

UNIDADE I – A ESTATÍSTICA E SEUS MÉTODOS

1.1 - Introdução Histórica

Os homens desde a antiguidade faziam registros de dados que consideravam informações importantes, como o número de habitantes, de nascimentos e de óbitos, faziam estimativas do estoque de alimentos necessários a sua sobrevivência durante o inverno. Além da finalidade social e econômica, existe também, por exemplo, a bélica. Em guerras, era extremamente importante avaliar o armamento e o número de guerreiros disponíveis para a luta, tanto da própria tribo, quanto da tribo adversária.

A partir do século XVI, surgiram as primeiras tábuas e registros organizados de fatos sociais do tipo: batizados, casamentos, nascimentos, etc.

No século XVIII, Godofredo Achenwall denominou Estatística o estudo matemático de catalogação de dados numéricos coletivos. As tabelas tornaram-se mais completas, surgiram às representações gráficas de probabilidades. Com base no desenvolvimento da teoria das probabilidades, verificou-se que a estatística poderia ser utilizada para tirar conclusões e tomar decisões baseadas na análise de dados.

A Estatística é a ciência que estuda os métodos de coleta, análise, interpretação e apresentação de dados experimentais. A Estatística denominada Estatística Descritiva cuida da organização e descrição dos dados e a inferência estatística se refere à análise e interpretação dos mesmos. As técnicas de inferência estatística usam conceitos de probabilidade e distribuições de probabilidade. Então as técnicas estatísticas são usadas em pesquisa de todas as áreas do conhecimento (exatas, humanas e biológicas) que envolvam coleta e análise de dados.

1.2 – Estatística e Métodos Estatísticos

A estatística fornece métodos para a coleta, organização, descrição, análise e interpretação de dados. Os resultados podem ser utilizados para planejamentos, tomadas de decisões ou formulações de soluções.

Por exemplo: Queremos obter informações sobre a preferência de meios extra-hoteleiros (como camping), investigar a importância da acessibilidade, da qualidade das instalações ou das facilidades de meios de pagamento ou a preferência dos eleitores para a votação para presidente nas eleições de 2010.

O estudo estatístico inicia-se com o planejamento da pesquisa, que representa a organização do plano geral do trabalho que estabelece os objetivos e a utilização dos meios estatísticos.

A coleta de dados vai obter informações sobre a realidade a ser estudada. Na educação, os instrumentos mais utilizados para a coleta de dados são os questionários e as entrevistas.

Após a coleta dos dados, é necessário classificá-los, isso significa estabelecer categorias que permitem a reunião das informações coletadas.

O simples levantamento da informação e sua apresentação em tabelas estatísticas, limitando-se a apresentar os dados e gráficos de dado fenômeno, levam o Nome de estatística descritiva.

É necessária que a informação seja organizada para sua posterior análise e interpretação. A análise dos dados permite caracterizar como os elementos se distribuem, quais os valores de tendência central, qual a variabilidade, qual a relação entre as variáveis e a verificação das semelhanças e diferenças entre os elementos.

A interpretação dos dados consiste nas comparações, ligações lógicas, estabelecimento de princípios e generalizações que indiquem as desvantagens dos resultados obtidos.

Estatística Descritiva: coleta, organização e descrição dos dados.

Estatística Indutiva (ou Inferencial): análise e interpretação dos dados.

1.3 – População ou Universo Estatístico

É o conjunto da totalidade de indivíduos que apresentam uma característica comum, cujo comportamento se quer analisar (inferir). Ou ainda, podemos dizer que a população é caracterizada por ser o conjunto dos elementos que formam o universo de nosso estudo.

Quanto ao número de elementos, a população pode ser finita ou infinita. Por exemplo, a população constituída por todos os alunos da UFPA - Cametá que estavam presentes em um determinado evento, durante um certo dia, é uma população finita, enquanto que a população constituída de todos os resultados (cara ou coroa) em sucessivos lances de uma moeda é uma população infinita.

Em grandes populações, torna-se interessante a realização da chamada amostragem, que ocorre na impossibilidade de colher informações sobre a população total.

1.3.1 – Amostra

É um subconjunto finito de uma população. A amostra é uma parte da população, necessária quando se tratar de uma população com uma quantidade muito grande de elementos. A amostra permite que se trabalhe com uma parte dos elementos de uma população, quando existe dificuldade de fazer a pesquisa com todos os elementos da população.

Exemplo: Queremos obter informações sobre a audiência do programa “Globo Ciência”.

A população corresponde ao número total de domicílios que possuem TV.

A amostra é o conjunto dos domicílios que serão visitados.

1.3.2 – Variáveis

São as características que podem ser observadas (ou medidas) em cada elemento da população, ou ainda, é um conjunto de resultados possíveis de um fenômeno.

A variável pode ser:

a) Qualitativa: quando os valores são expressos por uma qualidade ou atributo.

Ex.: sexo, cor da pele, estado civil.

b) Quantitativa: quando os valores são expressos por números. Ex.: idade, salários, notas da avaliação, etc.

A variável quantitativa pode ser contínua ou discreta:

- Variável Contínua: assume inúmeros valores entre dois limites. Ex.: Peso das malas num aeroporto.
- Variável Discreta: assume apenas os valores de um conjunto enumerável. Ex.: Número de alunos que freqüentam diariamente o campus da UFPA de Cametá.

1.4 – Números aproximados e arredondamento de dados

A norma **NBR 5891** da Associação Brasileira de Normas Técnicas (ABNT) estabelece as regras fixas de arredondamento na numeração decimal, em uso na atualidade. Estas regras estão de acordo com a resolução **886/66** do IBGE.

a) Quando o primeiro algarismo a ser abandonado no arredondamento é 0, 1, 2, 3 ou 4, fica inalterado o último algarismo a permanecer.

Exemplos:

1) 25, 32 → 25,3 2) 409,04 → 409,0 3) 3,021 → 3,02

b) Quando o primeiro algarismo a ser abandonado no arredondamento é 6, 7, 8 ou 9, aumenta-se uma unidade ao último algarismo a permanecer.

Exemplos:

1) 19, 417 → 19,42 2) 2,09 → 2,1 3) 2,99 → 3,0

c) Quando o primeiro algarismo a ser abandonado no arredondamento é 5, há dois procedimentos:

- Se após o algarismo 5 seguir em qualquer casa um número diferente de 0, aumenta-se em uma unidade o algarismo que o antecede o 5;

Exemplos:

1) 237,85001 → 237,9

2) 5,5256 → 5,53

- Se após o algarismo 5 não seguir (em qualquer casa) um número diferente de 0, ao algarismo que antecede o 5 será acrescentada uma unidade, se for ímpar, e permanecerá como está, se for par.

Exemplos:

1) 246,35 → 246,4

2) 246,85 → 246,8

3) 12,1250 → 12,12

Observação:

Nos softwares de computadores (como o Excel) e calculadoras científicas, porém, não é aplicado o critério indicado no item c. Nesses casos, se o primeiro algarismo a ser abandonado for o algarismo 5, o arredondamento será feito com o aumento de uma unidade ao algarismo que antecede o 5.

Exemplos:

1) 246,35 → 246,4

2) 246,85 → 246,9

3) 12,1250 → 12,13

1.5 – Cálculo de Porcentagem

Porcentagens são razões em que um valor total está associado a uma quantidade de 100% e, por meio de uma regra de três, podemos estabelecer a correspondência entre uma parcela do valor total e seu valor percentual.

Total → 100%

Parcela → X%

Exercícios

Nos exercícios 1 ao 5, responda os itens i, ii e iii abaixo:

- Estabeleça a variável em cada caso;
 - Classifique as variáveis em qualitativas ou quantitativas;
 - Diga quais das variáveis são contínuas e quais são discretas.
- A cor dos olhos dos alunos da turma de Pedagogia – 2008 da UFPA- Cametá.
 - Os salários dos funcionários da UFPA-Cametá.
 - A quantidade de alimento, em gramas, ingerida por estudante em um determinado colégio.
 - O sexo dos filhos de casais residentes em Mocajuba-Pa.
 - O número de pessoas da terceira idade, durante um ano, no turismo de Fortaleza.
 - Escreva cada número com arredondamento para décimos.
 - 238,4575
 - 71,21
 - 4,8976
 - 0,03424

- 7) Escreva cada número com arredondamento para centésimos.
- a) 149,1209
 - b) 0,246
 - c) 12,0001
 - d) 0,03498
- 8) Escreva cada número com arredondamento para a unidade.
- a) 0,03516
 - b) 0,3
 - c) 17,50015
 - d) 0,04458
- 9) Calcule 12% de R\$ 2400,00.
- 10) R\$36,00 correspondem a 24% de uma quantidade. Qual o valor da quantidade total?
- 11) Qual a porcentagem que R\$45,00 representam num total de R\$2560,00.
- 12) Calcule a porcentagem que 93,02 representa em 458,50.

1.6 – Amostragem

Existem métodos ou técnicas adequadas para recolher amostras, de forma a garantir (tanto quanto possível) o sucesso da pesquisa e dos resultados.

Devemos estabelecer um número mínimo de elementos para compor a amostra. Essa quantidade não deve ser menor que 10% do total de elementos da população. Por exemplo, numa população de 500 elementos, devemos, por um critério de seleção, selecionar um mínimo de 50 elementos (10% de 500) para compor a amostra.

Mas que método devemos utilizar para realizar a escolha?

Podemos recorrer a diferentes formas de amostragem: amostragem aleatória simples, amostragem sistemática e amostragem estratificada proporcional.

1.6.1 – Amostragem Aleatória Simples

Nesse tipo de amostragem, a primeira providência a ser tomada é a elaboração de uma lista com os 500 nomes dos elementos da população, numerados de 1 a 500 para serem submetidos a um sorteio. Bolas ou cartões, também numerados de 1 a 500, são então colocados em uma urna, na qual os números devem ser bem misturados. Em cada etapa do sorteio, todo número ainda não escolhido tem a mesma probabilidade de ser sorteado.

Esse processo não é muito prático para grandes populações. Nesse caso, podemos trabalhar com uma numeração de 0 a 9, sorteando os números por meio de blocos de três algarismos e tomando cuidado de repor na urna todo algarismo dela retirado. Como temos dez algarismos, cada um deles tem $\frac{1}{10}$ de

probabilidade de aparecer em determinada posição. Sempre que um bloco que não exista na população, ele será descartado.

Suponhamos que, ao efetuar o sorteio, obtenhamos os algarismos:

2	4	3	5	6	4	7	2	0	0	3	5	8	1	1	0	0	5
1	9	8	6	4	3	5	2	4	7	8	9	7	7	6	5	4	2
2	3	0	1	2	1	1	6	7	8	9		7	7	6	5	4	2
2	2	8	8	1	9	0	0	6	0	7	2	1	0	5	6	4	3

Agrupando-se em blocos de três algarismos, teremos os números:

243	564	720	035	811	005	198	643	524	789	776	542	230
121	167	891	034	567	228	819	006	072	105	643		

Desses números sorteados devemos descartar 811, 891 e 819, porque não são elementos da população, e 643 porque já foi selecionado.

E assim, sucessivamente, vamos procedendo ao sorteio, até conseguirmos selecionar os 50 elementos da amostra.

1.6.2 – Amostragem Sistemática

Continuemos a considerar a população de 500 elementos de nossa lista numerada. Para organizar uma amostragem sistemática, sorteamos um número de 1 a 10, ao acaso. Suponhamos que tenha sido obtido o número 6. Ele será o primeiro elemento da amostra, e os demais serão determinados em intervalos de dez unidades. Assim, nossa amostra será:

6	16	26	36	46	56	66	76	86	96	106	...	406
---	----	----	----	----	----	----	----	----	----	-----	-----	-----

A amostragem sistemática é simples de ser realizada e, no caso de amostras muito grandes, acarreta economia de tempo e dinheiro.

Exemplo:

Na escola Professor Sebastião Torres, deseja-se fazer um estudo sobre o peso dos alunos de 7 anos de idade. Sabendo-se que há 120 crianças na faixa dos 7 anos, selecione uma amostra por:

- amostragem aleatória simples;
- amostragem sistemática.

(Devemos considerar todas as crianças entre 6 e meio e 7 anos e meio porque, em geral, não encontramos muitas que tenham exatamente 7 anos no dia da pesquisa)

Solução:

a) Para proceder a uma amostragem aleatória simples, devemos:

- Elaborar uma lista com os 120 nomes das crianças na faixa dos 7 anos, numerados de 1 a 120;
- Sortear 12 números (10% de 120).

Logo as 12 crianças cujos nomes correspondem aos números sorteados constituem a amostra procurada.

b) Para conseguirmos uma mostra, por meio de amostragem sistemática, devemos proceder da seguinte forma:

- Elaborar uma lista dos nomes, numerando-os de 1 a 120;
- Sortear um número de 1 a 10.

Se o número sorteado for 5, por exemplo, nossa amostragem sistemática procurada será:

5 15 25 35 45 55 65 75 85 95 105 115

1.6.7 – Amostragem Estratificada Proporcional

A amostragem estratificada proporcional é recomendada quando existe uma divisão natural da população em grupos com números de elementos diversos. Vejamos o exemplo abaixo:

Exemplo:

Suponhamos que na mesma escola do exemplo anterior as 120 crianças na faixa de 7 anos de idade estejam distribuídas em cinco classes, com quantidades diferentes de alunos. A primeira série **A** tem 20 alunos com 7 anos, a primeira **B** tem 15, a **C** tem 35, a **D**, 30 e a **E** tem 20. Como faríamos a mesma seleção do primeiro exemplo?

Solução:

Podemos, nesse caso, sortear os nomes em quantidades proporcionais ao número de crianças com 7 anos de cada classe, considerando as porcentagens em relação ao conjunto total (120 elementos). As crianças sorteadas de cada classe constituem um **estrato** da amostra, que é uma amostra estratificada proporcional.

Neste problema, a população é de 120 pessoas distribuídas em cinco classes. Sabemos que a amostra deve ter, no mínimo, 12 (10% de 120) elementos. Em primeiro lugar, precisamos calcular a porcentagem de crianças com 7 anos em cada classe. Organizando os dados numa tabela, temos:

Classe	População	%
A	20	16,7
B	15	12,5
C	35	29,1
D	30	25,0
E	20	16,7
Total	120	100

Estatística Aplicada à Educação

A primeira série **A**, com 20 alunos, tem 16,7% dos elementos da população, logo 16,7% dos elementos da amostra sairão dessa classe. A primeira **B** tem 12,5% dos elementos da população, logo 12,5% dos alunos da amostra serão sorteados entre os alunos dessa classe. O cálculo para as demais classes segue o mesmo raciocínio.

Como necessitamos de uma mostra de 12 elementos, devemos calcular:

A: 16,7% de 12 $\Leftrightarrow 0,167 \cdot 12 \cong 2,004 = 2$

B: 12,5% de 12 $\Leftrightarrow 0,125 \cdot 12 \cong 1,5 = 2$

C: 29,1% de 12 $\Leftrightarrow 0,291 \cdot 12 \cong 3,493 = 3$

D: 25% de 12 $\Leftrightarrow 0,25 \cdot 12 = 3$

E: 16,7% de 12 $\Leftrightarrow 0,167 \cdot 12 \cong 2,004 = 2$

Observação: Note que os dados foram arredondados para o inteiro mais próximo, já que esses números indicam a quantidade de crianças.

Colocando os dados em uma tabela, temos:

Classe	População	%	Amostra
A	20	16,7	2
B	15	12,5	2
C	35	29,1	3
D	30	25,0	3
E	20	16,7	2
Total	120	100	12

Na última coluna está representada a quantidade de elementos de cada estrato e total da amostra.

Exercícios

1) Na escola São Leopoldo, para estudar a preferência em relação a refrigerantes, sortearam-se 150 estudantes, entre os 1000 matriculados. Responda:

a) Qual é a população envolvida na pesquisa?

b) Qual tipo de amostragem foi utilizado e qual é a amostra considerada?

2) Na escola São Miguel, deseja-se fazer um estudo sobre a altura dos alunos de 10 anos de idade. Sabendo-se que há 100 crianças na faixa de 10 anos de idade, selecione uma amostra por amostragem sistemática.

3) Em uma certa cidade, quer-se estudar o interesse despertado por um programa de TV entre os alunos de 7 anos de idade das escolas de ensino fundamental. Para isso, pretende-se levantar uma amostra de 300 crianças. A partir dos dados abaixo, estratifique a amostra:

Escola	A	B	C	D	E
População	400	300	350	450	520

UNIDADE II - DISTRIBUIÇÃO DE FREQUÊNCIA

2.1 – Definições Básicas

Frequência: É a quantidade de vezes que um mesmo valor de um dado é repetido.

Dados brutos: São os dados originais ainda não numericamente organizados após a coleta.

Rol: É a ordenação dos valores obtidos (dados brutos) em ordem crescente ou decrescente de grandeza numérica ou qualitativa.

Exemplo:

A distribuição de idade das 60 crianças de um acampamento, promovido por um determinado colégio, compõe os dados brutos.

Em forma de tabela, temos:

Faixa etária das crianças do acampamento X

6	10	9	14	7	4	13	11	5	7
8	10	12	5	9	13	11	14	9	10
9	11	8	6	7	14	10	8	7	13
11	6	12	11	15	13	9	4	9	8
12	11	4	10	7	13	9	8	10	5
10	9	8	12	13	7	8	6	15	11

Nessa tabela fica difícil estabelecer em torno de qual valor tendem a se concentrar as idades das crianças, ou ainda avaliar quantos alunos estão acima ou abaixo de determinada idade. Em função disso, é conveniente a organização das idades na tabela por meio de ordem crescente (ou decrescente). Essa tabela organizada é chamada de **rol**.

Em **rol** a tabela acima fica da seguinte forma:

4	6	7	8	9	9	10	11	12	13
4	6	7	8	9	10	10	11	12	14
4	6	7	8	9	10	10	11	13	14
5	6	7	8	9	10	11	11	13	14
5	7	8	8	9	10	11	12	13	15
5	7	8	9	9	10	11	12	13	15

2.2 – Distribuição de Freqüência

A organização das idades na tabela (rol) facilita a visualização das repetições ocorridas para uma mesma idade. Dessa forma, podemos verificar a concentração de crianças em cada faixa etária e, assim, construir uma tabela onde consta para cada idade a respectiva freqüência (número de vezes em que a idade é repetida). Essa tabela recebe o nome de **distribuição de freqüência**.

Distribuição de idade das 60 crianças de um acampamento

Idade	Freqüência
4	3
5	3
6	4
7	6
8	7
9	8
10	8
11	7
12	4
13	5
14	3
15	2

2.3 – Elementos de uma Distribuição de Freqüência

2.3.1 – Classe

Caso as colunas da tabela de distribuição de freqüência contenham muitos valores elencados, podemos reduzir a quantidade desses valores elencados agrupando-os em intervalos.

Os valores da última tabela feita podem ser agrupados dispondo-se as faixas etárias em intervalos que abranjam diferenças, por exemplo, de dois anos de idade. Em vez de agrupar as idades de dois em dois anos, a escolha feita poderia ser de agrupar as idades de três em três anos, de quatro em quatro anos ou ainda um outro intervalo qualquer.

Esses agrupamentos de valores num intervalo de abrangência são chamados de **classes**. E a tabela passa a se chamar distribuição de freqüência com intervalos de classe.

Essa última tabela mostra a distribuição de frequência simples. Caso os elementos sejam agrupados em classes, a tabela passa a ser uma distribuição de frequência com intervalos de classe.

A tabela seguinte representa uma tabela de distribuição de frequência com intervalos de classe. Observamos que a primeira classe (ou o primeiro grupo) reúne as crianças com a idade de 4 a 5 anos, a segunda classe reúne as crianças com idade de 6 e 7 anos, a terceira classe reúne as crianças com idade de 8 e 9 anos, e assim por diante.

Distribuição da frequência em faixas etárias (das 60 crianças de um acampamento)

Idades	Freqüências
4 – 6	6
6 – 8	10
8 – 10	15
10 – 12	15
12 – 14	9
14 – 16	5

Algumas dúvidas podem surgir em relação a alguns valores. Uma criança com sete anos, onze meses e vinte e nove dias e uma criança com oito anos exatos devem permanecer ao mesmo grupo? Nesse nosso exemplo, não (veja as a seguir).

2.3.2 – Limite Inferior e Superior

Limite Inferior (l_i): O menor número que aparece em um determinado intervalo é o limite inferior da classe. No caso se considerarmos 4 – 6, temos que $l_1 = 4$ é o limite inferior da classe.

Limite Superior (L_i): O maior número que aparece em um determinado intervalo é o limite superior da classe. No caso se considerarmos 4 – 6, temos que $L_1 = 6$ é o limite superior da classe.

Logo podemos representar:

$$l_i \text{ – } L_i$$

Este símbolo utilizado (–) estabelece inclusão e exclusão para os valores limites no intervalo de classe.

O intervalo 4 – 6 indica inclusão do limite inferior quatro (ou seja, a partir da idade de quatro anos exatos a criança está incluída nessa classe) e indica

exclusão do limite superior (significa que a partir da idade de seis anos exatos a criança está excluída dessa classe).

Segundo a resolução 886/66 do IBGE, os intervalos de classe devem empregar o símbolo de inclusão e exclusão \pm entre os valores extremos de um intervalo.

2.3.3 – Amplitude de um Intervalo de Classe (h_i)

A amplitude de um intervalo de classe (h_i) é a diferença entre o limite superior e inferior de uma classe:

$$h_i = L_i - l_i$$

Na última tabela temos que:

$$h_1 = 6 - 4 = 2 \text{ anos}$$

$$h_2 = 8 - 6 = 2 \text{ anos}$$

$$h_3 = 10 - 8 = 2 \text{ anos}$$

$$h_4 = 12 - 10 = 2 \text{ anos}$$

$$h_5 = 14 - 12 = 2 \text{ anos}$$

$$h_6 = 16 - 14 = 2 \text{ anos}$$

Nesse caso, as amplitudes são iguais, porém, não é obrigatório que elas sejam; podemos eventualmente amplitudes diferentes $h_1 \neq h_2 \neq h_3 \neq \dots$

Embora as amplitudes possam ser diferentes, é mais conveniente que as classes mantenham amplitudes iguais, pois facilita a visualização do fato estudado e agiliza os cálculos realizados.

2.3.4 – Amplitude Total da Distribuição (ΔT)

Amplitude total de distribuição (ΔT) é a diferença entre o limite superior (máximo) da última classe e o limite inferior (mínimo) da primeira classe, logo:

$$\Delta T = L_{(m\acute{a}x.)} - l_{(m\acute{i}n.)}$$

Onde:

$L_{(m\acute{a}x.)}$ = limite superior da última classe;

$l_{(m\acute{i}n.)}$ = limite inferior da primeira classe.

Na última tabela, temos que a amplitude total da distribuição é:

$$\Delta T = 16 - 4 = 12 \text{ anos.}$$

2.3.5 – Amplitude Amostral (AA)

Amplitude amostral é a diferença entre o valor máximo e o valor mínimo dos dados na amostra:

$$AA = X_{(m\acute{a}x.)} - X_{(m\acute{i}n.)}$$

Na tabela de Faixa etária das crianças do acampamento **X** já organizada em rol (exemplo anterior), observamos que o valor da amplitude amostral é:

$$AA = 15 - 4 = 11 \text{ anos}$$

2.3.6 – Ponto Médio de uma Classe (x_i)

Ponto médio de uma classe (x_i) é ponto que, por situar-se numa posição média da distribuição de valores do intervalo de classe, divide o intervalo em duas partes iguais.

$$x_i = \frac{l_i + L_i}{2}$$

Já na última tabela, temos que o ponto médio da primeira classe é:

$$x_1 = \frac{4 + 6}{2} = 5$$

O ponto médio da segunda classe é:

$$x_2 = \frac{6 + 8}{2} = 7$$

O ponto médio da sexta classe é:

$$x_6 = \frac{14 + 16}{2} = 15$$

2.3.7 – Freqüência Simples ou Absoluta (f_i)

Freqüência simples ou absoluta é o numero de observações de um valor individual (ou de uma classe).

A tabela a seguir representa a distribuição de freqüência simples com intervalos de classe de fato observado.

Distribuição da freqüência em faixas etárias (das 60 crianças de um acampamento)

Idade	Quantidade de crianças por faixa etária (freqüência simples)
4 † 6	6
6 † 8	10
8 † 10	15
10 † 12	15
12 † 14	9
14 † 16	5
	Total = 60

A soma de todas as freqüências é $n = f_1 + f_2 + f_3 + f_4 + f_5 + f_6$, logo:
 $n = 6 + 10 + 15 + 15 + 9 + 5 = 60$ (nesse caso, $n = 60$)

Observamos que n é o número total de dados da amostra.

Podemos representar a soma de freqüências pelo símbolo de somatória, ou seja:

$$n = f_1 + f_2 + f_3 + \dots = \sum_{i=1}^k f_i$$

No exemplo:

$$n = \sum_{i=1}^6 f_i = 60$$

A tabela pode ser reescrita de forma mais adequada ao conceito de freqüência:

Distribuição em faixas etárias (das 60 crianças de um acampamento)

I	Idade	f_i
1	4 † 6	$f_1 = 6$
2	6 † 8	$f_2 = 10$
3	8 † 10	$f_3 = 15$
4	10 † 12	$f_4 = 15$
5	12 † 14	$f_5 = 9$
6	14 † 16	$f_6 = 5$
		$n = \sum f_i = 60$

2.4 – Determinação do Número de Classes

É importante a escolha do número de classes e amplitude. Se o número de observações é pequeno, devemos restringir a amplitude; por outro lado, se o número de observações é grande, as amplitudes também serão maiores.

Uma sugestão para estabelecer o número de classes é utilizar a fórmula desenvolvida pelo matemático Sturges, também conhecida como regra de Sturges.

$$i = 1 + 3,3 \times \log n$$

Sendo:

i = número de classes.

n = número total de dados.

Com base na regra de Sturges, Vaugh desenvolveu uma tabela onde resumiu as sugestões para estabelecer o número de classes a partir do número total de dados:

Sugestão para estabelecer o número de classes em função do número de dados (regra de Sturges)

n = casos observados	i = número de classes a usar (pela regra de Sturges)
1	1
2	2
3 – 5	3
6 – 11	4
12 – 22	5
23 – 45	6
46 – 90	7
91 – 181	8
182 – 362	9
363 – 724	10
725 – 1.448	11
1.449 – 2.896	12
...	...

Além da regra de Sturges, existem outras fórmulas, mas também são apenas sugestões; a decisão final depende do bom-senso pessoal de cada um para estabelecer o número de classes.

2.5 – Determinação da Amplitude do Intervalo de Classe (h)

Pode-se obter a amplitude do intervalo de classe (h) por meio da divisão da amplitude amostral (AA) pelo número de classes (i):

$$h = \frac{AA}{i} = \frac{X_{(Máx.)} - X_{(Mín.)}}{i}$$

Caso o resultado da divisão não seja um resultado exato, sugerimos que seja feito o arredondamento para um valor acima.

Supondo que o resultado tenha sido $h = 4,4$ devemos arredondar para $h = 5$.

Devemos procurar os valores adequados para os limites dos intervalos, valores que forneçam números naturais no cálculo dos pontos médios (dentro das possibilidades).

Exercício

Numa equipe de recreação juvenil, foram coletadas as alturas de 50 adolescentes, onde os dados brutos encontram-se na tabela abaixo:

Estatuta de 50 adolescentes (em cm)

156	157	150	158	142	159	160	157	160	158
151	148	156	147	156	154	149	159	148	150
145	153	144	154	161	152	154	152	153	154
154	155	152	160	149	157	156	155	156	143
159	161	162	155	155	153	145	149	152	157

- Organize os valores obtidos nessa tabela em Rol, em ordem crescente.
- Qual é a menor altura?
- Qual é a maior altura?
- Qual é a amplitude amostral da distribuição?
- Considerando que o número total de dados é $n = 50$, e pela tabela (última) anterior a esse exercício a sugestão é estabelecer sete intervalos de classes ($i = 7$). Qual é a amplitude do intervalo de classe?
- Complete a tabela de distribuição a seguir:

Distribuição da estatura de 50 adolescentes (em cm)

I	Estaturas (cm)	Freqüência (f_i)
1	142 † 145	3
2	145 †
3 † 151
4 †
5 †
6 †
7 † 163
		$n = \sum f_i = 50$

Considerando essa nova tabela de distribuição de freqüência, responda aos próximos itens:

- Qual o limite inferior da terceira classe?
- Qual o limite superior da sexta classe?
- Qual o limite inferior d sétima classe?
- Qual o limite superior da segunda classe?
- Qual a amplitude total da distribuição (ΔT)?
- Foram obtidos valores diferentes para a amplitude amostral da distribuição (AA) e amplitude total da distribuição (ΔT)? Explique por quê.

2.6 – Tipos de Freqüências

2.6.1 – Freqüência Simples ou Absoluta (f_i)

É o número de repetições de um valor individual ou de uma classe. A soma das freqüências simples resulta o número total de observações (n), ou seja:

$$\sum f_i = n$$

2.6.2 – Frequência Relativa (fr_i)

É a razão entre a frequência simples e a frequência total. A frequência relativa representa a proporção de observações de um valor (ou de uma classe) em relação ao número total de observações, o que facilita as comparações.

$$fr_i = \frac{f_i}{\sum f_i}$$

Na tabela dada anteriormente de distribuição da frequência em faixas etárias (das 60 crianças de um acampamento), consideramos a frequência absoluta da quarta classe $f_4 = 15$ e o número de observações $n = 60$. Ao calcular a frequência relativa da quarta classe, obtemos:

$$fr_4 = \frac{f_4}{\sum f_i} = \frac{15}{60} = 0,25 \text{ (esta classe representa 25\% do número total de observações).}$$

2.6.3 – Frequência Acumulada (F_i)

A frequência acumulada (F_i) é a soma de todas as frequências abaixo do limite superior de uma classe considerada.

$$F_K = f_1 + f_2 + \dots + f_{K-1} + f_K \Leftrightarrow F_K = \sum_{i=1}^K f_i$$

Também na tabela mencionada anteriormente, consideramos a frequência da quarta classe $f_4 = 15$ e número de observações $n = 60$. Ao calcular a frequência acumulada da quarta classe, obtemos:

$$F_4 = f_1 + f_2 + f_3 + f_4 = 6 + 10 + 15 + 15 = 46 \Leftrightarrow F_4 = 46$$

Isto significa que existem 46 crianças com idade abaixo de 12 anos no acampamento (12 anos é o limite superior da quarta classe).

2.6.4 – Frequência Acumulada Relativa (Fr_i)

A frequência acumulada relativa (Fr_i) é a razão entre a frequência acumulada da classe e a frequência total.

A frequência acumulada relativa representa a proporção de observações da frequência acumulada da classe em relação ao número total de observações, o que facilita as comparações.

$$Fr_i = \frac{F_i}{\sum f_i}$$

Novamente na tabela de distribuição da frequência em faixas etárias (das 60 crianças de um acampamento), consideramos a frequência absoluta da quarta classe $f_4 = 15$ e número de observações $n = 60$ e a frequência

acumulada da quarta classe $F_4 = 46$. Calculando a frequência acumulada relativa da quarta classe, obtemos:

$$Fr_4 = \frac{F_4}{\sum f_i} = \frac{F_4}{n} = \frac{46}{60} = 0,77 \Leftrightarrow Fr_4 = 0,77$$

Como base nessa mesma tabela, podemos escrever uma nova tabela e incluir nela os cálculos do ponto médio de cada classe (x_i), da frequência relativa (fr_i), da frequência acumulada (F_i) e da frequência relativa (Fr_i).

Distribuição em faixas etárias (das 60 crianças de um acampamento)

I	Idades	Frequência (f_i)	Ponto médio (x_i)	Frequência relativa (fr_i)	Frequência acumulada (F_i)	Frequência acumulada relativa (Fr_i)
1	4 † 6	6	5	0,10	6	0,10
2	6 † 8	10	7	0,17	16	0,27
3	8 † 10	15	9	0,25	31	0,52
4	10 † 12	15	11	0,25	46	0,77
5	12 † 14	9	13	0,15	55	0,92
6	14 † 16	5	15	0,08	60	1,00
		$\sum = 60$		$\sum = 1,00$		

Exercício

Em um colégio X, foi feita uma pesquisa sobre o salário recebido pelos seus funcionários, sendo consultados 130 funcionários e obtidos os resultados na tabela a seguir:

Número de salários mínimos	Trabalhadores (f_i)
1 † 3	18
3 † 5	24
5 † 7	21
7 † 9	19
9 † 11	15
11 † 13	14
13 † 15	11
15 † 17	5
17 † 19	3
	$\sum = 130$

Complete a tabela a seguir, incluindo nessa nova tabela os cálculos do ponto médio de cada classe (x_i), da frequência relativa (fr_i), da frequência acumulada (F_i) e da frequência acumulada relativa (Fr_i).

Distribuição de salários recebidos (130 funcionários de um determinado colégio X).

I	Nº de salários mínimos	Trabalhadores (f_i)	Ponto médio (x_i)	fr _i	F _i	Fr _i
1	1 – 3	18				
2	3 – 5	24				
3	5 – 7	21				
4	7 – 9	19				
5	9 – 11	15				
6	11 – 13	14				
7	13 – 15	11				
8	15 – 17	5				
9	17 – 19	3				
		$\Sigma = 130$		$\Sigma = 1,00$		

2.7 – Distribuição de Frequência sem Intervalos de Classe

Numa coleta de dados, nem sempre os dados com repetição são agrupados em intervalos de classe. Pode ocorrer que cada conjunto de repetições de um mesmo dado corresponda a um “intervalo de classe”.

Exemplo:

Na tabela a seguir, os dados representam o número de filhos por família de 30 famílias entrevistadas em determinado bairro X.

Número de filhos por família (de 30 famílias do bairro X)

0	0	1	1	2	2	2	2	3	3
0	1	1	1	2	2	2	3	3	4
0	1	1	1	2	2	2	3	3	4

De acordo com as repetições, podemos escrever a seguinte tabela de frequências:

Distribuição do número de filhos por família (de 30 famílias do bairro X).

Nº de filhos por família (X_i)	Famílias do bairro X (f_i)
0	4
1	8
2	10
3	6
4	2

Estatística Aplicada à Educação

Essa tabela pode ser completa com os cálculos da frequência relativa (fr_i), da frequência acumulada (F_i) e da frequência acumulada relativa (Fr_i).

Distribuição do número de filhos por família (de 30 famílias do bairro X).

I	X_i	f_i	fr_i	F_i	Fr_i
1	0	4	0,13	4	0,13
2	1	8	0,27	12	0,40
3	2	10	0,33	22	0,73
4	3	6	0,20	28	0,93
5	4	2	0,07	30	1,00
		$n = \sum f_i = 30$	$\sum = 1,00$		

Exercício

Uma empresa de pesquisa colheu a opinião de 1200 pessoas e elaborou um relatório especificando a preferência sobre qual Estado brasileiro essas pessoas gostariam de conhecer. Os valores constam na tabela a seguir:

Distribuição dos estados brasileiros que as pessoas gostariam de conhecer

Estados brasileiros	Preferência de pessoas (frequência)
Rio Grande do Sul	80
Minas Gerais	170
Rio de Janeiro	380
São Paulo	320
Bahia	190
Paraná	60
$n = \sum f_i = 1200$	

Inclua na tabela a seguir os cálculos da frequência relativa (fr_i), da frequência acumulada (F_i) e da frequência acumulada relativa (Fr_i).

Distribuição dos estados brasileiros que as pessoas gostariam de conhecer

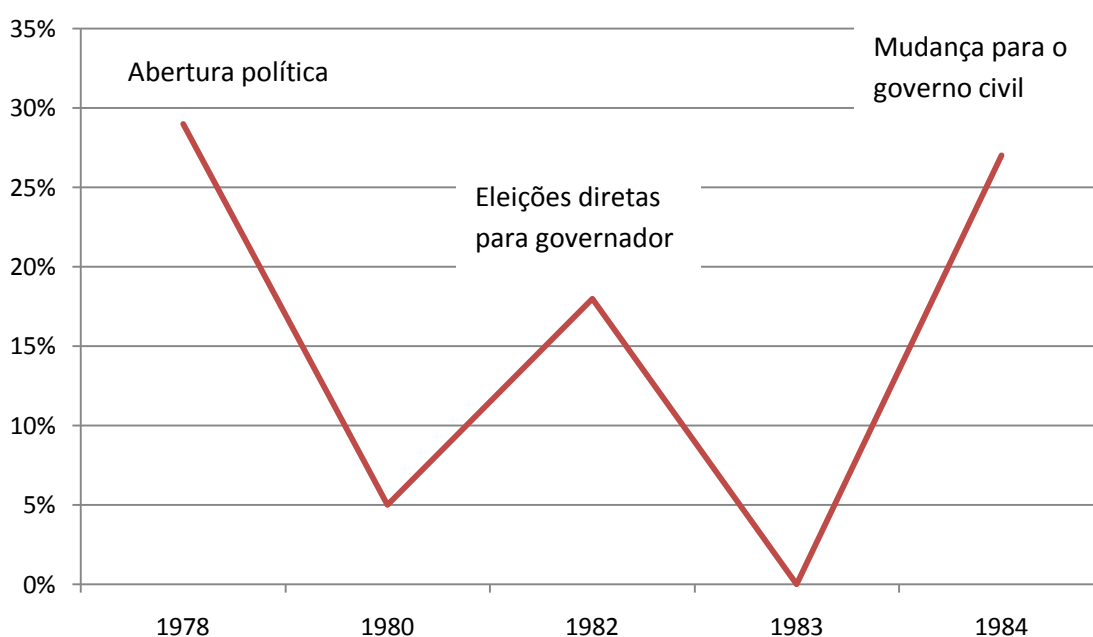
I	X_i	f_i	fr_i	F_i	Fr_i
1	Rio Grande do Sul	80			
2	Minas Gerais	170			
3	Rio de Janeiro	380			
4	São Paulo	320			
5	Bahia	190			
6	Paraná	60			
		$n = \sum f_i = 1200$	$\sum = 1,00$		

UNIDADE III - GRÁFICOS ESTATÍSTICOS

3.1 – Representação Gráfica

Dados estatísticos podem ser representados tanto por tabelas e quadros de distribuição por freqüência quanto por **gráficos**. Observe o seguinte exemplo: Exemplo:

Uma pesquisa do Instituto Gallup, realizada em vinte Estados do Brasil, registra uma alta no otimismo do povo brasileiro (publicação da revista *veja*, 02-01-1985). A “taxa de otimismo” foi estabelecida pela diferença entre o percentual de entrevistados que esperavam um ano seguinte melhor, em relação ao ano anterior, e o percentual daqueles que tinham opinião contrária a essa.



Pelo gráfico deduzimos que em 1978, ano em que o general Figueiredo foi escolhido presidente, a taxa era de 29%, caindo para 5% em 1980. Em 1982, quando foram realizadas eleições diretas para governador, o otimismo dos brasileiros chegou a 18%, caindo para zero em 1983, quando o Brasil recorreu ao FMI (Fundo Monetário Internacional). Nessa ocasião 40% dos entrevistados consideravam que o ano seguinte seria melhor, e os outros 40% que seria pior. Em 1984, o último ano do regime militar instalado em 1964, a taxa de otimismo volta a subir, atingindo 27%.

Além de revelar o fenômeno estatístico, o gráfico tem a função de facilitar sua compreensão, por meio do efeito visual imediato que lhe é próprio. Essa característica é uma vantagem que os gráficos têm sobre as tabelas, já que a impressão que eles produzem é mais rápida e viva.

A Estatística pode recorrer a vários tipos de gráfico. Os principais são os **gráficos de linha ou de curva e os diagramas de área** (que incluem os gráficos

de colunas, de barras e de setores, entre outros).

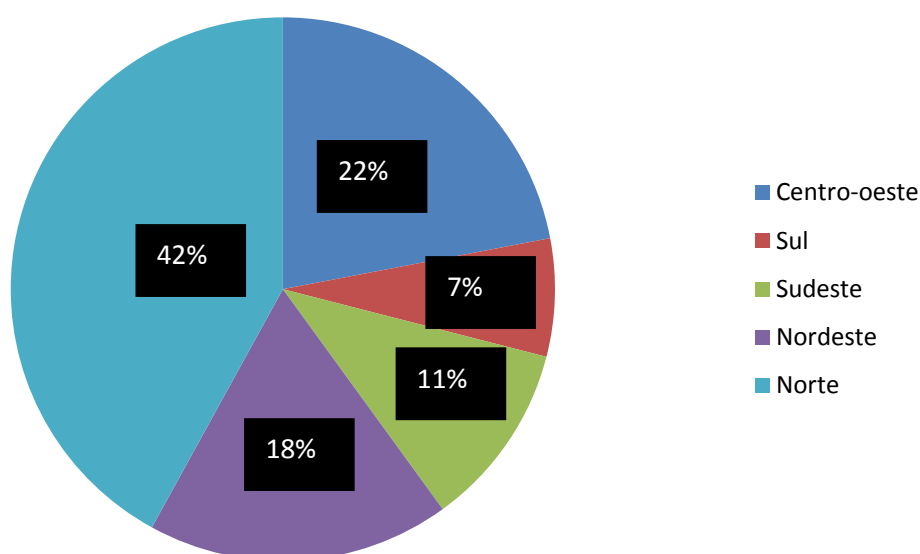
Estudaremos quatro tipos de representações gráficas: o gráfico de setores (ou “pizza”), o gráfico de barras (verticais ou horizontais), o histograma e o gráfico de linhas (poligonal).

1) Gráfico de Setores

A representação por setores consiste em dividir um círculo em partes (setores circulares), com os ângulos de medida proporcional à porcentagem da variável tabelada.

Exemplo:

ÁREA DAS REGIÕES DO BRASIL



Fonte: IBGE

Exercício

Com o objetivo de traçar um perfil dos alunos freqüentadores da biblioteca de certa universidade foram entrevistados 20 alunos, obtendo:

Tabela de Freqüência do Estado Civil

Estado Civil	Freqüência absoluta	Freqüência Relativa	Porcentagem %
Separado	1	$\frac{1}{20} = 0,05$	5
Solteiro	12	$\frac{12}{20} = 0,6$	60
Casado	6	$\frac{6}{20} = 0,3$	30
Viúvo	1	$\frac{1}{20} = 0,05$	5
Total	$n = 20$	1	100

Construa a partir dessa tabela o gráfico dos setores representando o estado civil dos entrevistados.

2) Gráfico de Barras

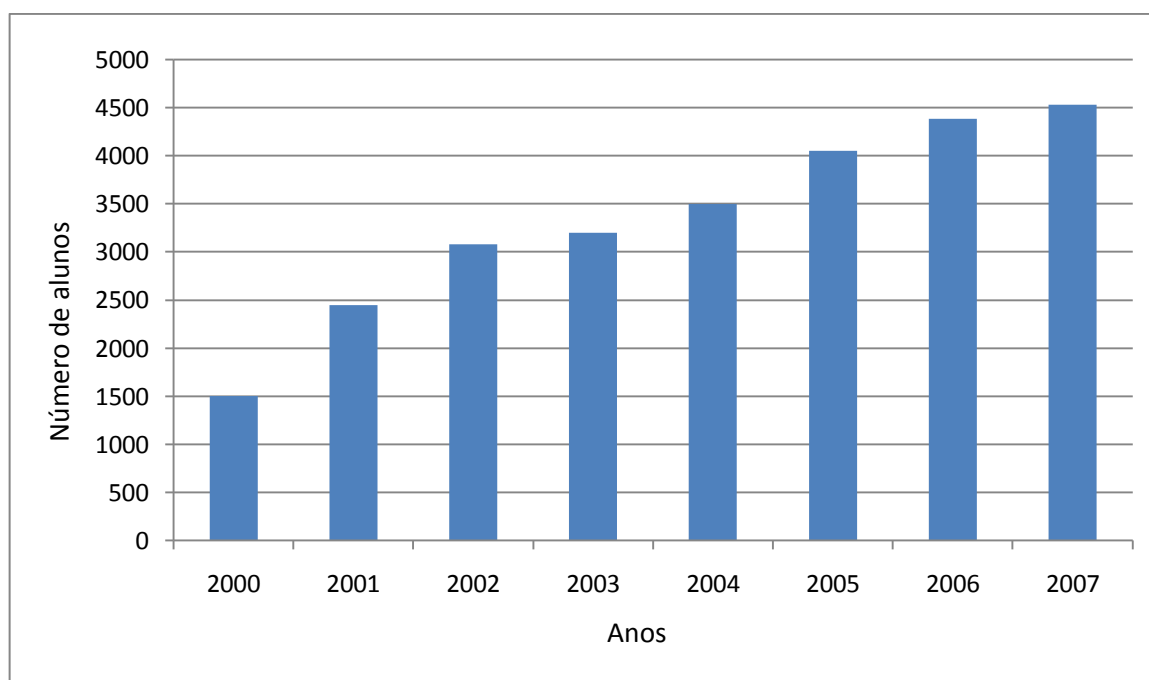
Para construir esse tipo de gráfico, basta estabelecer uma escala conveniente para definir o tamanho da barra e usar a frequência de cada ocorrência da variável em estudo na representação.

Exemplo:

Número de alunos que freqüentam uma biblioteca A

Ano	Número de alunos que usaram a biblioteca
2000	1500
2001	2350
2002	3100
2003	3250
2004	3500
2005	4050
2006	4300
2007	4550

Logo temos o gráfico de barras abaixo, ilustrando a frequência dos alunos na biblioteca por ano.



3) Histograma

Consiste em um conjunto de retângulos que apresentam:

- As bases sobre um eixo horizontal (eixo dos X) com centro nos pontos médios e as larguras iguais às amplitudes dos intervalos das classes;
- As áreas dos retângulos são proporcionais às freqüências das classes;

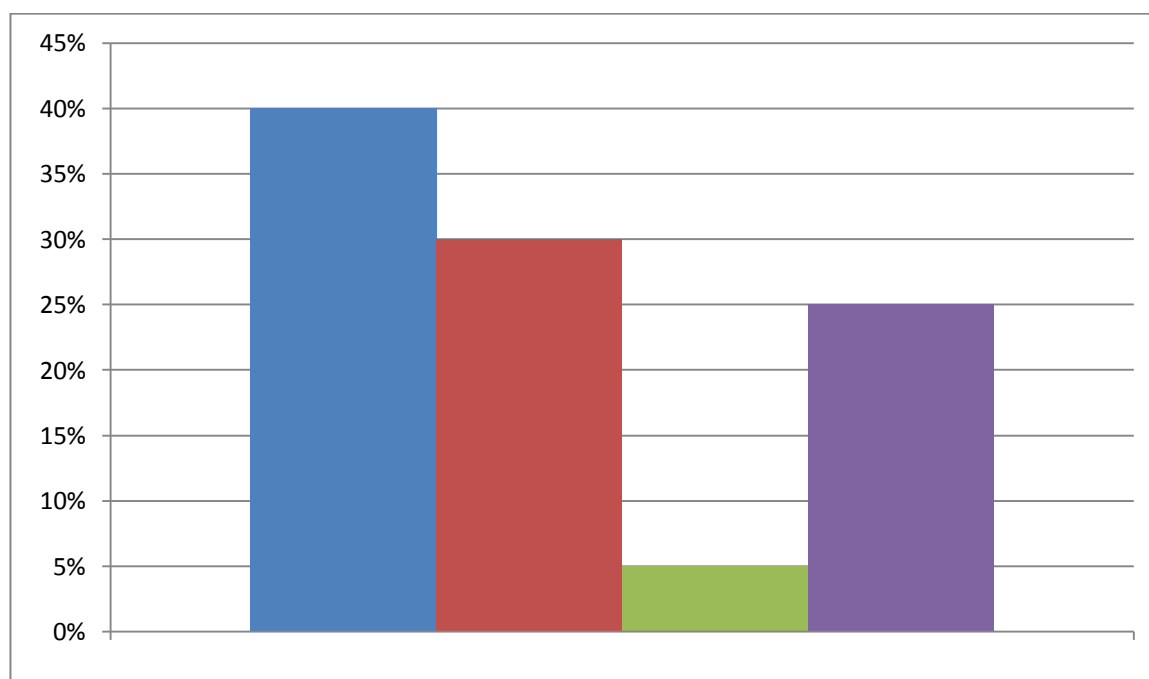
Se todos os intervalos tiverem a mesma amplitude, as alturas dos retângulos serão proporcionais às freqüências das classes.

Exemplo: Suponhamos válida a tabela abaixo:

Tabela de freqüência do tempo de permanência do aluno na biblioteca, onde a amplitude das classes considerada é 55

Tempo de permanência (em minutos)	Freqüência Absoluta	Freqüência Relativa	Porcentagem %
30 ┆ 65	8	$\frac{8}{20} = 0,4$	40
65 ┆ 100	6	$\frac{6}{20} = 0,3$	30
100 ┆ 135	1	$\frac{1}{20} = 0,05$	5
135 ┆ 170	5	$\frac{5}{20} = 0,25$	25
Total	20		100

Logo o Histograma ilustrando o tempo de permanência dos alunos a biblioteca é dado por:



4) Gráfico de Linhas

Neste tipo de gráfico determinamos diversos pontos, que são, unindo-os por segmentos de reta, construindo desta forma, uma curva poligonal. É importante lembrar que esse tipo de gráfico representa a função entre as variáveis envolvidas.

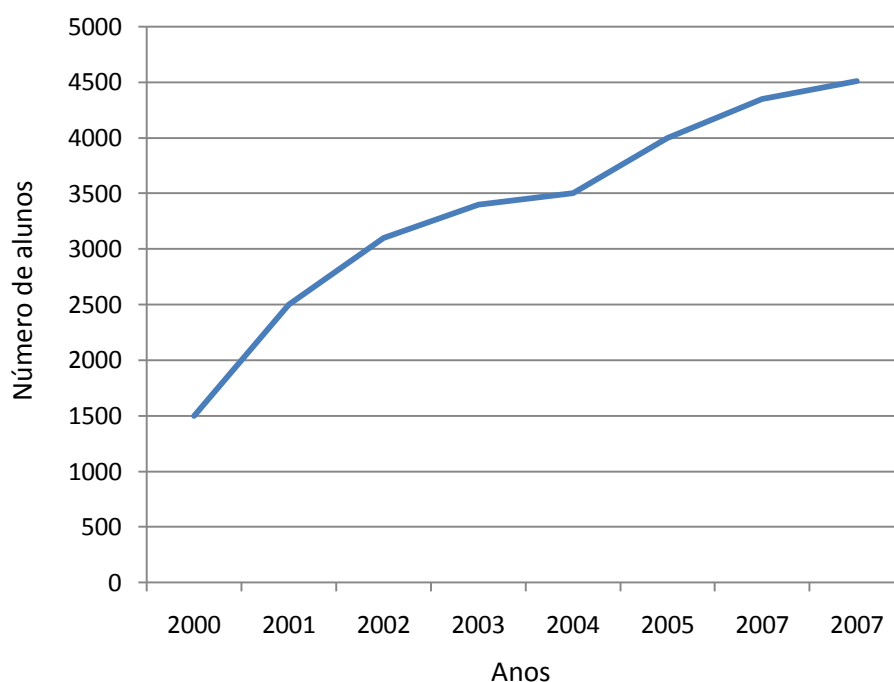
Exemplo:

Considere a tabela abaixo:

Dados do número de alunos que freqüentam a biblioteca por ano

Ano	Número de alunos que usaram a biblioteca
2000	1500
2001	2350
2002	3100
2003	3250
2004	3500
2005	4050
2006	4300
2007	4550

Logo o gráfico de linhas ilustrando a freqüência dos alunos na biblioteca é dado por:



UNIDADE IV - MEDIDAS DE TENDÊNCIA CENTRAL

4.1. Introdução

O estudo que fizemos sobre distribuições de freqüências, até agora, permite-nos descrever, de modo geral, os grupos de valores que uma variável pode assumir. Dessa forma, podemos localizar a maior concentração de valores de uma dada distribuição, isto é, se ela se localiza no início, no meio ou no final, ou ainda, se há uma distribuição por igual.

Porém, para ressaltar as tendências características de cada distribuição, isoladamente, ou em confronto com outras, necessitamos introduzir conceitos que se expressem através de números, que nos permitam traduzir essas tendências. Esses conceitos são denominados elementos típicos da distribuição e são as:

- a) Medidas de posição;
- b) Medidas de variabilidade ou dispersão;
- c) Medidas de assimetria;
- d) Medidas de curtose.

Dentre os elementos típicos, estudaremos nessa unidade, as medidas de posição, que são estatísticas que representam uma série de dados, orientando-nos quanto à posição da distribuição em relação ao eixo horizontal. Também estudaremos a medidas de variabilidade ou dispersão.

De um modo geral, qualquer conjunto de dados estatísticos agrupados ou não dependendo do estudo a que se propõe, ocupam uma posição específica dentro de uma distribuição.

Através de tabelas e gráficos construídos anteriormente, vimos como resumir e apresentar um conjunto de dados. Contudo, podemos resumir ainda mais este conjunto, apresentando um ou alguns valores que “representam” todo o conjunto. Esses valores são chamados de medidas de posição.

As medidas de posição mais importantes são as medidas de tendência central, que recebem tal denominação pelo fato de os dados observados tenderem, em geral, a se agrupar em torno dos valores centrais. As principais medidas de tendência central são:

- a) Média (aritmética, geométrica, harmônica, quadrática)
- b) Moda
- c) Mediana
- d) Separatrizes

4.2. Medidas de Posição para Dados Não Agrupados em Classes

4.2.1. Média Aritmética

A média aritmética de um conjunto de números pode ser de dois tipos: simples ou ponderada.

4.2.1.1. Média Aritmética Simples (\bar{x})

A média aritmética simples de um conjunto de números é igual ao quociente entre a soma dos valores do conjunto e o número total de valores.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

onde:

\bar{x} - média aritmética simples;

x_i - valores da variável;

n - número de observações.

Exemplo:

Sejam os valores abaixo correspondente aos salários de 5 funcionários de uma empresa. Calcular a média aritmética simples.

$$x_1 = 800,00; x_2 = 780,00; x_3 = 820,00; x_4 = 810,00; x_5 = 790,00$$

$$\bar{x} = \frac{800 + 780 + 820 + 810 + 790}{5} = 800,00 \Rightarrow \bar{x} = 800,00$$

A média aritmética simples será calculada sempre que os valores vierem representados individualmente.

Exercício

Calcule a média aritmética dos números 8, 3, 6, 12 e 10.

4.2.1.2. Média Aritmética Ponderada

A média aritmética é considerada ponderada quando os valores do conjunto tiverem pesos diferentes. No caso da média aritmética simples, todos os valores possuem o mesmo peso. A média aritmética ponderada é o quociente entre o produto dos valores da variável pelos respectivos pesos e a soma dos pesos.

$$\bar{x}_p = \frac{\sum_{i=1}^k x_i p_i}{\sum_{i=1}^k p_i} = \frac{x_1 p_1 + x_2 p_2 + \dots + x_k p_k}{p_1 + p_2 + \dots + p_k} = \frac{x_1 p_1 + x_2 p_2 + \dots + x_k p_k}{n}$$

onde:

\bar{x}_p - média aritmética ponderada;

x_i - valores da variável;

p_i - pesos dos valores da variável (nº de vezes que cada valor ocorre);

$\sum_{i=1}^k p_i = n$ - número de observações;

k - número de classes ou de valores individuais diferentes da variável.

Os pesos dos valores da variável correspondem ao número de vezes que cada valor ocorre.

Exemplo:

Sejam os valores abaixo correspondentes aos salários de 10 funcionários de uma empresa:

$x_1 = 800,00; x_2 = 770,00; x_3 = 790,00; x_4 = 800,00; x_5 = 760,00;$

$x_6 = 750,00; x_7 = 760,00; x_8 = 780,00; x_9 = 800,00; x_{10} = 770,00$

$$\bar{x}_p = \frac{(750,00 \times 1) + (760,00 \times 2) + (770,00 \times 2) + (780,00 \times 1) + (790,00 \times 1) + (800,00 \times 3)}{1 + 2 + 2 + 1 + 1 + 3}$$

$$\therefore \bar{x}_p = 778,00$$

Exercício

Para ingressar em uma determinada Instituição de Ensino um candidato deve fazer três provas, uma de matemática, uma de português e outra de conhecimentos gerais e deve obter no mínimo média 6,0 para ser classificado, disputando a vaga com os demais candidatos segundo a ordem de classificação. O que podemos afirmar sobre um candidato que obteve as notas mostradas na tabela a seguir?

Provas	Pesos das Provas	Notas do Candidato A
Matemática	7	8,0
Português	6	6,0
Conhecimentos Gerais	5	4,0

4.2.2. Média Geométrica (\bar{x}_g)

A média geométrica de “n” valores é definida, genericamente, como a raiz n-ésima do produto de todos eles.

Dados “n” valores x_1, x_2, \dots, x_n , a média geométrica desses valores será:

$$\bar{x}_g = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} \quad \text{ou} \quad \bar{x}_g = \sqrt[n]{\prod_{i=1}^n x_i}$$

Onde:

\prod – letra grega (pi) que indica o produto dos valores da variável.

Exemplo:

Calcular a média geométrica simples do conjunto $x = \{1, 4, 16, 64\}$.

$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[4]{x_1 \times x_2 \times x_3 \times x_4}$$

$$\bar{x}_g = \sqrt[4]{1 \times 4 \times 16 \times 64} = 8 \quad \Rightarrow \quad \bar{x}_g = 8$$

Ou

$$\log \bar{x}_g = \log \sqrt[4]{4096} = \frac{\log 4096}{4} = \frac{3,612360}{4} = 0,903090$$

$$\bar{x}_g = \text{anti log } 0,903090 = 8$$

Exercício

Determine a média geométrica dos números 2, 4 e 8.

4.2.3. Média Harmônica (\bar{x}_h)

A média harmônica de um conjunto de valores x_i é o inverso da média aritmética dos inversos dos valores.

$$\bar{x}_h = \frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} \quad \text{ou} \quad \bar{x}_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Exemplo:

Calcular a média harmônica simples do seguinte conjunto de números:

$x = \{10, 60, 360\}$

$$\bar{x}_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} = \frac{3}{\frac{1}{10} + \frac{1}{60} + \frac{1}{360}} = \frac{3}{\frac{36+6+1}{360}} = \frac{3 \times 360}{43} = 25,12 \quad \Rightarrow$$

$$\bar{x}_h = 25,12$$

Exercício

A pontuação de um aluno em uma seqüência de prova foi: 8,4; 9,1; 7,2; 6,8; 8,7 e 7,8. Determine a média harmônica.

4.2.4. Média Quadrática (\bar{x}_q)

A média quadrática de um conjunto de “n” valores x_i é a raiz quadrada da média aritmética dos quadrados.

$$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$$

onde:

x_i - valores da variável;

n - número de observações.

Exemplo:

Calcular a média quadrática do conjunto:

$$x = \{2, 3, 4, 5\}$$

$$\bar{x}_q = \sqrt{\frac{2^2 + 3^2 + 4^2 + 5^2}{4}} = 3,67 \quad \Rightarrow \quad \bar{x}_q = 3,67$$

Exercício

O valor da hora-aula de 5 professores de um determinado colégio está mostrado na tabela abaixo. Determine a média quadrática do valor da hora-aula.

Professores	Valor da hora-aula (R\$)
A	7,50
B	12,50
C	17,00
D	25,00
E	32,00

4.2.5. Moda (Mo)

A moda é outra medida de tendência central, definida como o valor mais freqüente, quando comparado sua freqüência com a dos valores contíguos de um conjunto ordenado.

Considerando um conjunto ordenado de valores, a moda será o valor predominante, o valor mais freqüente desse conjunto. Esse conjunto de valores pode ser:

- Amodal: não apresenta uma moda, isto é, todos os valores da variável em estudo ocorreram com a mesma intensidade (freqüência).
- Plurimodal: quando houver mais de um valor predominante.

Exemplo:

Calcular a moda dos seguintes conjuntos de valores:

$$x = \{4, 5, 5, 6, 6, 6, 7, 7, 8, 8\} \quad Mo = 6$$

$$y = \{4, 4, 5, 5, 6, 6\}$$

Amodal, pois seus três valores apareceram 2 vezes cada um.

$z = \{1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6\}$ $Mo_1 = 2$ e $Mo_2 = 5$, conjunto bimodal, pois tanto o valor 2 como o valor 5 apresentaram o maior número de ocorrências.

$$w = \{1, 2, 3, 4, 5\} \quad \text{Amodal}$$

4.2.6. Mediana (Me)

Mediana é um valor central de um rol, ou seja, a mediana de um conjunto de valores ordenados (crescente ou decrescente) é a medida que divide este conjunto em duas partes iguais, cujo valor está sucedido de 50% e antecedido de 50% desse conjunto de observações.

A mediana também é considerada uma medida separatriz, pois divide a distribuição (a série) ou conjunto de dados em partes iguais. É uma medida muito utilizada na análise de dados estatísticos, especialmente quando se atribui pouca importância aos valores extremos da variável.

A mediana é um valor que ocupa uma determinada ordem ou posição na série ordenada.

Estando ordenados os valores de uma série e sendo “n” o número de elementos da série, o valor mediano será:

- Se “n” for ímpar: a mediana será o termo de ordem:

$$P = \frac{n+1}{2}$$

onde:

P – elemento mediano (Posição);
 n – número de elementos do conjunto.

- Se “ n ” for par: a mediana será a média aritmética dos termos de ordem $n/2$ e $n/2 + 1$:

$$P_1 = \frac{n}{2} \qquad P_2 = \frac{n}{2} + 1 \qquad Me = \frac{P_1 + P_2}{2}$$

Exemplos:

1) Para a série {2, 5, 6, 9, 10, 13, 15, 16, 18} $n = 9 \Rightarrow$ A mediana será o termo de ordem

$$P = \frac{9+1}{2} = 5$$

Assim, $Me = 10$.

2) Para a série {2, 6, 7, 10, 12, 13, 18, 21} $n = 8 \Rightarrow$ A mediana será o termo de ordem:

$$P_1 = \frac{n}{2} = \frac{8}{2} = 4 \qquad P_2 = \frac{n}{2} + 1 = \frac{8}{2} + 1 = 5$$

$$\therefore Me = \frac{10+12}{2} = 11$$

Observação: A mediana depende da “posição” e não dos valores dos elementos na série ordenada. Essa é uma das diferenças marcantes entre a mediana e a média (que se deixa influenciar, e muito, pelos valores extremos).

Exercícios

1) O quadro abaixo apresenta os salários recebidos por funcionários de uma empresa segundo o cargo que ocupam. Determine a mediana dos salários que a empresa paga para seus funcionários.

Cargos	Salários (R\$)
Diretor	2.500,00
Chefe de Departamento	1.700,00
Agente Administrativo	800,00
Serviços Gerais	500,00
Segurança	450,00
Apoio Técnico	600,00

2) Uma instituição de pesquisa fez um levantamento de vários produtos adquiridos normalmente em uma feira popular. O resultado está na tabela a seguir. Determine a média, mediana e a moda do preço dos produtos pesquisados.

Produtos	Preço (R\$)
Tomate	2,50
Cebola	2,00
Batata	2,30
Cenoura	1,70
Repolho	2,00
Pimentão	1,90
Abobora	2,70

4.3. Medidas de Posição para Dados Agrupados em Classes

4.3.1. Média Aritmética Ponderada

O valor de x_i passa a ser o ponto médio do intervalo.

$$\bar{x}_p = \frac{\sum_{i=1}^k P_{m_i} \cdot f_i}{n} = \frac{P_{m_1} \cdot f_1 + P_{m_2} \cdot f_2 + \dots + P_{m_k} \cdot f_k}{f_1 + f_2 + \dots + f_k}$$

onde:

P_{m_i} é o ponto médio do intervalo de classe i ;

f_i é a frequência absoluta do intervalo classe i .

Exemplo:

Classes	f_i	P_{m_i}	$P_{m_i} \cdot f_i$
10 — 20	5	15	75
20 — 30	10	25	250
30 — 40	15	35	525
40 — 50	10	45	450
50 — 60	5	55	275
	$n = \sum_{i=1}^5 f_i = 45$		$\sum_{i=1}^5 P_{m_i} \cdot f_i = 1575$

$$\bar{x}_p = \frac{\sum_{i=1}^5 P_{m_i} \cdot f_i}{n} = \frac{1575}{45} \Rightarrow \bar{x}_p = 35.$$

Propriedades da Média

1ª) A soma algébrica dos afastamentos (ou desvios, ou resíduos) de um conjunto de números tomados em relação à média aritmética é zero. Simbolicamente:

$$\sum_{i=1}^n d_i = \sum (x_i - \bar{x}) = 0 \quad \text{ou} \quad \sum_{i=1}^k d_i p_i = \sum (x_i - \bar{x}) p_i = 0$$

2ª) Se multiplicarmos ou dividirmos todas as informações por uma constante, a média aritmética também ficará multiplicada ou dividida por essa constante.

3ª) Somando-se ou subtraindo-se uma constante a todos os valores de um conjunto de informações, a média aritmética ficará somada ou subtraída dessa constante.

4ª) A soma dos quadrados dos desvios tomados em relação à média aritmética é um mínimo.

Uso da Média:

É a mais utilizada dos valores médios, pela simplicidade e rapidez de seu cálculo.

- Quando se deseja obter um valor médio estável e significativo que inclui no seu cálculo todos os valores;
- É usada na determinação de índices de grande importância estatística;
- Quando se deseja maior precisão na determinação de uma medida, realiza-se várias induções e toma-se como resultado a média aritmética.

Exercício

A tabela a seguir contém informações da renda familiar mensal de um grupo de estudantes. Determine a renda média desse grupo.

Renda familiar mensal (em salário mínimo)	Frequência Relativa $fr_i = \frac{f_i}{\sum f_i}$
5 — 6,7	0,1
6,7 — 8,3	0,1
8,3 — 10	0,2
10 — 11,6	0,6

4.3.2. Média Geométrica Ponderada

A média geométrica ponderada de um conjunto de números dispostos em uma tabela de freqüências é calculada por intermédio da seguinte expressão:

$$\bar{x}_g = \sqrt[n]{x_1^{p_1} \times x_2^{p_2} \times \dots \times x_k^{p_k}} \quad \text{ou} \quad \bar{x}_g = \sqrt[n]{\prod_{i=1}^k x_i^{p_i}}$$

onde: $n = \sum_{i=1}^k p_i$

Exemplo: Calcular a média geométrica ponderada dos valores constantes da seguinte tabela:

x_i	p_i
1	2
3	4
9	2
27	1
	$\sum_{i=1}^k p_i = 9$

Temos que $k = 4$ e $n = 9$, logo:

$$\bar{x}_g = \sqrt[9]{1^2 \times 3^4 \times 9^2 \times 27^1} = 3,82955 \quad \Rightarrow \quad \bar{x}_g = 3,83$$

Exercício

A tabela a seguir contém informações da renda familiar mensal de funcionários de uma determinada empresa. Determine a média geométrica da renda desses funcionários.

Renda Familiar mensal	Número de Funcionários
1 — 3	30
3 — 5	12
5 — 7	6
7 — 9	3

Propriedades da Média Geométrica

1ª) O produto dos quocientes de cada valor de um conjunto de números pela média geométrica do conjunto é igual a um.

$$\frac{x_1}{\bar{x}_g} \times \frac{x_2}{\bar{x}_g} \times \dots \times \frac{x_n}{\bar{x}_g} = 1$$

2ª) Séries que apresentam o mesmo número de elementos com a mesma soma total têm a mesma média aritmética, enquanto séries que apresentam o mesmo número de elementos com o mesmo produto têm a mesma média geométrica.

3ª) A média geométrica é menor ou igual à média aritmética.

A desigualdade $\bar{x}_g < \bar{x}$ sempre se verifica, quando os valores da série forem positivos e nem todos iguais. Se entre eles houver um ou mais zeros, a média geométrica será nula.

A igualdade $\bar{x}_g = \bar{x}$ só ocorrerá quando todos os valores da série forem iguais.

4ª) Quanto maior for a diferença entre os valores originais maior será a diferença entre as médias aritmética e geométrica.

Uso da Média Geométrica

- a) Quando um dos valores é nulo não se aplica a média geométrica;
- b) Para o cálculo do índice do custo de vida;
- c) Crescimento demográfico.

Exemplo:

Conjunto	Média Aritmética (\bar{x})	Média Geométrica (\bar{x}_g)
x = {2, 2}	2	2
y = {14, 16}	15	14,97
w = {8, 12}	10	9,80
z = {2, 50}	26	10

4.3.3. Média Harmônica Ponderada (\bar{x}_h)

A média harmônica ponderada de um conjunto de números, dispostos em uma tabela de frequências, é dada pela seguinte expressão:

$$\bar{x}_h = \frac{1}{\sum_{i=1}^k \frac{1}{x_i} \cdot p_i} = \frac{\sum_{i=1}^k p_i}{\sum_{i=1}^k \frac{1}{x_i} \cdot p_i} = \frac{n}{\sum_{i=1}^k \frac{p_i}{x_i}}$$

onde:

n - número de observações;

x_i - valores da variável;

p_i - pesos dos valores da variável.

Exemplo:

Calcular a média harmônica dos dados constantes da tabela abaixo:

Classes	p _i	x _i	1/x _i	p _i /x _i
1 — 3	2	2	1/2	1,00
3 — 5	4	4	1/4	1,00
5 — 7	8	6	1/6	1,33
7 — 9	4	8	1/8	0,50
9 — 11	2	10	1/10	0,20
	$\sum_{i=1}^5 p_i = 20$			$\sum_{i=1}^5 \frac{p_i}{x_i} = 4,03$

$$\bar{x}_h = \frac{20}{4,03} = 4,96 \quad \Rightarrow \quad \bar{x}_h = 4,96$$

Exercício

A tabela a seguir mostra a distribuição, em toneladas, das cargas máximas suportadas por certos cabos fabricados por uma companhia. Determine a média harmônica.

Carga Máxima (toneladas)	Número de cabos
9,3 † 9,7	2
9,8 † 10,2	5
10,3 † 10,7	12
10,8 † 11,2	17
11,3 † 11,7	14
11,8 † 12,2	6
12,3 † 12,7	3
12,8 † 13,2	1
	$n = 60$

Propriedades da Média Harmônica

A média harmônica é menor ou igual à média geométrica para valores da variável diferentes de zero.

$$\bar{x}_h \leq \bar{x}_g$$

Por extensão de raciocínio e de acordo com a terceira propriedade da média geométrica, podemos escrever:

$$\bar{x}_h \leq \bar{x}_g \leq \bar{x}$$

Uso da Média Harmônica

- a) Muito utilizada em fatores de ordem física (aceleração, velocidade);
- b) Custo médio de artigos comprados com uma quantia fixa.

4.3.4. Média Quadrática Ponderada

Quando os valores da variável estiverem dispostos em uma tabela de freqüências, a média quadrática será determinada pela seguinte expressão:

$$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^k x_i^2 \cdot p_i}{n}} = \sqrt{\frac{x_1^2 \cdot p_1 + x_2^2 \cdot p_2 + \dots + x_k^2 \cdot p_k}{n}}$$

onde:

x_i - valores da variável;

p_i - pesos dos valores da variável;

n - número de observações.

Exemplo:

Calcular a média quadrática dos valores da tabela abaixo:

Classes	p_i	x_i	x_i^2	$x_i^2 \cdot p_i$	$x_i \cdot p_i$
2 — 4	5	3	9	45	15
4 — 6	10	5	25	250	50
6 — 8	12	7	49	588	84
8 — 10	10	9	81	810	90
10— 12	5	11	121	605	55
	$n = 42$			2298	294

$$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^k x_i^2 \cdot p_i}{n}} = \sqrt{\frac{2298}{42}} = 7,40 \Rightarrow \bar{x}_q = 7,40$$

Exercício

A tabela a seguir contém a altura de 100 estudantes do sexo masculino de uma determinada universidade. A partir das informações contidas na tabela determine as médias aritmética, geométrica, harmônica e quadrática altura dos estudantes.

Altura (cm)	Número de Estudantes
1,51 † 1,58	5
1,59 † 1,66	18
1,67 † 1,74	42
1,75 † 1,82	27
1,83 † 1,90	8

Propriedades da Média Quadrática

1ª) A média quadrática de uma constante é igual a constante.

2ª) Multiplicando ou dividindo todos os valores de um conjunto de números por um valor constante arbitrário, a média quadrática fica multiplicada ou dividida por essa constante.

3ª) Sempre que os valores de “x” forem positivos é válida a relação:

$$\bar{x}_q \geq \bar{x} \geq \bar{x}_g \geq \bar{x}_h$$

A igualdade se verifica quando os valores da variável forem iguais (constantes).

$$\bar{x}_q = \bar{x} = \bar{x}_g = \bar{x}_h \quad \text{para} \quad x_1 = x_2 = \dots = x_0$$

A média quadrática é largamente utilizada em Estatística, principalmente quando se pretende calcular a média de desvios ($x - \bar{x}$) em vez de a média dos valores originais. Neste caso, a média quadrática é denominada desvio-padrão, que é uma importante medida de dispersão, que será estudada mais adiante.

4.3.5. Moda para Dados Agrupados

Os valores da variável dispostos em uma tabela de freqüências podem apresentar-se individualmente ou agrupados em classes. No primeiro caso, a determinação da moda é imediata, bastando, para isso, consultar a tabela, localizando o valor que apresenta a maior freqüência. Esse valor será a moda do

conjunto. Assim, a moda do conjunto apresentado na tabela abaixo é $Mo = 3$, indicando que a rejeição de 3 peças defeituosas por mês foi o resultado mais observado.

Exemplo:

Número de Peças de Precisão Defeituosas devolvidas mensalmente pelo Controle de Qualidade.

Nº de Peças com Defeito x_i	Nº de meses p_i
0	2
1	4
2	6
3	8
4	4
5	2
6	1
	$\sum_{i=1}^7 p_i = 27$

Tratando-se de uma tabela de freqüências com valores tabulados e agrupados em classes, o procedimento não é imediato, sendo disponíveis alguns métodos de cálculo distintos. Qualquer que seja o método adotado, o primeiro passo para determinar a moda é localizar a classe que apresenta a maior freqüência, comumente chamada de classe modal. Um dos métodos para o cálculo da Moda é o Método de Czuber.

Método de Czuber

O método de Czuber, para o cálculo da moda elaborada, leva em consideração não apenas as freqüências das classes adjacentes, mas também a freqüência da classe modal. O ponto que corresponde à moda divide o intervalo da classe modal em duas partes, as quais são proporcionais às diferenças entre a freqüência da classe modal e as das respectivas classes adjacentes. Assim:

$$Mo = Li + h \frac{F_{Mo} - F_{ant}}{2F_{Mo} - (F_{ant} + F_{post})}$$

onde:

Estatística Aplicada à Educação

L_i - limite inferior da classe modal;

h - amplitude do intervalo de classe;

F_{Mo} - frequência simples da classe modal;

F_{ant} - frequência simples da classe anterior à classe modal;

F_{post} - frequência simples da classe posterior à classe modal.

Exemplo:

Determinar a moda pelo método de Czuber para os dados apresentados abaixo:

Classes	p_i ou f_i
10 — 20	2
20 — 30	3
30 — 40	10
40 — 50	9
50 — 60	4
	$n = 28$

Classe Modal: 30 |— 40

$L_i = 30$

$$Mo = 30 + 10 \frac{10 - 3}{2 \times 10 - (3 + 9)} = 38,75$$

$h = 10$

$F_{Mo} = 10$

$Mo = 38,75$

$F_{ant} = 3$

$F_{post} = 9$

Exercício

Determine a moda pelo método de Czuber para os dados apresentados abaixo:

Distribuição de Estaturas de um acampamento infantil

i	Estaturas (cm)	Frequência
1	120 — 128	6
2	128 — 136	12
3	136 — 144	16
4	144 — 152	13
5	152 — 160	7
		$n = \sum f_i = 54$

4.3.6. Mediana

Para o caso de uma distribuição, a ordem, a partir de qualquer um dos extremos, é dada por:

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

1º passo: Determina-se as freqüências acumuladas;

2º passo: Calcula-se a posição da mediana;

3º passo: Marca-se a classe correspondente à freqüência acumulada imediatamente superior a posição calculada e, em seguida, emprega-se a fórmula:

$$M_e = l_{M_e} + \left(\frac{\frac{\sum f_i}{2} - F_{ant}}{f_{M_e}} \right) \cdot h_{M_e}$$

onde:

l_{M_e} – é o limite inferior da classe mediana;

F_{ant} - freqüência acumulada até a classe anterior à classe mediana;

f_{M_e} – freqüência simples da classe mediana;

h_{M_e} – amplitude do intervalo da classe mediana.

Exemplo:

i	Estaturas (cm)	Fi	Fa
1	150 — 154	4	4
2	154 — 158	9	13
3	158 — 162	11	24
4	162 — 166	8	32
5	166 — 170	5	37
6	170 — 174	3	40
		$\sum f_i = 40$	

$\frac{\sum f_i}{2} = \frac{40}{2} = 20$
← Classe mediana de ordem 3 (i = 3)

Como há 24 valores incluídos nas três primeiras classes da distribuição e como pretendemos determinar o valor que ocupa o 20º lugar, a partir do início da série, vemos que esse deve estar localizado na terceira classe (i = 3), supondo que as freqüências dessas classes estejam uniformemente distribuídas.

Como há 11 elementos nessa classe e o intervalo de classe é igual a 4, devemos tomar, a partir do limite inferior, a distância:

$$\frac{20-13}{11} \times 4 = \frac{7}{11} \times 4 \quad \text{E a mediana será:}$$

$$Me = 158 + \frac{7}{11} \times 4 = 158 + \frac{28}{11} = 158 + 2,54 = 160,54$$

$$Me = 160,5 \text{ cm}$$

Exercício

Determine a mediana da tabela de distribuição de freqüências abaixo:

Distribuição de Estaturas de um acampamento infantil

i	Estaturas (cm)	Freqüência
1	120 — 128	6
2	128 — 136	12
3	136 — 144	16
4	144 — 152	13
5	152 — 160	7
		$n = \sum f_i = 54$

Uso da Mediana:

- a) É usada em fenômenos educacionais quando se quer tornar objetiva a avaliação de uma classe;
- b) Quando se quer exatamente o valor que divide a metade da distribuição;
- c) Quando a distribuição tem resultados discrepantes e pairam dúvidas sobre sua validade e correção.

Comparação entre média aritmética, mediana e moda.

Há um momento em que o pesquisador fará a seguinte pergunta: Qual a medida de tendência central que representa melhor o conjunto de dados em estudo?

Assim, mostraremos as vantagens e desvantagens de cada uma das medidas de tendência central.

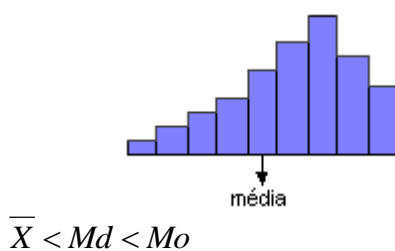
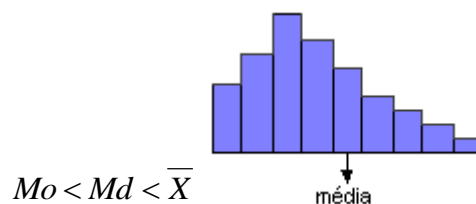
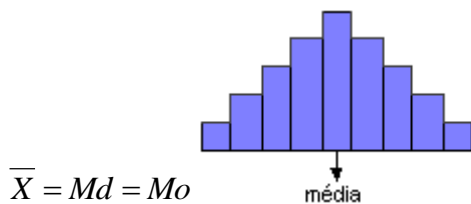
- A moda é uma medida que requer apenas o conhecimento da freqüência absoluta e pode ser utilizada para qualquer tipo de variáveis, tanto qualitativas, quanto quantitativas.
- A mediana é uma medida que exige uma ordenação de categorias, da mais alta a mais baixa, assim ela só pode ser obtida para variáveis

qualitativas ordinais ou para as quantitativas, jamais para variáveis qualitativas nominais. Além disso, a mediana não é influenciada por valores extremos.

- A média aritmética trabalha com todos os elementos do conjunto de dados, enquanto a mediana utiliza apenas um ou dois valores. No entanto a média sofre influência de valores extremos (muito alto ou baixo).
- A média é uma medida que pode ser calculada apenas para variáveis quantitativas. E embora a média seja um valor mais fácil de entender, tem o defeito de nos induzir em erro se a nossa amostra tiver valores muito extremos.

Assim, no caso das variáveis quantitativas, quando o valor da Mediana é muito diferente da Média, é aconselhável considerar sempre a Mediana como valor de referência mais importante.

Quando a distribuição dos dados é considerada "normal", então a melhor medida de localização do centro, é a média. Ora sendo a Distribuição Normal uma das distribuições mais importantes e que surge com mais frequência em aplicações, esse fato justifica a grande utilização da média. Esquemáticamente podemos posicionar a média da forma seguinte, tendo em conta a representação gráfica na forma de histograma.



Outras Medidas de Posição

4.3.7. Quartis, decis e percentis

Verificamos que a mediana separa a série em duas partes iguais, em que cada parte contém o mesmo número de elementos, porém, uma mesma série pode ser dividida em duas ou mais partes que contenham a mesma quantidade

de elementos. O nome da medida de posição separatriz de acordo com a quantidade de partes que é dividida a série.

- Mediana: divide a série em duas partes iguais (M_d);
- Quartis: divide a série em quatro partes iguais (Q_1, Q_2, Q_3);
- Decis: divide a série em 10 partes iguais ($D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9$);
- Percentis: divide a série em 100 partes iguais ($P_1, P_2, P_3, \dots, P_{99}$).

4.3.7.1. Quartis (Q_K)

Nos quartis a série é dividida em quatro partes iguais. Os elementos separatrizes da série são Q_1, Q_2 e Q_3 .

Q_1 : é o primeiro quartil, corresponde à separação dos primeiros 25% de elementos da série;

Q_2 : é o segundo quartil, coincide com a mediana ($Q_2 = M_d$);

Q_3 : é o terceiro quartil, corresponde à separação dos últimos 25% de elementos da série.

Para o cálculo dos quartis utilizam-se técnicas semelhantes àquelas do cálculo da mediana. Conseqüentemente, podemos utilizar as mesmas fórmulas do cálculo da mediana, levando em conta que onde houver a expressão $\frac{\sum f_i}{2}$ será substituída por $\frac{K \sum f_i}{4}$, sendo K o número da ordem do quartil, em que $K = 1$ corresponde ao primeiro quartil. Se $K = 2$, temos o segundo quartil e se $K = 3$, temos o terceiro quartil.

Determina-se, inicialmente, a classe que contém o valor quartil a ser calculado. A identificação da classe é feita por meio do termo da ordem calculada pela expressão.

$$\frac{K \sum f_i}{4} (K = 1, 2, 3)$$

Esse termo está localizado numa classe que recebe o nome de classe quartil. Assim, temos:

$$Q_K = l_{Q_K} + \left[\frac{\frac{K \sum f_i}{4} - F_{ant}}{f_{Q_K}} \right] \cdot h_{Q_K}$$

Sendo:

l_{Q_K} = limite inferior da classe do quartil considerado;

F_{ant} = freqüência acumulada da classe anterior à classe do quartil considerado;

h_{Q_K} = amplitude do intervalo de classe do quartil considerado;

f_{Q_K} = freqüência simples da classe do quartil considerado.

Exemplo:

Para o cálculo dos quartis de dados agrupados com intervalos de classe, consideramos a distribuição dos pesos de um grupo de turistas que visita um parque temático. Será acrescentada uma coluna com os valores da frequência acumulada.

Distribuição dos pesos de um grupo de turistas de um parque temático

i	Pesos (Kg)	Frequência (f_i)	Frequência acumulada (F_i)
1	10 † 30	8	8
2	30 † 50	26	34
3	50 † 70	57	91
4	70 † 90	42	133
5	90 † 110	27	160
6	110 † 130	16	176
		$\sum f_i = 176$	

Calcula-se primeiro a classe a que pertence o quartil Q_1 ($K = 1$).

$$\frac{1 \cdot \sum f_i}{4} = \frac{176}{4} = 44$$

O primeiro quartil corresponde ao quadragésimo quarto termo da série. Observando a coluna de frequência acumulada, verificamos que o quadragésimo termo pertence à terceira classe (a frequência acumulada da terceira classe abrange do 35º termo ao 91º termo).

Sabendo que classe do primeiro quartil é a terceira classe, podemos verificar qual o valor numérico do primeiro quartil.

$$Q_1 = l_{Q_1} + \left[\frac{1 \cdot \sum f_i - F_{ant}}{f_{Q_1}} \right] \cdot h_{Q_1} = 50 + \left[\frac{44 - 34}{57} \right] \cdot 20 = 53,5 \text{ kg}$$

Os cálculos para os quartis Q_2 e Q_3 são feitos de modo análogo ao do primeiro quartil.

$$K = 2 \rightarrow \frac{2 \cdot \sum f_i}{4} = \frac{2 \times 176}{4} = 88 \text{ (o segundo quartil pertence à terceira classe).}$$

Logo:

$$Q_2 = l_{Q_2} + \left[\frac{2 \cdot \sum f_i - F_{ant}}{f_{Q_2}} \right] \cdot h_{Q_2} = 50 + \left[\frac{88 - 34}{57} \right] \cdot 20 = 68,95 \text{ kg}$$

$$K = 3 \rightarrow \frac{3 \cdot \sum f_i}{4} = \frac{3 \times 176}{4} = 132 \text{ (o terceiro quartil pertence à quarta classe)}$$

Logo:

$$Q_3 = l_{Q_3} + \left[\frac{3 \cdot \sum f_i - F_{ant}}{f_{Q_3}} \right] \cdot h_{Q_3} = 70 + \left[\frac{132 - 91}{42} \right] \cdot 20 = 89,52 \text{ kg}$$

Portanto:

$$Q_1 = 53,5 \text{ kg}, Q_2 = 68,95 \text{ kg e } Q_3 = 89,52 \text{ kg}$$

4.3.7.2. Decis (D_K)

Nos decis, a série é dividida em 10 partes iguais (D_1, D_2, \dots, D_9).

D_1 : é o primeiro decil, corresponde à separação dos primeiros 10% de elementos da série;

D_5 : é o quinto decil, coincide com a mediana ($D_5 = M_d$);

D_9 : é o nono decil, corresponde à separação dos últimos 10% de elementos da série.

Determina-se a classe que contém o valor decil a ser calculado pela expressão:

$$\frac{K \sum f_i}{10} \quad (K = 1, 2, 3, \dots, 9)$$

Esse termo está localizado numa classe que recebe o nome de classe decil. Para o cálculo dos decis utilizamos técnicas semelhantes às do cálculo dos quartis. E utilizamos a fórmula:

$$D_K = l_{D_K} + \left[\frac{\frac{K \sum f_i}{10} - F_{ant}}{f_{D_K}} \right] \cdot h_{D_K}$$

Sendo:

l_{D_K} = limite inferior da classe do decil considerado;

F_{ant} = freqüência acumulada da classe anterior à classe do decil considerado;

h_{D_K} = amplitude do intervalo de classe do decil considerado;

f_{D_K} = freqüência simples da classe do decil considerado.

Exemplo:

O cálculo dos decis será exemplificado com os dados da próxima tabela, que organiza as estaturas de adolescentes, colhidas durante o período em que participaram de um acampamento, durante as férias.

Distribuição de estaturas de um acampamento infantil

i	Estaturas (cm)	Freqüência (f_i)	Freqüência acumulada (F_i)
1	120 – 128	6	6
2	128 – 136	12	18
3	136 – 144	16	34
4	144 – 152	13	47
5	152 – 160	7	54
		$\sum f_i = 54$	

Calculam-se os decis $D_1, D_2, \dots, D_7, \dots$, de forma semelhante ao cálculo dos quartis.

Primeiro decil ($K = 1$): $\frac{1 \cdot \sum f_i}{10} = \frac{54}{10} = 5,4$ (o primeiro decil pertence à primeira classe).

$$D_1 = l_{D_1} + \left[\frac{\frac{1 \cdot \sum f_i}{10} - F_{ant}}{f_{D_1}} \right] \cdot h_{D_1} = 120 + \left[\frac{5,4 - 0}{6} \right] \cdot 8 = 127,5 \text{ cm}$$

Segundo decil ($K = 2$): $\frac{2 \cdot \sum f_i}{10} = \frac{2 \times 54}{10} = 10,8$ (o segundo decil pertence à segunda classe).

$$D_2 = l_{D_2} + \left[\frac{\frac{2 \cdot \sum f_i}{10} - F_{ant}}{f_{D_2}} \right] \cdot h_{D_2} = 128 + \left[\frac{10,8 - 6}{12} \right] \cdot 8 = 131,2 \text{ cm}$$

Dessa forma, podemos calcular os outros decis. Por exemplo, o cálculo do sétimo decil ($K = 7$): $\frac{7 \cdot \sum f_i}{10} = \frac{7 \times 54}{10} = 37,8$ (o sétimo decil pertence à quarta classe).

$$D_7 = l_{D_7} + \left[\frac{\frac{7 \cdot \sum f_i}{10} - F_{ant}}{f_{D_7}} \right] \cdot h_{D_7} = 144 + \left[\frac{37,8 - 34}{13} \right] \cdot 8 = 146,3 \text{ cm}$$

4.3.7.2. Percentis (D_K)

Nos percentis, a série é dividida em 100 partes iguais ($P_1, P_2, P_3, \dots, P_{99}$).

P_1 : é o primeiro percentil, corresponde à separação do primeiro 1% de elementos da série.

P_{50} : é o quinquagésimo percentil, coincide com a mediana ($P_{50} = M_d$).

Para o cálculo dos percentis, utilizamos técnicas semelhantes às do cálculo dos quartis e decis. Inicialmente, determina-se a classe que contém o valor a ser calculado pela expressão:

$$\frac{K \sum f_i}{100} \quad (K = 1, 2, 3, \dots, 99)$$

Para a obtenção do percentil, utilizamos a fórmula:

$$P_K = l_{P_K} + \left[\frac{\frac{K \sum f_i}{100} - F_{ant}}{f_{P_K}} \right] \cdot h_{P_K}$$

l_{P_K} = limite inferior da classe do percentil considerado;

F_{ant} = freqüência acumulada da classe anterior à classe do percentil considerado;

h_{P_K} = amplitude do intervalo de classe do percentil considerado;

f_{P_K} = freqüência simples da classe do decil considerado.

Exemplo:

Na tabela do exemplo anterior, vamos calcular o 36º percentil.

$K = 36$: $\frac{36 \sum f_i}{100} = \frac{36 \times 54}{100} = 19,4$ (o trigésimo sexto percentil pertence à terceira classe).

$$P_{36} = l_{P_{36}} + \left[\frac{\frac{36 \sum f_i}{100} - F_{ant}}{f_{P_{36}}} \right] \cdot h_{P_{36}} = 136 + \left[\frac{19,4 - 18}{16} \right] \cdot 8 = 136,7 \text{ cm}$$

Exercício

Uma rede de hotéis tem um gasto salarial com seus funcionários de acordo com a tabela abaixo.

Gasto salarial de uma rede de hotéis com seus funcionários

i	Número de salários mínimos	Número de funcionários (f_i)	Ponto médio (x_i)
1	0 - 2	14	6
2	2 - 4	28	18
3	4 - 6	19	34
4	6 - 8	15	47
5	8 - 10	16	54
6	10 - 12	17	
7	12 - 14	13	
8	14 - 16	9	
9	16 - 18	6	
10	18 - 20	3	
		$\sum f_i =$	

Calcule:

- a) Os quartis;
- b) O 2º, 6º e o 9º decil;
- c) O 18º, 29º, 58º, 72º e o 93º percentil.

UNIDADE V – MEDIDAS DE DISPERSÃO OU DE VARIABILIDADE

5.1. Introdução

A interpretação de dados estatísticos exige que se realize um número maior de estudos, além das medidas de posição. O estudo das médias, medianas, moda, quartis e percentis são válidos, mas não suficientes para estudos comparativos ou conclusões qualitativas.

As medidas de dispersão ou de variabilidade servem para verificar a representatividade das medidas de posição.

Das medidas de dispersão ou de variabilidade estudamos:

- Amplitude total (já estudado na unidade II);
- Variância e desvio-padrão;
- Coeficiente de variação.

5.2. Variância (S^2 ou σ^2)

A variância (S^2) leva em consideração os valores extremos e os valores intermediários, isto é, expressa melhor os resultados obtidos.

A variância relaciona os desvios em torno da média, ou, mais especificamente, é a média aritmética dos quadrados dos desvios.

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n} = \frac{\sum d_i^2}{n} \text{ (variância de uma população)}$$

Sendo:

S^2 = variância

\bar{x} = valor da média aritmética

$d_i = (x_i - \bar{x})$

$n = \sum f_i$

Observação:

Caso a variância represente uma descrição da amostra e não da população (este tipo de ocorrência é mais comum na estatística), o denominador passa a ser $(n - 1)$ em vez de n . O motivo dessa modificação é porque melhora a estimativa do parâmetro de população.

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{\sum d_i^2}{n-1} \text{ (variância de uma amostra)}$$

5.3. Desvio-Padrão (S ou σ)

O desvio-padrão (S) é a medida mais usada na comparação de diferenças entre conjuntos de dados, por ter grande precisão. O desvio-padrão determina a dispersão dos valores em relação à média.

O desvio-padrão é determinado pela fórmula:

$$S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum d_i^2}{n}} \text{ (desvio-padrão de uma população)}$$

Da mesma forma que para a variância, caso o desvio-médio representar uma descrição da amostra e não da população, o denominador passa a ser $(n - 1)$, logo:

$$S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum d_i^2}{n-1}} \text{ (desvio-padrão de uma amostra)}$$

5.3.1. Fórmula Alternativa para o Cálculo do Desvio-Padrão

O valor médio em algumas séries resulta números decimais, conseqüentemente, o cálculo do desvio-padrão pode-se estender numa somatória do quadrado de números decimais. Com o objetivo de simplificar os cálculos matemáticos, utilizamos uma fórmula alternativa para o cálculo do desvio-padrão.

Como $\bar{x} = \frac{\sum x_i}{n}$ e $\sum(x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$, então substituindo esses

valores em $S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$, obtemos:

$$S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{\sum x_i^2}{n} - \frac{(\sum x_i)^2}{n^2}} = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$$

$$\therefore S = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$$

Propriedades do Desvio-Padrão

- Somando ou subtraindo um mesmo valor de todos os valores de uma variável, o desvio-padrão não se altera;
- Multiplicando (ou dividindo-se) todos os valores de uma variável por um mesmo número (diferente de zero), o desvio-padrão fica multiplicado (ou dividido) por esse número.

5.3.2. Desvio-Padrão para Dados Não Agrupados

Exemplo:

Durante determinada semana, os nove vendedores de uma agência de turistas venderam as seguintes quantidades de passagens aéreas: 20; 25; 28; 31; 37; 42; 45; 49; 53. Calcule o valor do desvio-padrão utilizando:

- a) A fórmula convencional;
- b) A fórmula alternativa.

Solução:

a) Utilizando a fórmula convencional, temos:

$$\bar{x} = \frac{20 + 25 + 28 + 31 + 37 + 42 + 45 + 49 + 53}{9} = \frac{330}{9} = 36,67$$

$$S = \sqrt{\frac{\sum_{i=1}^9 (x_i - \bar{x})^2}{9}}$$

$$= \sqrt{\frac{(20 - 36,67)^2 + (25 - 36,67)^2 + \dots + (53 - 36,67)^2}{9}} = \sqrt{\frac{1038}{9}}$$

∴ S = 10,74.

b) Montando a tabela de frequência com os dados fornecidos, obtemos:

Distribuição da venda de passagens aéreas por nove vendedores de uma agência de turismo

Vendedor	Número de passagens vendidas (x_i)	x_i^2
1	20	400
2	25	625
3	28	784
4	31	961
5	37	1.369
6	42	1.764
7	45	2.025
8	49	2.401
9	53	2.809
	\sum 330	\sum 13.138

$$\begin{aligned} \therefore S &= \sqrt{\frac{\sum_{i=1}^9 x_i^2}{9} - \left(\frac{\sum_{i=1}^9 x_i}{9}\right)^2} = \sqrt{\frac{13.138}{9} - \left(\frac{330}{9}\right)^2} \\ &= \sqrt{1459,78 - (36,67)^2} = \sqrt{1459,78 - 1344,44} = 10,74 \end{aligned}$$

5.3.3. Desvio-Padrão para Dados Agrupados Sem Intervalos de Classe

Para dados agrupados sem intervalos de classe, devemos levar em conta as repetições, ou seja, as frequências. O cálculo pode ser feito de duas formas equivalentes:

$$S = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2} \text{ ou } S = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{n}}$$

Se o desvio-padrão representar uma descrição da amostra e não da população, o denominador passa a ser $(n - 1)$ em vez de n .

Exemplo:

Em um encontro de estudantes, foi feito um levantamento sobre o número de dias de permanência para 37 hóspedes, os resultados estão apresentados na tabela abaixo.

Para o cálculo do desvio-padrão aplicando a fórmula

$S = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}$ é conveniente inserir na tabela as colunas contendo os produtos $f_i x_i$ e $f_i x_i^2$.

Número de dias de permanência num hotel para 37 hóspedes

Número de dias de permanência (x_i)	Número de hóspedes (f_i)	f _i x _i	f _i x _i ²
2	3	6	12
4	6	24	96
6	4	24	144
8	2	16	128
10	7	70	700
12	5	60	720
14	4	56	784
16	6	96	1.536
	$\sum f_i = 37$	$\sum f_i x_i = 352$	$\sum f_i x_i^2 = 4.120$

$$\begin{aligned} \therefore S &= \sqrt{\frac{\sum_{i=1}^8 f_i x_i^2}{37} - \left(\frac{\sum_{i=1}^8 f_i x_i}{37}\right)^2} = \sqrt{\frac{4120}{37} - \left(\frac{352}{37}\right)^2} = \sqrt{111,35 - (9,51)^2} \\ &= 4,57 \end{aligned}$$

5.3.4. Desvio-Padrão para Dados Agrupados com Intervalos de Classe

No cálculo do desvio-padrão para dados agrupados com intervalos de classe, considera-se o ponto médio do intervalo.

$$S = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}$$

Para o cálculo do desvio-padrão aplicando a fórmula anterior, é conveniente que a tabela tenha colunas contendo o valor médio do intervalo de classe (x_i) e o produto $f_i x_i$ e $f_i x_i^2$.

Exemplo:

Foi feita uma pesquisa entre as hospedagens e hotéis de uma cidade praiana em que procurou-se apurar os valores praticados na cobrança de uma diária (com direito ao café da manhã).

Valores da diária com café da manhã em algumas hospedagens e hotéis de uma cidade praiana

I	Valor da diária (em R\$)	Hospedagens ou hotéis (f_i)	Ponto médio x_i	$f_i x_i$	$f_i x_i^2$
1	30 † 50	4	40	160	6400
2	50 † 70	8	60	480	28.800
3	70 † 90	5	80	400	32.000
4	90 † 110	7	100	700	70.000
5	110 † 130	3	120	360	43.200
6	130 † 150	5	140	700	98.000
		$\sum f_i = 32$		$\sum f_i x_i = 2.800$	$\sum 278.400$

$$S = \sqrt{\frac{\sum_{i=1}^6 f_i x_i^2}{n} - \left(\frac{\sum_{i=1}^6 f_i x_i}{n}\right)^2} = \sqrt{\frac{278.400}{32} - \left(\frac{2.800}{32}\right)^2} = \sqrt{8.700 - (87,50)^2}$$

$$= 32,31$$

$$\therefore S = 32,31$$

5.4. Coeficiente de variação (CV)

O coeficiente de variação é a relação entre o desvio-padrão (S) e a média aritmética (\bar{x}), multiplicada por 100.

$$CV = \frac{S}{\bar{x}} \times 100$$

Utilizamos o coeficiente de variação na comparação do grau de concentração em torno da média para séries distintas.

Exemplos:

1) Numa distribuição de valores, em que o valor médio é $\bar{x} = R\$ 450,00$ e o desvio-padrão é $S = R\$19,00$, temos o coeficiente de variação:

$$CV = \frac{S}{\bar{x}} \times 100 = \frac{19}{450} \times 100 = 4,222$$

$$\therefore CV = 4,2\%$$

O coeficiente de variação que caracteriza a dispersão dos dados em termos relativos ao valor médio é de 4,2%.

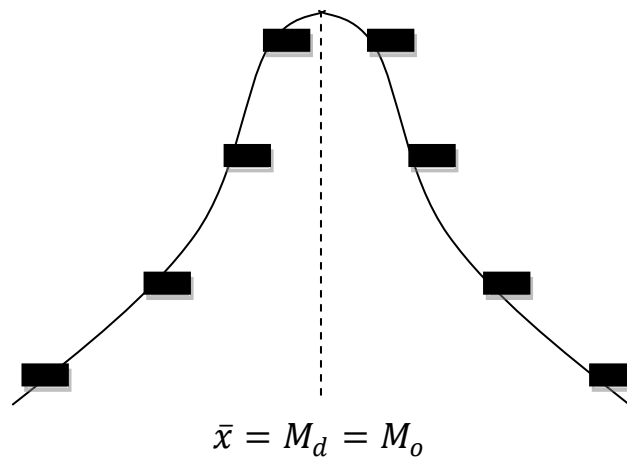
2) Numa cidade **A**, a temperatura média do ano é $\bar{T}_A = 27^\circ\text{C}$ e o desvio-padrão é 8°C . Numa cidade **B**, a temperatura média do ano é $\bar{T}_A = 24^\circ\text{C}$ e desvio-padrão é 6°C . Qual cidade apresenta a temperatura mais homogênea: a cidade **A** ou a cidade **B**.

$$CV_A = \frac{8}{27} \times 100 = 29,6\% \quad \text{e} \quad CV_B = \frac{6}{24} \times 100 = 25\%$$

Observamos que a cidade **B** apresenta maior homogeneidade nos valores de temperatura ao longo do ano.

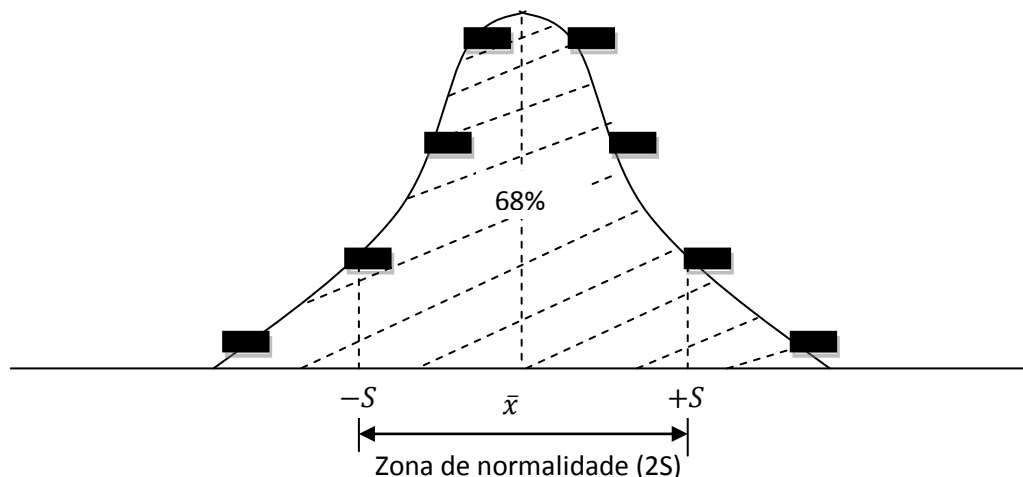
5.5. Significado prático do desvio-padrão

Numa distribuição simétrica, a construção gráfica em forma de sino corresponde a uma curva normal (ou curva de Gauss). Na curva simétrica os valores de média, mediana e moda coincidem com o pico da curva.



5.5.1. Zona de normalidade

É definida por um conjunto de valores (ou uma região) em torno da média aritmética, contidos num intervalo de amplitude “**2S**” (duas vezes o desvio-padrão), ou seja, $-S$ (antes da média) e $+S$ (depois da média). De acordo com alguns estudos matemáticos, essa região engloba 68,26% dos valores da série.



Por outro lado, se for considerado o intervalo de amplitude “**4S**” (quatro vezes o desvio-padrão), abrange-se em torno de 95% dos elementos da série. O intervalo de amplitude “**6S**” (seis vezes o desvio-padrão) abrange em torno de 100% da série.

Exemplo:

Um restaurante cobra o almoço de cada cliente mediante peso (por quilo) da quantidade de alimento consumida. Foi observado, durante um mês, que as quantidades de alimento consumidas são normalmente distribuídas. Se a média consumida for 550g e o desvio-padrão 200g, calcule:

- a) A amplitude do intervalo da zona de normalidade;
- b) A amplitude dos 95% centrais.

Solução:

a) Zona de normalidade : de $(\bar{x} - S)$ até $(\bar{x} + S)$.

Sendo $\bar{x} = 550$ g e $S = 200$ g, temos o intervalo:

$$(\bar{x} - S) = 550 - 200 = 350\text{g}$$

$$(\bar{x} + S) = 550 + 200 = 750\text{g}$$

A amplitude do intervalo da zona de normalidade é de 350g até 750g. Isso significa que: 68% dos clientes do restaurante consomem entre 350g e 750g.

b) Amplitude dos 95% centrais: de $(\bar{x} - 2S)$ até $(\bar{x} + 2S)$.

$$(\bar{x} - 2S) = 550 - 2 \times 200 = 550 - 400 = 150\text{g}$$

$$(\bar{x} + 2S) = 550 + 2 \times 200 = 550 + 400 = 900\text{g}$$

A amplitude dos 95% centrais é de 150g até 900g. Essa amplitude indica que 95% dos clientes consomem entre 150g e 900g.

Exercícios

1) Complete a tabela, calcule o desvio-padrão e o coeficiente de variação das séries com dados agrupados sem intervalo de classe.

Distribuição de freqüência

x_i	f_i	$f_i x_i$	$f_i x_i^2$	F_i
110	2			
140	4			
160	5			
190	8			
	$\sum f_i =$	$\sum f_i x_i =$	$\sum f_i x_i^2 =$	

2) Complete a tabela, calcule o desvio-padrão e o coeficiente de variação das séries de dados agrupados com intervalos de classe.

Distribuição de freqüência

i	Classes	f_i	x_i	$f_i x_i$	$f_i x_i^2$	F_i
1	30 † 56	1				
2	56 † 82	3				
3	82 † 108	4				
4	108 † 134	3				
5	134 † 160	5				
6	160 † 186	2				
		$\sum f_i =$		$\sum f_i x_i =$	$\sum f_i x_i^2 =$	

Observação: O estudo das medidas de assimetria e Curtose são feitos em um curso mais aprofundado de estatística, por esse motivo a parte teórica desse curso termina aqui.

Obrigado pela companhia!
Saiba que você não teve
Apenas mais um professor em sua vida,
Mas você tem agora um amigo,
Para todo sempre!
Até algum dia!Tchau...

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] SPIEGEL, Murray R. **Estatística**. Tradução de Pedro Consentino. São Paulo: McGraw-Hill do Brasil, 1977.
- [2] ARA, Amilton Braio; MUSETTI, Ana Villares; SCHNEIDERMAN, Boris. **Introdução à Estatística**. São Paulo: Edgard Blücher: Instituto Mauá de Tecnologia, 2003.
- [3] NAZARETH, Helenalda. **Curso Básico de Estatística**. São Paulo: Ática, 1995.
- [4] CRESPO, Antonio A. **Estatística Fácil**. 18ª edição. São Paulo: Saraiva, 2002.
- [5] FONSECA, J.S. da; MARTINS G. de A. **Curso de Estatística**. 3ª edição. São Paulo: Atlas, 1982.
- [6] AZEVEDO, A.G. de; CAMPOS, P.H.B. de. **Estatística Básica**. 5ª edição. Rio de Janeiro: LTC, 1987.
- [7] BARBETTA, Pedro Alberto. **Estatística Aplicada à Ciências Sociais**. 3ª edição. Florianópolis: UFSC, 1999.