5

Medidas de resumo

SAIBA MAIS

"Numa comparação grosseira, é como se a mensuração fosse o processo de fotografar e medida, a fotografia resultante" (COSTA, 2004, p. 36).

Seção 1: Introdução

É preciso iniciar nossa última Unidade de estudo, fazendo a importante distinção entre **mensuração** e **medida**. **Mensuração** é processo do qual resulta uma **medida;** medida é valor, número resultante do processo de mensuração.[42] Medir algo é, portanto, atribuir um número.

Há quatro níveis de medidas:

Os níveis de medidas

Níveis	Variáveis
1º nível	**Nominal**, pois, apesar de expressa em números, é apenas um nome. Exemplos: número de telefone, RG, CIC, CPF etc. Esses números não são objetos de operações matemáticas.
2º nível	**Ordinal**, quando os itens podem ser colocados em ordem de grandeza. As notas escolares são um bom exemplo desse nível.
3º nível	**Intervalar**. Aqui, faz sentido quantificar. Na escala intervalar, *adição* e *subtração* são permitidas (mas *multiplicação* e *divisão* não). Escalas termométricas são um bom exemplo.
4º nível	**Racional** ou **de razão**. Nesse nível, *todas as operações matemáticas* são permitidas. Medidas tomadas com régua, fita métrica, balança, litro são bons exemplos, pois o *medido* corresponde ao *real* e não a uma correspondência.

Quadro 3: Níveis de medidas
Fonte: COSTA (2004, p. 36-40).

Pelos níveis de medidas acima, é fácil notar que um professor, ao atribuir uma nota bimestral a um aluno, está, na verdade, lidando com uma variável ordinal. Assim, ele está, apenas, indicando em uma escala, por exemplo, de 0 a 10, onde o aluno se encontra. Essa nota bimestral não é, portanto, uma medida racional, isto é, não possui a qualidade de uma medida obtida com uma fita métrica onde o resultado expressa a realidade.

Além disso, ao final do ano, os professores costumam tirar média das notas bimestrais. Isso é matematicamente sem sentido, pois, as notas não são reais, isto é, não representam a totalidade do conhecimento do aluno. Sendo assim, a

42 COSTA (2004, p. 36).

matemática não autoriza a operação com variáveis *ordinais*. Os professores costumam tirar média de notas. Por tradição e desconhecimento, não sabem que a Matemática não autoriza esse tipo de cálculo. Imagine que a nota de um aluno no 1º bimestre seja 5, o que isso significa? Significa que no processo de mensuração a resultante pode ser expressa pelo número 5 (medida). Isto é, numa escala de 0 a 10, o aluno pode ser colocado no posto 5. Somente isso, trata-se de uma variável ordinal, pois, pode ser colocado em uma ordem (ordem 5, na escala de 0 a 10). Não tem significado algum realizar operações com as notas do 1º e 2º bimestre para produzir uma resultante final. (COSTA, 2004).

IMPORTANTE

Esse é um problema que, a meu ver, tarda em ser enfrentado. Mas fique sabendo que

"existe, hoje, embora com pouca divulgação entre nós, uma teoria capaz de dar conta dos problemas apontados: trata-se da Teoria de Resposta ao Item (TRI), extremamente complexa e fortemente dependente de conhecimentos probabilísticos. Pouco a pouco, essa teoria vai ganhando espaço, graças, entre outros fatores, à rápida evolução de recursos computacionais. Em países como Estados Unidos, Holanda e Espanha, a TRI já conta com forte adesão" (COSTA, 2004, p. 40).

A Teoria de Resposta ao Item (TRI) já possui vasta aplicação no Brasil. Consulte o endereço eletrônico abaixo, para ver a aplicação da TRI na produção de indicadores socioeconômicos. http://www.scielo.br/pdf/pope/v25n1/24252.pdf

Sem perder de vista a importante diferenciação entre **mensuração** e **medida**, passemos ao estudo das **medidas**. Em Estatística Descritiva,[43] alguns conceitos são fundamentais para analisarmos os dados, se quisermos uma análise responsável. As medidas podem ser divididas em:[44]

a) medidas de tendência central (média, moda e mediana);

b) medidas de dispersão (desvio-padrão e coeficiente de variação);

c) medidas de posição (*quartis*, *decis* e *percentis*).

Como a finalidade dessas medidas é resumir as informações, essas medidas são chamadas **medidas de resumo**.[45] Por essa

43 Ver Seção 2: Estatística Descritiva e Estatística Indutiva ou Inferencial, p. 42.
44 Segundo PEREIRA (2004, p. 11)
45 PEREIRA (2004).

razão, a *média*, por exemplo, é um valor que resume as informações de um conjunto maior de dados. Por exemplo, "quando um jornalista diz na TV que o salário médio do brasileiro é algo que gira em torno de R$ 450,00 é porque muitos salários foram considerados, em todo o país, e o valor de R$ 450,00 expressa esse conjunto de salários." (PEREIRA, 2004, p. 11).

No nosso estudo, nesta **Unidade V**, enfocaremos algumas dessas medidas. Começaremos com as *medidas de tendência central*; nessa parte, *seção 2*, estudaremos a *média* e a *média aritmética ponderada*, a *mediana*, a *moda* e, por fim, a *relação entre média, mediana e moda*. Depois, na *seção 3*, estudaremos as *medidas de dispersão*, especialmente, os conceitos de *dispersão* e *variação*, *desvio padrão* e *coeficiente de variação*. Por último, na *seção 4*, estudaremos as *medidas de posição* conhecidas como *quartis*, *decis* e *percentis*.

Bom estudo a todos!

Seção 2: Medidas de Tendência Central

A **média** *é a mais importante das medidas estatísticas.*

A média é um valor típico de um conjunto de dados que tende a se localizar em um ponto central. Por essa razão, medidas com essa tendência são também denominadas **medidas de tendência central**. Vários tipos de médias podem ser definidos, sendo as mais comuns a **média aritmética**, a **média aritmética ponderada**, a **mediana** e a **moda**.[46]

Média Aritmética

Para se calcular a **média aritmética**, ou simplesmente média, de um conjunto depende do tipo de dados. Para **dados não-agrupados** é muito simples. Observe o exemplo:

46 Existem outras médias, tais como a Média Geométrica e a Média Harmônica, que não serão estudadas por nós.

As notas de um estudante em seis provas foram 8,4; 9,1; 7,2; 6,8; 8,7 e 7,8. Determinar a média aritmética das notas.

Solução:

$$\textit{Média Aritmética} = \frac{8{,}4 + 9{,}1 + 7{,}2 + 6{,}8 + 8{,}7 + 7{,}8}{6} = \frac{48}{6} = 8{,}0$$

Figura 24: Média Aritmética: Exemplo
Fonte: Adaptado de SPIEGEL (1975, p. 80)

IMPORTANTE

Observe que, na prática, o que realizamos foi *somar todas as notas* (48) e dividir pela *quantidade total de notas* (6).

Já que os números servem para "resumir" as informações, que tal diminuir a quantidade de dados por meio de fórmulas?

Estatísticos e matemáticos gostam muito de fórmulas. Isso se deve ao fato de elas "economizarem" quantidade de informações. Eles são muito práticos.

SAIBA MAIS

***Soma, Total ou $\sum$, são maneiras diferentes de representar a mesma coisa: a* soma total.**

Assim, ao invés de escreverem *"média aritmética"*, na resolução de um exercício, eles utilizam a letra *"x"*, com uma barra em cima ($\bar{x}$); cada elemento do conjunto eles chamam de *"x_i"*; todos os elementos, *"n"* e, para representarem uma soma de todos os elementos de um conjunto, eles utilizam o símbolo chamado *"somatório"* ($\sum$).

Dessa maneira, a fórmula para a *média aritmética* fica assim representada:

$$\bar{x} = \frac{\sum x_i}{n}, \text{ onde } \begin{cases} \bar{x} = \textit{Média Aritmética} \\ x_i = \textit{Valores da Variável} \\ \sum x_i = \textit{Soma Total dos Valores da Variável} \\ n = \textit{Número Total de Valores} \end{cases}$$

ATENÇÃO

Fórmula 1: Média Aritmética

Vamos realizar outro exercício para *dados não-agrupados* utilizando, desta vez, a Fórmula 1. Considere as aprovações na disciplina de matemática do professor João, de uma turma, nos últimos anos, representadas na série histórica abaixo:

Tabela 23: Série Histórica: Exercício

Total de aprovados em matemática – Professor João				
2001	**2002**	**2003**	**2004**	**2005**
35	38	32	40	37

Pergunta-se: qual a média aritmética dos aprovados nessa disciplina, no período considerado?

Solução:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{35+38+32+40+37}{5} = 36,4$$

Então, $\bar{x}$= 36,4.

Você notou que não existe o número 36,4 no conjunto de dados? Quando isso acontece, dizemos que a média **não tem existência concreta.**[47] *O que esse valor significa? Significa que, considerando todas as grandezas, dentro do conjunto de dados ordenados, esse valor tende a uma posição central, por isso, a média é uma medida de tendência central.*

Vejamos, agora, como se calcula a média aritmética para **dados agrupados**. *Os dados agrupados podem se apresentar sem intervalos de classe ou com intervalos de classes.*[48]

Vamos calcular a média aritmética para dados agrupados **sem intervalos de classe**. *Considere a distribuição de freqüência relativa a 34 famílias de quatro filhos, tomando como variável o número de filhos do sexo masculino,*[49] *abaixo.*

47 CRESPO (1995, p. 80).
48 Ver Unidade 4: Distribuição de Freqüência, particularmente, a Seção 2: Distribuição de freqüência e Aprofundamento: regras para a elaboração de uma distribuição de freqüência, p. 61.
49 CRESPO (1995, p. 82).

Tabela 24: Distribuição de Freqüência: Exercício

Número de filhos do sexo masculino	
Nº de meninos	Freqüência (f_i)
0	2
1	6
2	10
3	12
4	4
	$\sum = 34$

Fonte: CRESPO (1995, p. 82)

IMPORTANTE

O levantamento foi realizado em 34 famílias, todas com 4 filhos. A coluna da esquerda, número de meninos, é a coluna indicadora. A coluna da direita, freqüência, é a coluna numérica.[50] De acordo com a Tabela de Distribuição de Freqüência, de todas as famílias em estudo, 2 famílias não possuíam meninos; 6 famílias apresentaram 1 menino; 10 famílias, 2 meninos; 12 famílias, 3 meninos e, por fim, 4 famílias possuíam 4 meninos.

Dessa forma, as freqüências são indicadoras da *intensidade* de cada valor da variável número de meninos. Esse é um caso de **ponderação,** o que nos leva a calcular a **média aritmética ponderada**, porque cada variável possui intensidade diferente.

Para o cálculo da média, precisaremos de outra Fórmula:

SAIBA MAIS

Quando na Tabela aparece, por exemplo, que para 1 menino a freqüência é 6, é o mesmo que dizer que existem 1+1+1+1+1+1 meninos ou 6 vezes 1. Viu? Ponderar nada mais é do que considerar as repetições.

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}, onde \begin{cases} \bar{x} = Média\ Aritmética \\ x_i = Valores\ da\ Variável \\ f_i = Freqüência \\ x_i f_i = Ponderação \end{cases}$$

ATENÇÃO

Fórmula 2: Média Aritmética Ponderada

50 Ver Unidade 3: Variáveis, Tabelas e Gráficos, Seção 4: Tabelas, p. 46.

O modo mais prático para calcular uma média ponderada[51] é construir na Tabela de Distribuição de Freqüência mais uma coluna com os produtos "nº de meninos" vezes "freqüência" (ou, segundo a fórmula, $x_i f_i$). Veja:

Tabela 25: Distribuição de Freqüência: Exercício: Ponderação

Número de filhos do sexo masculino		
Nº de meninos	Freqüência (f_i)	$x_i f_i$
0	2	0
1	6	6
2	10	20
3	12	36
4	4	16
	$\sum = 34$	$\sum = 78$

Agora ficou fácil. Temos, então, que:

$$\sum x_i f_i = 78 \text{ e } \sum f_i = 34$$

Logo, pela Fórmula 2:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} \Rightarrow \bar{x} = \frac{78}{34} = 2,29 \Rightarrow \bar{x} = 2,3$$

A média de 2,3 nos indica que as famílias têm em média 2 meninos e 2 meninas, sendo que existe uma tendência geral de uma leve superioridade numérica dos meninos em relação ao número de meninas.

Por fim, vamos calcular a média aritmética para *dados agrupados* **com intervalos de classes**. Quando os dados são apresentados em uma distribuição de freqüência, todos os valores incluídos num certo intervalo de classe são considerados coincidentes com o ponto médio do intervalo.[52] Para o cálculo da média aritmética ponderada, utilizamos a Fórmula 2:

51 Para falar a verdade, sempre que formos aplicar uma Fórmula, construiremos tabelas de auxílio. Desse modo, identificamos os dados da Fórmula e, depois, encontramos o resultado.

52 SPIEGEL (1975, p. 73).

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$ *, onde xi é o ponto médio da classe.*

Qual o ponto médio do intervalo de 0 até 2?
A resposta é 1.
Qual é o ponto médio do intervalo de 2 a 4?
A resposta é 3.
Viu?
Ponto médio é o ponto que está no meio do intervalo.

Dessa forma, o raciocínio é o mesmo para a média aritmética ponderada sem intervalos de classe.

Vamos realizar um exercício. Você se lembra do professor Paulo? Bem, vamos retornar às notas dos alunos dele.[53]

Tabela 26: Distribuição de Freqüência: Exercício: Ponderação: Ponto Médio

Notas dos alunos do professor Paulo			
Notas	f_i	x_i	$x_i f_i$
0 a 2	30	1	30
2 a 4	28	3	84
4 a 6	28	5	140
6 a 8	37	7	259
8 a 10	45	9	405
	$\sum = 168$		$\sum = 918$

A Tabela 26, acima, recuperou a distribuição de freqüência do professor Paulo, acrescentando, apenas, o ponto médio dos intervalos de classe (x_i) e a ponderação, isto é, o produto dos pontos médios pela freqüência ($x_i f_i$). Bem, sabemos, portanto, que:

$$\sum x_i f_i = 918 \text{ e } \sum f_i = 168$$

Logo, utilizando a Fórmula 2 para o cálculo da média aritmética ponderada, temos que:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} \Rightarrow \bar{x} = \frac{918}{168} = 5,46 \Rightarrow \bar{x} = 5,5$$

53 Tabela 22, p. 70.

REFLEXÃO

O que isso indica? Indica que temos que mudar nossa opinião sobre o trabalho do professor Paulo. E por quê? Porque a "análise" que realizamos, naquele momento, nos levou a afirmar "que o resultado do trabalho do professor Paulo é satisfatório, pois, há mais alunos com notas acima do intervalo de 4 a 6 do que abaixo dele". Você se lembra?[54]

E o que mudou de lá para cá? Bem, a média das notas do professor sendo 5,5, indica que praticamente, metade dos alunos do professor estão com notas abaixo de 5,0, com uma tendência para notas acima de 5,0. Ora, isso não parece tão satisfatório, não é mesmo? Diante disso, não é ilícito afirmar que o professor Paulo precisa rever seus processos de mensuração.[55]

PRATIQUE

Calcule a média dos acidentes de trânsito, na Região Centro-Oeste, em 2002.

Tabela 27: Vítimas de Acidentes de Trânsito, por 10.000 veículos, em 2002

Unidade da Federação	Vítimas de acidentes
Distrito Federal	11.256
Brasília	6.747
Goiás	22.383
Goiânia	9.567
Mato Grosso	-
Cuiabá	-
Mato Grosso do Sul	7.346
Campo Grande	3.071

Fonte: Adaptado de Anuário Estatístico de Acidentes de Trânsito (2002)

54 Ver p. 70.
55 Sobre *mensuração* e *medida*, ver Seção 1: Introdução desta Unidade, p. 76.

Mediana e Média

Em um conjunto ordenado, o ponto central que divide esse conjunto em dois subconjuntos com o mesmo número de elementos chama-se **mediana**. Aqui, diferentemente da *média* (que nos fornece a concentração dos dados), a *mediana* nos fornece a posição que divide, exatamente, um conjunto em função da quantidade de seus elementos. Por exemplo:

IMPORTANTE

Vamos considerar o conjunto dos números

3, 4, 4, 5, 6, 8, 8, 8, 10

Quem está no meio do conjunto?

6

Então, os elementos antes de 6 são:

3, 4, 4 e 5

E depois de 6:

8, 8, 8 e 10

Observe que temos a mesma quantidade de elementos antes e depois de 6. A **mediana** indica isso: o número que divide o conjunto ao meio, isto é, a quantidade antes e depois dele é a mesma. Assim,

A **mediana** *é [...] definida como o número que se encontra no centro de uma série de números, estando estes dispostos segundo uma ordem. Em outras palavras, a mediana de um conjunto de valores, ordenados segundo uma ordem de grandeza, é o valor situado de tal forma no conjunto que o separa em dois subconjuntos de mesmo número de elementos. (CRESPO, 1995, p. 93).*

Não se esqueça que, para fazer isso, é preciso que os elementos estejam em um rol, isto é, apresentem-se em uma ordem crescente ou decrescente.

Para **dados não agrupados**, como no exemplo acima, calcula-se a mediana de duas maneiras:

1) quando os dados forem de número ímpar, basta encontrar o *ponto central*, isto é, encontrar o valor que antes dele e depois dele, tenham o mesmo número de elementos;
2) quando os dados forem de número par, não haverá um ponto central. Nesse caso, calcula-se o *ponto médio* dos dois valores centrais, com a ajuda da média aritmética.

Considere o conjunto:[56] 145, 68, 1, 2, 6, 5, 4, 3, 4, 8. Vamos calcular a *média* e a *mediana (md)*. A primeira coisa a fazer, nunca se esqueça, é colocar os elementos em ordem:

1, 2, 3, 4, 4, 5, 6, 8, 68, 145.

Efetuando os cálculos:

Média

Aplicando a Fórmula 1, temos:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1+2+3+4+4+5+6+8+68+145}{10} = 24,6$$

Mediana

Para conjunto de dados par, realizar a média dos dois pontos centrais:

$$md = \frac{4+5}{2} = 4,5$$

Observe que a *média* é muito diferente da *mediana*. Média igual a 24,6 significa que os dados do conjunto se concentram em torno desse número, isto é, "o problema da média é que *ela é afetada pelos grandes valores*" (PEREIRA, 2004, p. 19)[57]. Com o cálculo da *mediana (md)* igual a 4,5, podemos afirmar que metade dos valores está abaixo de 4,5 e, portanto, são muito baixos.

Embora ambas as medidas sejam de tendência central (ou seja, representem pontos que tendem para o centro dos dados), no nosso caso, os valores do conjunto estão mais próximos de 4,5 do que de 24,6, não concorda? Por isso dizemos que a média leva em conta os valores e a mediana não.

56 PEREIRA (2004, p. 20).

57 Um exemplo dessa importante informação: dizer que a *média* dos salários de três amigos meus é de R$ 1.900,00 não me indica quase nada, pois, eles podem receber R$ 350,00, R$ 350,00 e R$ 5.000,00. O que isso prova? Prova que a média é afetada pelos grandes valores.

Se os dados estão **agrupados**, para calcular a mediana utilizamos a fórmula:

$$\sum \frac{f_i}{2}$$

Fórmula 3: Mediana

Observe que, para freqüência, utilizamos o símbolo f_i. Quando queremos nos referir à freqüência acumulada, utilizamos F_i.

No caso de **dados agrupados sem intervalos de classe**, como é o caso da Tabela 28, abaixo, podemos utilizar um recurso que nos auxilia a calcular a *mediana*: a coluna de freqüências acumuladas (F_i). **Freqüência acumulada** nada mais é do que a soma das freqüências de cada variável. Observe que para a variável "0 menino", temos freqüência 2, logo, a freqüência acumulada é 2; para a variável "1 menino", temos freqüência 6, logo, a freqüência acumulada é 8, pois, 2 (freqüência acumulada anterior) + 6 (freqüência simples); para a variável "2 meninos", temos freqüência simples igual a 10, logo, a freqüência acumulada será 8 (anterior) + 10 = 18; e assim sucessivamente. Freqüência acumulada será então, a soma das freqüências simples.

Tabela 28: Distribuição de Freqüência: Exercício: Mediana: Freqüência Acumulada

Número de filhos do sexo masculino		
Nº de meninos	f_i	F_i
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
	$\sum = 34$	

Fonte: CRESPO (1995, p. 95).

Pois bem, como calcular o ponto que divide igualmente a quantidade de valores acima e abaixo dele, ou seja, como calcular a *mediana*? Para o cálculo da mediana, aplicamos a Fórmula 3. O resultado indica que a mediana será um dos valores da coluna da esquerda (0, 1, 2, 3 ou 4) correspondente à freqüência acumulada imediatamente superior.

Vamos resolver o exercício acima. Sabemos que

$$\sum f_i = 34$$

Aplicando a Fórmula 3, temos que

$$\frac{\sum f_i}{2} = \frac{34}{2} = 17$$

Pela Fórmula 3, a mediana é 17. Na Tabela existe freqüência acumulada 17? Não. Caso existisse, aquela seria a linha em se encontraria a mediana. Mas, no caso de não existir, como proceder? Simples, veja:

As freqüências acumuladas são 2, 8, 18, 30 e 34. Qual é a imediatamente superior a 17? Isso mesmo, 18. Então, vamos destacar a linha:

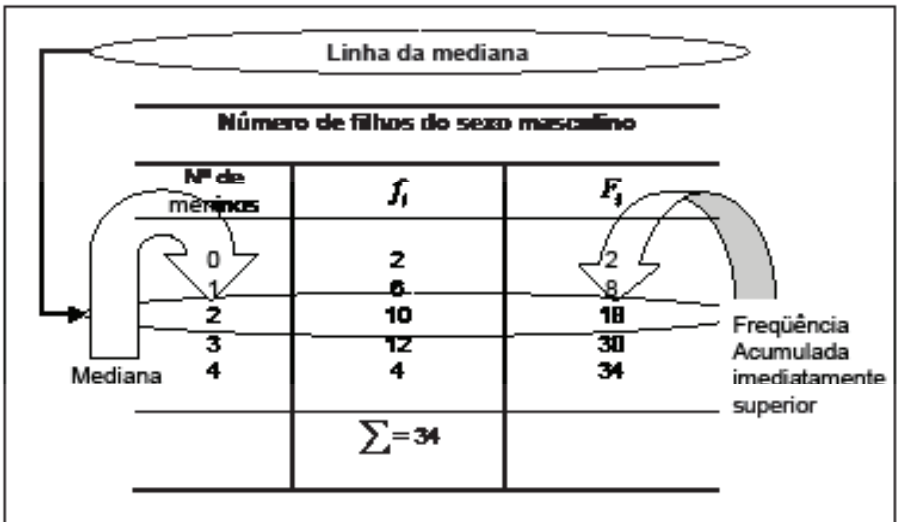

Figura 25: Linha Mediana

O número 17, conseguido com a Fórmula 3, indica que a mediana pertence à linha em que esse número se encontra. Mas como não há freqüência acumulada 17, como não é possível encontrar diretamente 17 na freqüência acumulada, então, consideramos a freqüência acumulada imediatamente superior. Nesse caso, essa freqüência é o 18. Destacamos a **linha mediana**, isto é, a linha onde a nossa mediana procurada se encontra. A mediana é, portanto, 2.