Tabela 13.

Tabela 13 – Tabela ampliada com dados sobre os valores armazenados em poupança para 65 famílias pesquisadas

k	*Valor da Poupança ($)*	$f(x_k)$	$F(x_k)$	$\bar{x}_k$	$\bar{x}_k.f(x_k)$
1	0,00 ⊢ 500,00	18	18	250,00	4.500,00
2	500,00 ⊢ 1.000,00	13	31	750,00	9.750,00
3	1.000,00 ⊢ 1.500,00	10	41	1.250,00	12.500,00
4	1.500,00 ⊢ 2.000,00	9	50	1.750,00	15.750,00
5	2.000,00 ⊢ 2.500,00	7	57	2.250,00	15.750,00
6	2.500,00 ⊢ 3.000,00	5	62	2.750,00	13.750,00
7	3.000,00 ⊢ 3.500,00	3	65	3.250,00	9.750,00
$\sum$		65			81.750,00

Fonte: Bravo, 2021.

Note que a mediana será o dado que separa o conjunto de dados em dois grupos com a mesma quantidade. Como são 65 elementos, será o elemento de posição 33. Observando os dados de frequência acumulada, notamos que se encontra na terceira classe.

$$\text{mediana} = x_{k(\text{inf})} + \frac{\left[\frac{n}{2} - F(x_{k-1})\right].\left(x_{k(sup)} - x_{k(inf)}\right)}{F(x_k) - F(x_{k-1})}$$

$$x_{k(inf)} = 1.000{,}00$$

$$x_{k(sup)} = 1.500{,}00$$

$$F(x_k) = 41$$

$$F(x_{k-1}) = 31$$

$$n = 65$$

$$\text{mediana} = 1.000{,}00 + \frac{\left[\frac{65}{2} - 31\right].(1.500{,}00 - 1.000{,}00)}{41 - 31} = 1.075{,}00$$

No caso da média, fazemos:

$$\bar{x} = \frac{\sum_{k=1}^{m} \bar{x}_k . f(x_k)}{n} = \frac{81.750{,}00}{65} = 1.257{,}69$$

Para a moda, identificamos a classe modal. No caso, é a primeira classe. Pelo primeiro modo, encontramos a moda dada por:

$$\frac{0 + 500{,}00}{2} = 250{,}00$$

No segundo modo, obtemos:

$$3 \cdot 1.075{,}00 - 2 \cdot 1.257{,}69 = 709{,}62$$

FINALIZANDO

Aprendemos como realizar as medidas de tendência central para os diversos tipos de dados.