MATERIAL DE APOIO CURSO DE ESTATÍSTICA DESCRITIVA

Professor Joselias – joselias@uol.com.br – (11) 996541153

CONCEITOS BÁSICOS

1.1 — ESTATÍSTICA

A estatística constitui uma parte da matemática aplicada que tem como finalidade obter conclusões sobre os verdadeiros parâmetros do universo, utilizando para isso a coleta, a organização, a descrição, a análise e a interpretação dos dados.

1.2 — ESTATÍSTICA DESCRITIVA

É o ramo da estatística que se preocupa apenas em descrever os dados observados da amostra, sem se preocupar em fazer previsões sobre os parâmetros do universo. Na estatística descritiva temos a coleta, organização e descrição dos dados.

1.3 — ESTATÍSTICA INFERENCIAL

A estatística inferencial ou estatística indutiva é a parte mais importante da estatística, pois é a inferência estatística que permite a análise e a interpretação dos dados através de estimativas de parâmetros do universo.

1.4 — POPULAÇÃO OU UNIVERSO

É qualquer conjunto de elementos ou indivíduos, com pelo menos uma característica comum ao objeto em estudo.

Exemplo:

- A população de alturas dos candidatos ao concurso de AFRF/2013;
- A população de escolas de estatística no Brasil em 2013;
- A população de computadores em São Paulo.

A população pode ser dita finita ou infinita conforme o número de elementos que possui. Por exemplo a população dos pesos dos candidatos ao concurso do ICMS/2002 é finita. Porém se cada aluno é sorteado e recolocado no conjunto para novo sorteio, teríamos a população de pesos infinita. Na prática consideramos como infinitas aquelas populações com número de elementos muito grande.

1.5 — CENSO

O censo é o processo que consiste no exame de todos os elementos da população. Na prática, a coleta de dados sobre a população requer:

I — Disponibilidade de tempo
III — Precisão dos dados coletados
III — Recursos financeiros
IV — Planejamento das etapas de coleta

Portanto, são muitas as dificuldades para a realização de um censo, logo, nos geralmente utilizamos os processos de amostragem.

1.6 — AMOSTRA

Amostra é qualquer subconjunto não vazio da população.
Para a seleção da amostra devemos tomar cuidado para que a amostra seja representativa da população, considerando a aleatoriedade da seleção e o tamanho da amostra.

1.7 — EXPERIMENTO ALEATÓRIO

Experimentos Aleatórios são aqueles que, repetidos nas mesmas condições, produzem resultados possíveis e diferentes.
Exemplo:
O lançamento de uma moeda honesta várias vezes nas mesmas condições, produz cara ou coroa como resultado, que só pode ser conhecido após o lançamento.

1.8 — EXPERIMENTO DETERMINÍSTICO

Quando o resultado do experimento já está determinado antes de sua realização, portanto não interessa ao estudo da Estatística.

CONCEITOS BÁSICOS

DISTRIBUIÇÃO DE FREQUÊNCIA

2.1 — DADOS ESTATÍSTICOS

Dados estatísticos são todas as informações levantadas (coletadas) que servirão como base para o estudo e análise estatística e que chamaremos de **Dados**.

2.2 — DADOS BRUTOS

Dados Brutos são dados inicialmente coletados que ainda não foram organizados sistematicamente.

2.3 — ROL

Rol é qualquer arranjo de dados brutos em ordem crescente ou decrescente.

2.4 — DISTRIBUIÇÃO DE FREQUÊNCIA

A Distribuição de Frequência é uma disposição de dados numéricos, de acordo com o tamanho ou magnitude dos mesmos. Neste tipo de série não variam local, tempo e o fato. A distribuição de frequência pode ser apresentada por valor (único) ou por grupo de escalares (classes), discriminando a frequência dos mesmos.

Exemplo:

a) Distribuição de freqüência por valor:

título

NOTAS DOS APROVADOS NO AFRF/2000-SP

cabeçalho

NOTAS	FREQÜÊNCIAS
0	2
1	4
2	5
3	4
4	8
5	3
6	9
7	5
8	2
9	8
TOTAL	50

FONTE: ALUNO

rodapé

b) Distribuição de freqüência por classe:

NOTAS DOS APROVADOS NO AFRF/2000-SP

	NOTAS	FREQÜÊNCIAS
classes	0 ⊢—— 20	10
	20 ⊢—— 40	15
	40 ⊢—— 60	50
	60 ⊢—— 80	20
	80 ⊢—— 100	5
	TOTAL	100

FONTE: ALUNO

Obs.: E ——⊣ I
I ⊢—— E
I ⊢——⊣ I
E —— E

E ⊗ **Excluindo**
I ● **Incluindo**

2.5 - ELEMENTOS DE UMA DISTRIBUIÇÃO DE FREQÜÊNCIA

A Tabela abaixo representa as notas de100 alunos aprovados no concurso de AFTN/94 em São Paulo.

NOTAS DOS ALUNOS APROVADOS NO CONCURSO AFTN/94-SP

NOTAS	f	X_i	Freq. Acumuladas		F_r	$F_r(\%)$
			Abaixo de	A partir de		
0 ⊢— 20	10	10	10	100	0,10	10%
20 ⊢— 40	15	30	25	90	0,15	15%
40 ⊢— 60	50	50	75	75	0,50	50%
60 ⊢— 80	20	70	95	25	0,20	20%
80 ⊢—100	5	90	100	5	0,05	5%
TOTAL	100				1,00	100%

Descreveremos a seguir cada coluna:

a) Classe de Frequência
As classes de frequência são os intervalos em que a variável **nota** foi agrupada.
Exemplo:

0 ⊢— 20 - representa as notas desde 0 até quase 20

b) Limites de uma Classe (Li , Ls)
Os limites de classe são os valores ínfimo e supremo da classe, sendo que o limite inferior (Li) o ínfimo da classe e limite superior (Ls) o supremo da classe.
Assim teremos:
O limite inferior da 2ª classe é 20
O limite superior da 3ª classe é 60

c) Intervalo de classe (amplitude de classe) – h
É a diferença entre o limite superior real da classe e o limite inferior real da classe.

$$h = Ls - Li$$

Obs: Quando o limite inferior da classe coincide com o limite superior da classe anterior, ele é chamado de limite real. Caso contrário será chamado de limite aparente, e o limite real será a média aritmética entre eles.

d) Amplitude Total (AT)
É a diferença entre o maior valor e o menor valor da amostra.
No exemplo acima: AT = 100 – 0 = 100.

e) Ponto médio da classe
É a média aritmética entre o limite inferior real e o limite superior real

$$X_i = \frac{\ell_I + L_S}{2}$$

f) Frequência absoluta simples (fi)

Frequência absoluta é o número de observações que ocorreram em determinada classe.

No exemplo acima a Frequência absoluta ou simplesmente Frequência da 3ª classe é 50.

g) Frequência Total (N)

A Frequência total é a soma de todas as frequências absolutas.

$$N = \sum_{i=1}^{k} f_i$$

$$\text{onde} \begin{cases} k - \text{nº de classes} \\ f_i - \text{Freqüência absoluta da i-ésima classe} \\ N - \text{Freqüência total.} \end{cases}$$

h) Frequências Acumuladas

Frequência Acumulada Crescente (ou Frequência acumulada “abaixo de”, ou Frequência acumulada “até”) que representaremos por F^+ é a soma das frequências absolutas anteriores de uma determinada classe. Por exemplo, na tabela acima, a Frequência acumulada crescente da 3ª classe é a Frequência acumulada abaixo de 60 que é 10 + 15 + 50 = 75. Como a classe é do tipo (40 ⊢ 60) poderíamos falar em Frequência acumulada crescente como sendo a Frequência acumulada **até** 60, que é: 10 + 15 + 50 = 75.

Frequência Acumulada Decrescente (ou Frequência acumulada “a partir de”, ou Frequência acumulada “acima de”) que representaríamos por F^- é a soma das frequências absolutas acima de determinado valor de classe. Assim, a Frequência acumulada decrescente de 3ª classe é a Frequência acumulada a partir de 40, que dá 75. Se a classe fosse do tipo ⊣ poderíamos falar em frequência acumulada decrescente como sendo frequência acumulada decrescente acima 40, seria: 50 + 20 + 5 = 75.

i) Frequência Relativa (fr)
É a razão entre a Frequência absoluta e a Frequência total.

$$f_r = \frac{f_i}{N} \quad \text{ou} \quad f_r = \frac{f_i}{\sum f_i}$$

Portanto, a Frequência relativa da 4ª classe 20/100 = 0,20, podemos representar a Frequência relativa em porcentagem que seria 20% e a somatória da frequência absoluta igual a 1ou 100%.

NORMAS PARA APRESENTAÇÃO TABULAR DE DADOS

A apresentação tabular é a apresentação dos dados (ou resultados) de determinado assunto de modo sintético a fim de se obter uma visão global do que vamos analisar.

Exemplo:

Título

CANDIDATOS APROVADOS NO AFRF/2000 POR ESTADO DO BRASIL

Cabeçalho →

ESTADOS	CANDIDATOS
MARANHÃO	50
PIAUÍ	100
CEARÁ	120
R. G. DO NORTE	40
PARAÍBA	110
MINAS GERAIS	80
SÃO PAULO	200
RIO DE JANEIRO	250
R. G. DO SUL	50
TOTAL	**1000**

Corpo

Rodapé → FONTE : CURSO PRÉ-FISCAL

SÉRIES ESTATÍSTICAS

Chamamos de série estatística ao quadro de distribuição de dados estatísticas em função da época, do local ou da espécie do fenômeno.
Sendo assim teremos:

a) série histórica, ou temporal, ou cronológica
Aquela em que a variável é o tempo, permanecendo fixos o local e a espécie do fenômeno.

Exemplo:

EXPORTAÇÕES BRASILEIRAS DE CARROS ENTRE 1997-2001

ANOS	VALOR (US$ 1 MILHÃO)
1997	100
1999	200
1999	400
2000	500
2001	300

FONTE: BANCO DO BRASIL

b) Série territorial, ou geográfica, ou de localização
Aquela em que a variável é o **local**, permanecendo fixos o tempo e a espécie do fenômeno.

Exemplo:

EXPORTAÇÕES BRASILEIRAS — 2000

PAÍSES DE DESTINO	VALOR (US$ 1 MILHÃO)
ESTADOS UNIDOS	200
CANADÁ	100
ALEMANHA	150
ITÁLIA	100
INGLATERRA	300

FONTE: BANCO DO BRASIL

c) Série categórica, ou específica
Aquela em que variam as espécies ou categórica do fenômeno mantendo-se fixos o tempo e o local.

Exemplo:

REBANHO BRASILEIRO

ESPÉCIE	QUANTIDADE (1.000 CABEÇAS)
BOVINOS	5.000
SUÍNOS	9.000
OVINOS	2.000
CAPRINOS	1.000

FONTE: MINISTÉRIO DA ECONOMIA

d) Série conjugada
Chamamos de séries conjugadas aquelas onde são cruzados 2 (dois) ou mais tipos de séries, podendo ter, assim, duas ou mais entradas.

Exemplo:

POPULAÇÃO DE CÃES BRASILEIROS (100 UNIDADES)

UNID. DA FEDERAÇÃO	ANOS			
	1998	1999	2000	2001
AMAZONAS	100	150	200	90
MARANHÃO	30	400	300	350
----------------	------	-----	------	------
----------------	------	-----	------	------
SÃO PAULO		195	------	------
----------------	------	------	------	------
----------------	------	------	------	------
----------------	------	------	------	------
----------------	------	------	------	------

FONTE: IBGE

REPRESENTAÇÃO GRÁFICA

Primeiramente veremos a representação gráfica de uma distribuição de Frequência.

a) Histograma

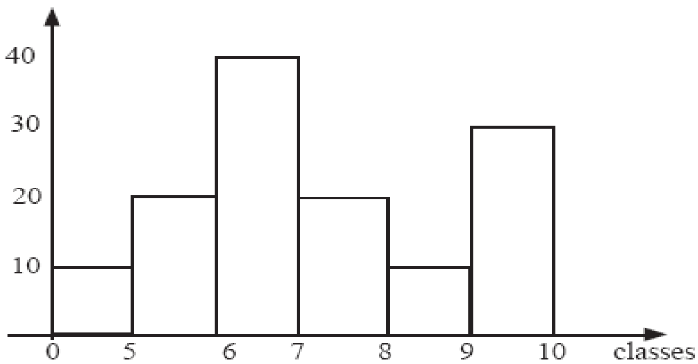
É a representação gráfica de uma **distribuição de Frequência** através de retângulos justapostos de forma que a área de cada retângulo é proporcional à Frequência da classe que ele representa e as bases de cada retângulo iguais às amplitudes das classes que elas representam.

Exemplo:

Seja a distribuição de freqüência:

CLASSES	f_i	X_i	F_{AC}^{+}	F_{AC}^{-}
0 ⊢ 5	10	2,5	10	130
5 ⊢ 6	20	5,5	30	120
6 ⊢ 7	40	6,5	70	100
7 ⊢ 8	20	7,5	90	60
8 ⊢ 9	10	8,5	100	40
9 ⊢ 10	30	9,5	130	30
Total	130			

Então teremos o histograma:

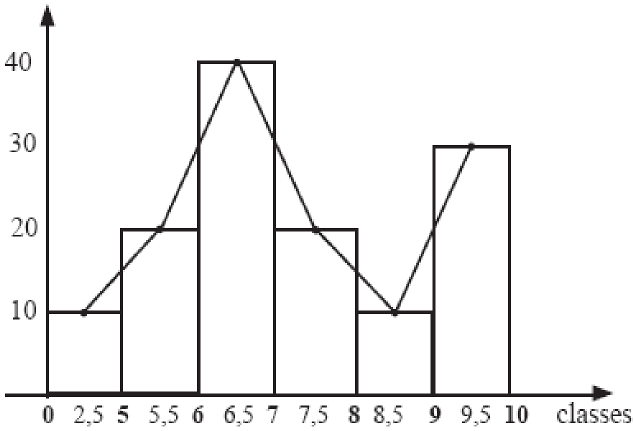


b) Polígono de frequência

É o gráfico obtido quando se une os pontos médios das bases superiores dos retângulos de um histograma, através de segmentos de retas consecutivas.

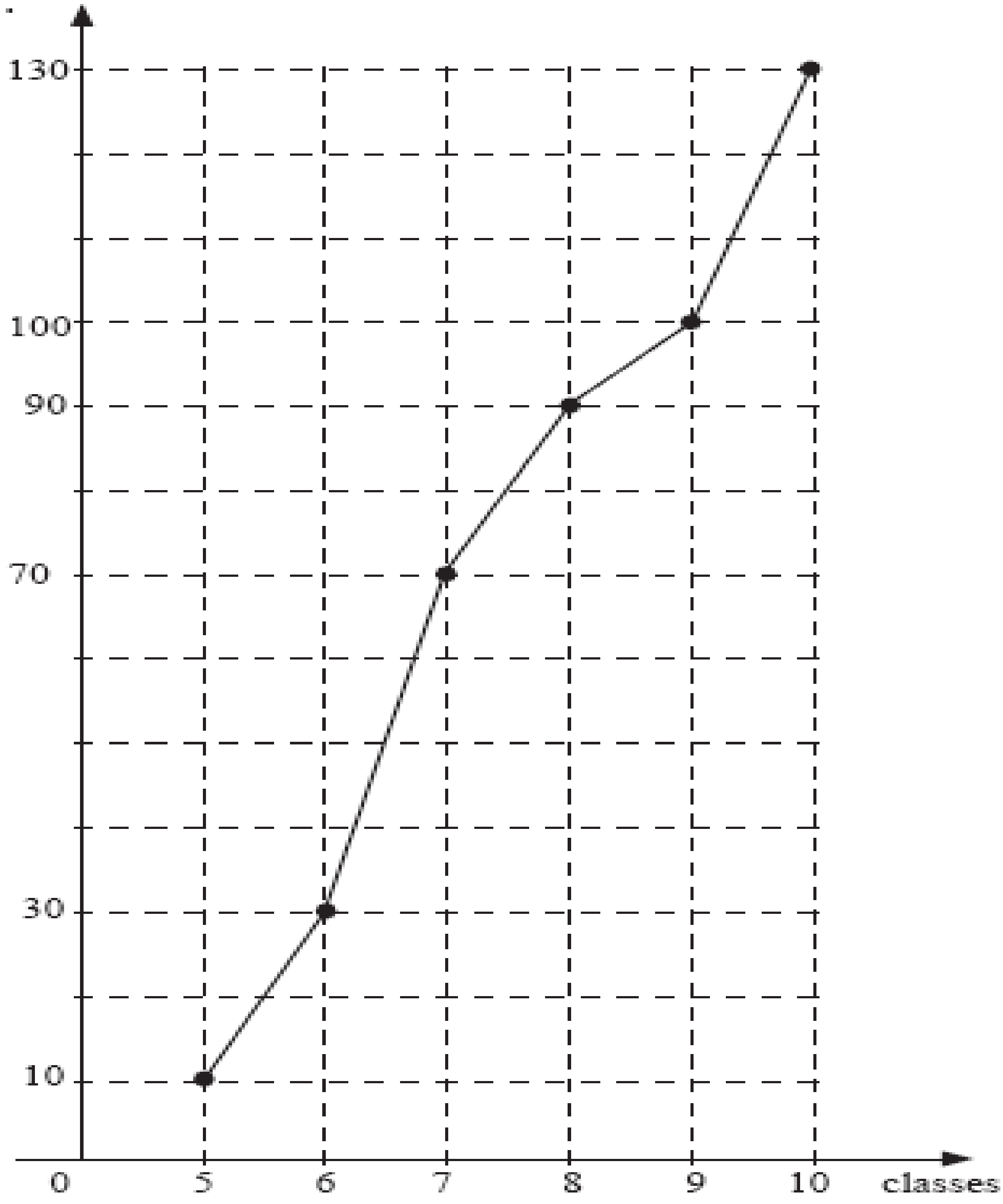
Exemplo:

O polígono de Frequência no exemplo anterior seria:



Observação: Para finalmente fechar o polígono de Frequência devemos unir os pontos do polígono aos pontos médios das classes anterior e posterior supostas, até atingir os limites e superfícies correspondentes.

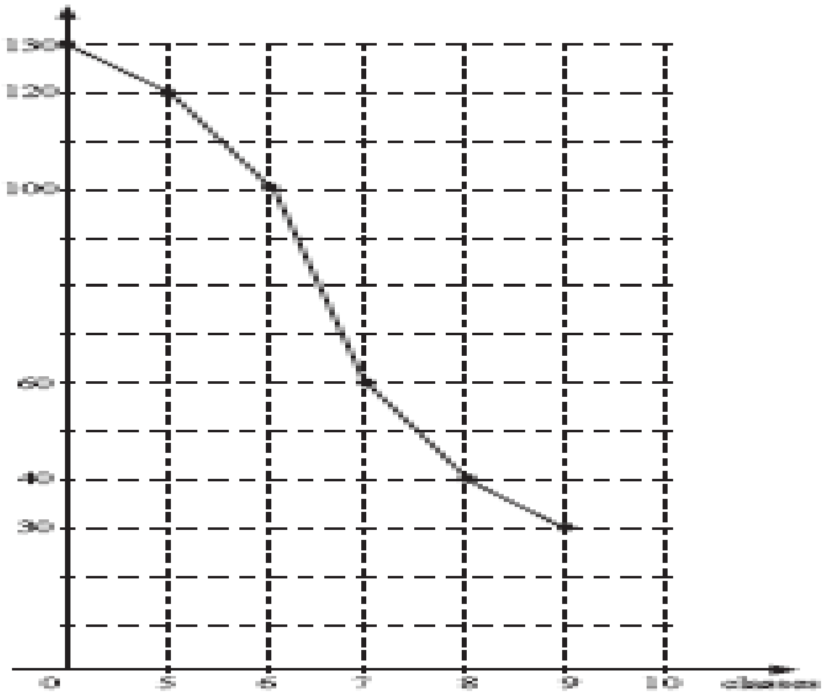
c) Ogivas Crescentes
É o gráfico construído através da frequência acumulada crescente.



c) Ogivas Decrescentes

É o gráfico construído através da freqüência acumulada decrescente.

Observação: Ogivas Crescentes e Decrescentes se cruzam na mediana.



GRÁFICOS DE SÉRIES ESTATÍSTICAS

a) Gráfico de Barras (ou colunas)

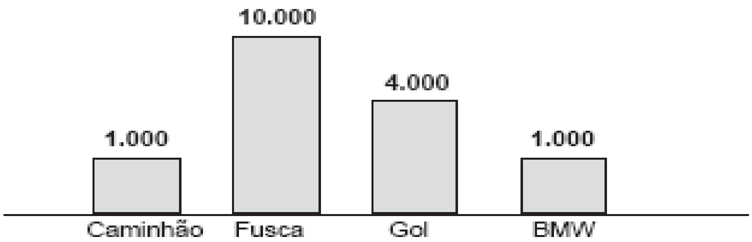
São gráficos que utilizam barras horizontais ou verticais para representar a magnitude dos dados e

estatísticos.

Exemplo:

Produção de carros 2000 - São Paulo

Espécie	Produção (mil)
Caminhão	1.000
Fusca	10.000
Gol	4.000
BMW	1.000



Obs.: Pode também ser feito na vertical.

b) Gráficos Pictóricos

São gráficos que utilizam figuras para representar a magnitude dos dados.

c) Gráficos de Setores

São gráficos que evidenciam uma parte do todo.

Exemplo:

Aprovados no INSS/2002

Estados	Aprovados
RJ	50
SP	500
MG	150
RS	200
CE	100

Fonte: ESAF

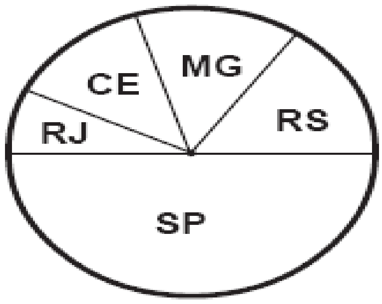
Modo de Calcular:

RJ

1.000 —— 360º

50 —— X

X = 18º

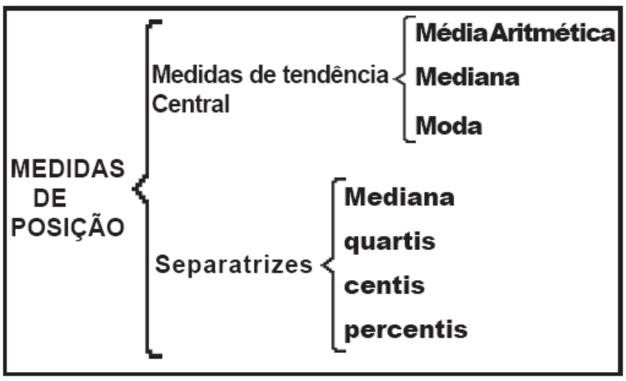


MEDIDAS DE POSIÇÃO

3.1 –INTRODUÇÃO

As medidas de posição irão nos mostrar como estão concentrados os nossos dados. Essas medidas dividem-se em medidas de tendência central, que se caracterizam pelo fato dos dados tenderem a se concentrar em valores centrais, e as medidas conhecidas como separatrizes.

Sendo assim teremos:



Inicialmente veremos as medidas de tendência central.

3.2 – MÉDIA ARITMÉTICA ($\overline{X}$)

Chamamos de média aritmética a razão entre a soma dos dados assumidos pela variável e o número de dados considerados.

Sendo assim temos:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

onde: $\overline{X}$ – *média aritmética*
X_i – *os valores observados da variável*
n – *número de valores*

Obs: Anotação $\sum_{i=1}^{n} X_i$ significa a soma de todos os valores X_is para **i** variando de **1** até **n**. Isto é,

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + X_3 + \ldots\ldots\ldots\ldots + X_n$$

3.2.1 – MÉDIA ARITMÉTICA PARA DADOS NÃO AGRUPADOS

No caso dos dados não estarem agrupados em uma distribuição, teremos a fórmula semelhante à da definição.

Exemplo 1:

Sabendo que a produção de pães diária em uma padaria, durante uma semana foi, 105, 102, 108, 104, 106, 107 e 103 pães, temos a produção média da semana como:

$$\overline{X} = \frac{105+102+108+104+106+107+103}{7}$$

$$\overline{X} = \frac{735}{7}$$

$$\boxed{\overline{X} = 105 \text{ pães}}$$

Exemplo 2:

O número de filhos de 5 funcionários de uma empresa é 1, 9, 2, 8 e 0 filhos, temos que a média de filhos, nesse escritório, é

$$\overline{X} = \frac{1+9+2+8+0}{5}$$

$$\overline{X} = \frac{20}{5}$$

$$\boxed{\overline{X} = 4 \text{ filhos}}$$

3.2.2 — DESVIO EM RELAÇÃO À MÉDIA ARITMÉTICA

Chamamos de desvio em relação à média aritmética, ou simplesmente de desvio em relação à média, a diferença entre cada valor observado da variável e a sua média aritmética.

Isto é,

Sendo $X_1, X_2, X_3, \ldots\ldots, X_n$, os dados observados da variável e $\overline{X}$ a respectiva média aritmética,

E os desvios em relação a média como sendo:

$$d_1 = X_1 - \overline{X}$$
$$d_2 = X_2 - \overline{X}$$
$$d_3 = X_3 - \overline{X}$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$d_n = X_n - \overline{X}$$

Teremos;

$$\mathbf{d_i = \overline{X_i} - X}$$

Onde d_i é o desvio do i-ésimo dado em relação à média aritmética.

Exemplo 8:

Suponhamos que cada funcionário do exemplo 2 tivesse mais 3 filhos, então:

$X_1 = 1,\ \ X_2 = 9,\ \ X_3 = 2,\ \ X_4 = 8,\ \ X_5 = 0$

Aumentando em mais 3 filhos, cada funcionário, teríamos

$Y_1 = 4, Y_2 = 12, Y_3 = 5, Y_4 = 11, Y_5 = 3$

Observe que $Y_i = X_i + 3$

Logo $\overline{Y} = \overline{X} + 3 \ \backslash \ \ \overline{Y} = 4 + 3 \ \ \backslash \ \ \boxed{\overline{Y} = 7 \text{ filhos}}$

3ª. PROPRIEDADE

Multiplicando-se (ou dividindo-se) por uma constante (**K**), todos os valores de uma variável, a nova média aritmética se altera, fica multiplicada (ou dividida) pela constante.

Isto é:

Se $\mathbf{Y_i = K \cdot X_i}$ então $\mathbf{\overline{Y} = K \cdot \overline{X}}$

Se $\mathbf{Y_i = \dfrac{X_i}{K}}$ então $\mathbf{Y = \dfrac{\overline{X}}{K}}$ (K ¹ 0)

Exemplo 9:

Suponha que o padeiro gostaria de dobrar a produção diária de pães, então teríamos:

$X_1 = 105, X_2 = 102, X_3 = 108, X_4 = 104, X_5 = 106, X_6 = 107, X_7 = 103$

dobrando a produção diária, teríamos:

$Y_1 = 210, Y_2 = 204, Y_3 = 216, Y_4 = 208, Y_5 = 212, Y_6 = 214, Y_7 = 206$

Observe que $Y_i = 2 \cdot X_i$

Logo $\overline{Y} = 2 \cdot \overline{X} \quad \backslash \quad \overline{Y} = 2 \cdot 105 \quad \backslash \quad \boxed{\overline{Y} = 210 \text{ pães}}$

Exemplo 10:

Suponhamos que os funcionários do **exemplo 2** triplicassem o nº de filhos, daí,

$X_1 = 1, \quad X_2 = 9, \quad X_3 = 2, \quad X_4 = 8, \quad X_5 = 0$

Triplicando o nº de filhos, teríamos:

$Y_1 = 3, \quad Y_2 = 27, \quad Y_3 = 6, \quad Y_4 = 24, \quad Y_5 = 0$

Observe que: $Y_i = 3 \cdot X_i$

Logo $\overline{Y} = 3 \cdot \overline{X} \quad \backslash \quad \overline{Y} = 2 \cdot 4 \quad \backslash \quad \boxed{\overline{Y} = 12 \text{ filhos}}$

3.3 –MÉDIA ARITMÉTICA PARA DADOS AGRUPADOS

3.3.1- CONSIDERE OS DADOS AGRUPADOS EM UMA DISTRIBUIÇÃO DE FREQÜÊNCIA ABAIXO:

VALORES (X_i)	FREQÜÊNCIA ABSOLUTA (f_i)
X_1	f_1
X_2	f_2
X_3	f_3
.	.
.	.
.	.
.	.
X_k	f_k
Total	$\sum_{i=1}^{K} f_i$

Então a média aritmética da distribuição acima terá a seguinte fórmula:

$$\overline{X} = \frac{\sum_{i=1}^{K} X_i f_i}{\sum_{i=1}^{K} f_i}$$

A partir desse ponto iremos suprimir os índices no símbolo de somatório para facilitar a notação, sendo assim, teremos:

$$\overline{X} = \frac{\sum X_i f_i}{\sum f_i} \quad \text{ou} \quad \overline{X} = \frac{\sum X_i f_i}{N}$$

Exemplo 11:

Consideremos a distribuição do número de filhos de uma determinada classe de alunos

NÚMEROS DE FILHOS (X_i)	ALUNOS (f_i)
0	5
1	10
2	20
3	10
4	5
Total	50

Então o método mais fácil de se calcular a média é fazer mais uma coluna $X_i f_i$, isto é:

X_i	f_i	$X_i f_i$
0	5	0
1	10	10
2	20	40
3	10	30
4	5	20
Total	50	100

Então, temos:

$$\sum X_i f_i = 100 \quad \text{e} \quad \sum f_i = 50, \quad \text{logo}$$

$$\overline{X} = \frac{\sum X_i \cdot f_i}{\sum f_i} \quad \therefore \quad \overline{X} = \frac{100}{50} \quad \therefore \quad \boxed{\overline{X} = 2 \text{ filhos}}$$

Exemplo 12:

Considere as notas de 100 alunos aprovados em concurso público.

NOTAS (X_i)	ALUNOS (f_i)
2	15
4	18
6	47
8	15
9	5
Total	100

Então vamos considerar mais uma coluna de $X_i f_i$

X_i	f_i	$X_i f_i$
2	15	30
4	18	72
6	47	252
8	15	120
9	5	45
Total	100	549

Logo:

$$\sum X_i f_i = 549 \quad e \quad \sum f_i = 100, \quad logo$$

$$\overline{X} = \frac{\sum X_i \cdot f_i}{\sum f_i} \quad \therefore \quad \overline{X} = \frac{549}{100} \quad \therefore \quad \boxed{\overline{X} = 5{,}49}$$

3.3.2 - CONSIDEREMOS AGORA OS DADOS AGRUPADOS ABAIXO CONFORME UMA DISTRIBUIÇÃO DE FREQÜÊNCIA POR CLASSE

Classes	Freqüências aparentes (f_i)
$\ell_1 \vdash L_1$	f_1
$\ell_2 \vdash L_2$	f_2
$\ell_3 \vdash L_3$	f_3
.	.
.	.
.	.
.	.
$\ell_k \vdash L_k$	f_k
Total	$\sum f_i$

Então a média aritmética para a distribuição de freqüência em classe acima terá a seguinte fórmula:

$$\overline{X} = \frac{\sum X_i f_i}{\sum f_i}$$

onde X_i — ponto médio da i-ésima classe.

Exemplo 13:

Considere as alturas de 50 indivíduos de uma empresa, conforme a distribuição abaixo:

Alturas (cm)	Indivíduos (f_i)
150 ├── 160	5
160 ├── 170	10
170 ├── 180	20
180 ├── 190	10
190 ├── 200	5
Total	50

Então o método mais fácil é considerar duas colunas, dos pontos médios (X_i) e do produto do ponto médio pela freqüência absoluta ($X_i f_i$), isto é,

Classe	f_i	X_i	$X_i f_i$
150 ├── 160	5	155	775
160 ├── 170	10	165	1.650
170 ├── 180	20	175	3.500
180 ├── 190	10	185	1.850
190 ├── 200	5	195	975
Total	50	X	8.750

Temos, então:

$$\sum X_i f_i = 8.750 \quad \text{e} \quad \sum f_i = 50, \quad \text{logo}$$

$$\overline{X} = \frac{\sum X_i \cdot f_i}{\sum f_i} \quad \therefore \quad \overline{X} = \frac{8.750}{50} \quad \therefore \quad \boxed{\overline{X} = 175 \text{ cm}}$$

Exemplo 14:

Considere os salários de 100 funcionários de uma empresa, conforme a distribuição abaixo:

Salários Mínimos (S.M.)	Funcionários (f_i)
0 ⊢ 2	20
2 ⊢ 4	40
4 ⊢ 6	20
6 ⊢ 8	15
8 ⊢ 10	5
Total	100

Considerando então mais duas colunas, dos pontos médios (X_i) e do produto do ponto médio pela freqüência absoluta ($X_i f_i$), temos:

Classe (SM)	f_i	X_i	$X_i f_i$
0 ⊢ 2	20	1	20
2 ⊢ 4	40	3	120
4 ⊢ 6	20	5	100
6 ⊢ 8	15	7	105
8 ⊢ 10	5	9	45
Total	100		390

Temos, então:

$$\sum X_i f_i = 390 \quad \text{e} \quad \sum f_i = 100, \quad \text{logo}$$

$$\overline{X} = \frac{\sum X_i \cdot f_i}{\sum f_i} \quad \therefore \quad \overline{X} = \frac{390}{100}$$

$$\boxed{\overline{X} = 3{,}9 \text{ salários mínimos}}$$

Capítulo 4

MODA (M_o)

Chamamos de ***moda*** o valor ou atributo que ocorre com maior freqüência em uma distribuição.

Por exemplo, a nota modal dos alunos de um concurso é a nota mais comum, isto é, a nota que a maioria dos alunos obtiveram.

4.1 - MODA PARA DADOS NÃO AGRUPADOS

Quando temos série de valores não agrupados, a moda é facilmente encontrada, pois pela definição, basta encontrar o valor que mais se repete.

Exemplo 17

4, 2, 6, 4, 3, 5, 7, 9, 4, 10, 8, 4, 3, 2, 4
Mo = 4. **(unimodal)**

Exemplo 18

3, 2, 3, 4, 5, 3, 4, 2, 3, 2, 5, 2.
Neste caso são dois valores (2 e 3) que mais se repetem, e na mesma quantidade.
Portanto, dizemos que a distribuição possui duas modas iguais a 2 e 3, e chamamos de bimodal.

Exemplo 18.1

1, 2, 3, 0, 7, 3, 2, 5, 1, 9, 15
Mo = 1 Mo = 2 Mo = 3 **(multimodal)**

Exemplo 18.2

2, 0, 1, 3, 4, 15, 7
Não existe Moda **(amodal)**

4.2 - MODA PARA DADOS AGRUPADOS

4.2.1 – MODA PARA DADOS AGRUPADOS EM UMA DISTRIBUIÇÃO DE FREQÜÊNCIA.

Quando os dados estiverem agrupados em uma distribuição de freqüência de valores, para acharmos a moda basta observar qual é o valor da variável que possui a maior freqüência.

Exemplo 19

Vamos considerar a distribuição do exemplo 11.

Número de filhos (X_i)	Alunos (f_i)
0	5
1	10
2	20
3	10
4	5

Observamos que o valor 2 filhos possui a maior freqüência (20), logo a moda é 2 filhos

Exemplo 20

Vamos considerar a distribuição do exemplo 12.

Notas (X_i)	Alunos (f_i)
2	15
4	18
6	47
8	15
9	5

Observamos que o valor da nota 6 possui a maior freqüência (47), portanto a nota modal é 6.

4.2.2 — MODA PARA DADOS AGRUPADOS EM UMA DISTRIBUIÇÃO DE FREQÜÊNCIA POR CLASSE

Quando os dados estiverem agrupados em distribuição de freqüência por classe, a moda estará evidentemente na classe que possui a maior freqüência (classe modal).

Se os dados forem agrupados em classe, perdemos o conhecimento dos dados e os respectivos cálculos da média, da moda e da mediana, nesse caso, fazemos uma estimativa entre os limites inferiores e superiores da classe da mesma.

No caso da moda, existem 3 métodos de cálculo da moda:

a) MODA BRUTA

Chamaremos de moda bruta ao ponto médio da classe modal (classe que contém a maior freqüência).

Sendo assim teremos uma fórmula para a moda bruta:

$$M_0 = \frac{\ell_i^* + L^*}{2}$$

onde: ℓ_i^* — limite inferior da classe modal

L^* — limite superior da classe modal

Exemplo 21

Vamos considerar a distribuição do exemplo 13

Alturas (cm)	Indivíduos (f_i)
150 ⊢ 160	5
160 ⊢ 170	10
170 ⊢ 180	20
180 ⊢ 190	10
190 ⊢ 200	5

27

Temos que a classe modal é a classe 170 ⊢ 180, logo $\ell_i^* = 170$ e $L^* = 180$ e a moda bruta será:

$$M_0 = \frac{\ell_i^* + L^*}{2} \rightarrow M_0 = \frac{170 + 180}{2} \therefore \mathbf{M_0 = 175\ cm}$$

Exemplo 22

Vamos considerar a distribuição do exemplo 14

Salários Mínimos (SM)	Funcionários (f_i)
0 ⊢ 2	20
2 ⊢ 4	40
4 ⊢ 6	20
6 ⊢ 8	15
8 ⊢ 10	5

logo, classe modal: 2 ⊢ 4 $\ell_i^* = 2$ $L^* = 4$

Moda bruta: $M_0 = \frac{2+4}{2} \therefore \mathbf{M_0 = 3}$ **salários mínimos**

b) MODA DE CZUBER

Trata-se que uma estimativa, na classe modal, através de uma regra de três, que resulta na seguinte fórmula:

$$M_0 = \ell_i^* + \frac{[f_{max} - f_{ant}].h^*}{[f_{max} - f_{ant}] + [f_{max} - f_{post}]}$$

onde: ℓ_i^* – é o limite inferior da classe modal

f_{max} – é a freqüência absoluta máxima (freqüência de classe modal)

f_{ant} – é a freqüência absoluta anterior à classe modal

f_{post} – é a freqüência posterior à classe modal

h^* – intervalo da classe modal

Exemplo 23

Vamos considerar a distribuição do exemplo 13.

Alturas (cm)	**Indivíduos (f_i)**
150 ├── 160	5
160 ├── 170	10
170 ├── 180	20
180 ├── 190	10
190 ├── 200	5

28

Então, temos:

Classe modal: 170 ⊢ 180

$\ell_i^* = 170$, $h^* = 180 - 170 = 10$

$f_{max} = 20$, $f_{ant} = 10$ $f_{pos} = 10$

logo:

$$M_O = \ell_i^* + \frac{[f_{max} - f_{ant}] \cdot h^*}{[f_{max} - f_{ant}] + [f_{max} - f_{post}]}$$

$$M_O = 170 + \frac{[20 - 10] \cdot 10}{[20 - 10] + [20 - 10]}$$

$$M_O = 170 + \frac{10 \cdot 10}{10 + 10} \rightarrow M_O = 170 + 5 \therefore \mathbf{M_O = 175\ cm}$$

Exemplo 24

Vamos considerar a distribuição do exemplo 14

Salários Mínimos (SM)	Funcionários (f_i)
0 ⊢ 2	20
2 ⊢ 4	40
4 ⊢ 6	20
6 ⊢ 8	15
8 ⊢ 10	5

Classe modal: 2 ⊢ 4

$\ell_i^* = 2$

$h^* = 4 - 2 = 2$

$f_{max} = 40$

$f_{ant} = 20$

$f_{post} = 20$

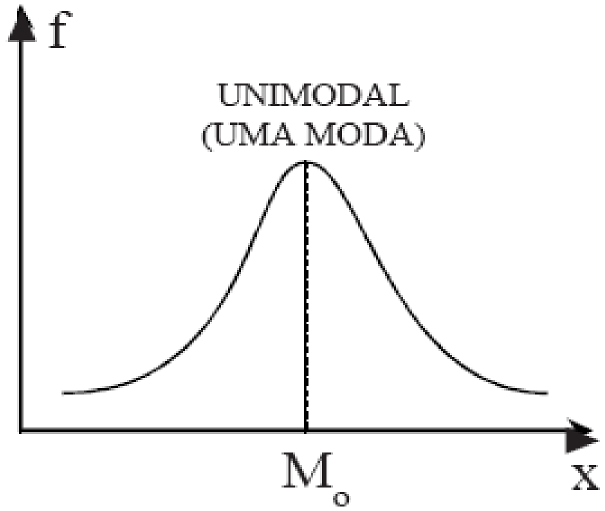
$$M_O = \ell_i^* + \frac{[f_{max} - f_{ant}] \cdot h^*}{[f_{max} - f_{ant}] + [f_{max} - f_{post}]}$$

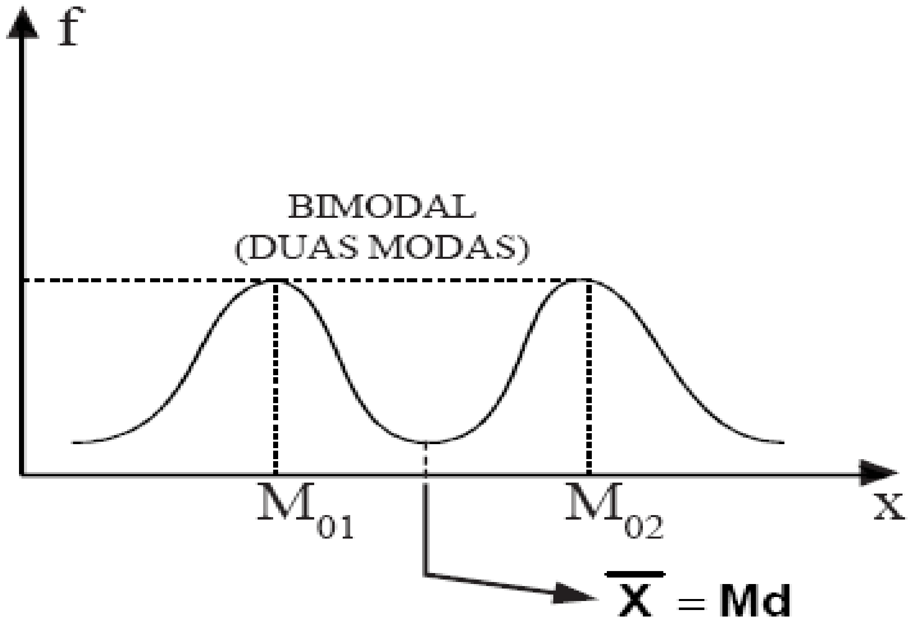
$$M_O = 2 + \frac{[40 - 20] \cdot 2}{[40 - 20] + [40 - 20]}$$

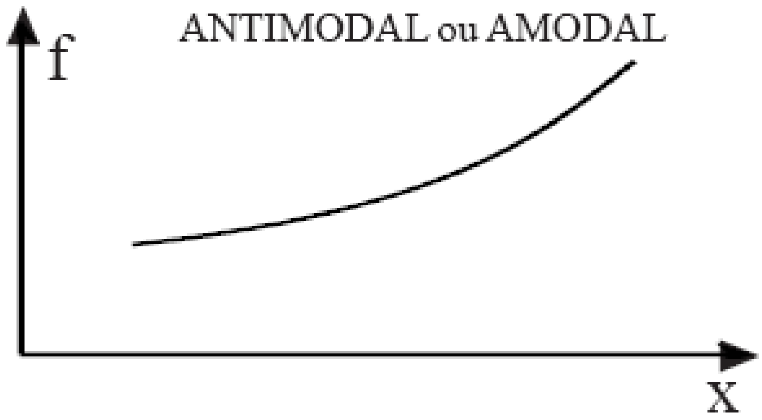
$\mathbf{M_O}$ **= 3 salários mínimos**

A MODA NA CURVA DE FREQÜÊNCIA.

Na curva de freqüência, a moda será o valor que corresponde, no eixo horizontal, ao ponto de freqüência máxima (vertical).







OBS.:

1. Em uma distribuição simétrica e unimodal a Média Aritmética é igual a moda e igual a Mediana.

2. Em uma distribuição simétrica e bimodal apenas a Média Artimética e a Mediana são iguais.

3. Em uma distribuição simétrica e multimodal a Média Artimética e a Mediana são iguais e coincidem apenas como uma das modas.

Capítulo 5

MEDIANA (MD)

A mediana é outra medida de posição, que representa o valor que divide a distribuição em dois conjuntos com o mesmo número de elementos.

5.1. MEDIANA PARA DADOS NÃO AGRUPADOS

A definição é bem clara e fácil de ser interpretada no caso de dados não agrupados.

Exemplo 26

Dado a série de valores

4, 12, 10, 3, 15, 5, 14, 8, 7

a mediana será fácil de ser identificada após encontrarmos o **Rol**.

Rol: 3, 4, 5, 7, 8, 10, 12, 14, 15.

Agora veremos qual é o valor central que divide a distribuição de forma que tenha o mesmo nº de elementos à esquerda e à direita desse valor.

Logo a mediana é 8 (Md = 8), pois temos quatros valores abaixo de 8 e quatro valores acima de 8.
Observe que no exemplo acima havia 9 valores, e nove é um número ímpar.

Portanto, se o rol tem um número ímpar de dados a mediana será justamente o termo central.

Se, porém, houvesse um número par de dados, a mediana, seria então valor entre os dois termos centrais, e por convenção adotamos a mediana como sendo a média aritmética entre os dois termos centrais.

Exemplo 27

Dada a série de valores

6, 4, 10, 5, 12, 3, 20, 7

Existem 8 dados, como o número de dados é par, existem dois termos centrais no Rol.

Rol: 3, 4, 5, 6, 7, 10, 12, 20

↑↑

termos centrais

Logo a mediana será:

$$Md = \frac{6+7}{2}$$

$$\mathbf{Md = 6{,}5}$$

RESUMINDO

- Se a série possui **n** elementos teremos:

a) Se **n** for ímpar, existe um termo central no rol e este termo central do rol será justamente a mediana, que será calculada como sendo o termo de ordem $\frac{n+1}{2}$ do rol.

b) Se **n** for par, existem dos termos centrais no rol, e a mediana será a média aritmética entre esses termos centrais, que será calculado como sendo a média aritmética entre os termos de ordem

$$\frac{n}{2} \text{ e } \frac{n}{2}+1$$

Portanto no ***exemplo 26***, temos n = 9 (ímpar), logo, a mediana será o 5º termo do rol $\left(\frac{n+1}{2}\right)$, isto é, Md = 5º termo do rol = 8.

No ***exemplo 27***, termos n = 8 (par), logo a mediana será a média aritmética entre o 4º e 5º termo do rol $\left(\frac{n}{2}\right)^{o}$ e $\left(\frac{n}{2}+1\right)^{o}$

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

5.2. MEDIANA PARA DADOS AGRUPADOS

5.2.1. MEDIANA PARA DADOS AGRUPADOS EM UMA DISTRIBUIÇÃO DE FREQÜÊNCIA POR VALOR

Vamos ver esta situação através de um exemplo.

Exemplo 28

Nº DE APROVADOS EM UM CONCURSO, POR NOTAS

NOTA	APROVADOS
6	10
7	15
8	2
9	14
10	10
TOTAL	51

Seja N a freqüência total $N = \sum f_i$, portanto N = 51 aprovados, isto é, existem 51 dados, como 51 é ímpar existe um termo central, que é o termo de ordem $26\left(\frac{N+1}{2}\right)$, no rol. Para ser fácil a identificação dele vamos construir a freqüência acumulada crescente.

Nota	freqüência	freq. Acum.
6	10	10
7	15	25
8	2	27
9	14	41
10	10	51
TOTAL	**51**	**X**

Observamos então, na coluna de freqüência acumulada, que o 26º elemento do rol é 8, logo Md = 8.

Se a freqüência total (**N**) fosse par, procederíamos da seguinte forma,

Exemplo 29

Nº de aprovados em um concurso, por nota

Nota	Aprovados
6	10
7	15
8	1
9	14
10	10
TOTAL	**50**

Como a freqüência total (**N**) é 50, existem dois termos centrais (25º e 26º), logo construindo a freqüência acumulada teríamos.

Nota	freqüência	freq. Acum.
6	10	10
7	15	25
8	1	26
9	14	40
10	10	50
TOTAL	**50**	**X**

Logo, a mediana será a média aritmética entre os termos de ordem

$$25^{o}\left(\frac{N}{2}\right) \text{ e } 26^{o}\left(\frac{N}{2}+1\right), \text{ isto é,}$$

$$Md = \frac{7+8}{2}$$

$$\mathbf{Md = 7{,}5}$$

5.2.2. MEDIANA PARA DADOS AGRUPADOS EM UMA DISTRIBUIÇÃO DE FREQÜÊNCIA POR CLASSE

Quando temos os dados agrupados em uma distribuição de freqüência por classe, teremos que determinar qual será o ponto do intervalo de classe em que está compreendida a mediana. Logo, precisamos achar qual a classe que contém a mediana.

Evidentemente a classe que contém a mediana será a que corresponde a freqüência acumulada imediatamente superior a $\left(\frac{N}{2}\right)$.

Tudo que foi dito acima está considerado na fórmula abaixo:

$$Md = \ell_i^* + \frac{\left[\frac{N}{2} - F_{AC.ANT}\right] \cdot h^*}{f^*}$$

Onde:

ℓ_i^* – é o limite inferior da classe que contém a mediana
h^* – o intervalo da classe que contém a mediana
f^* – a freqüência absoluta da classe que contém a mediana
N – Freqüência total
$F_{AC.ANT}$ – Freqüência acumulada crescente anterior à classe que contém a mediana.

Exemplo 30

Considerando o ***exemplo 13***

Altura (Cm)	**Indivíduo (f_i)**
150 ⊢ 160	5
160 ⊢ 170	10
170 ⊢ 180	20
180 ⊢ 190	10
190 ⊢ 200	5
TOTAL	**50**

Primeiramente precisamos fazer a freqüência acumulada crescente.

Alturas (Cm)	**Ìndivíduo (f_i)**	**Freq. Acum.**
150 ⊢ 160	5	5
160 ⊢ 170	10	15
170 ⊢ 180	20	35
180 ⊢ 190	10	45
190 ⊢ 200	5	50
TOTAL	**50**	**X**

Daí, N = 50 \ $\frac{N}{2} = 25$, então a classe que contém a mediana será a classe 170 ├── 180, logo, $\ell_i^* = 170$

$h^* = 10$ $f^* = 20$ $F_{AC.ANT} = 15$

$$Md = \ell_i^* + \frac{\left[\frac{N}{2} - F_{AC.ANT}\right]}{f^*} \rightarrow Md = 170 + \frac{[25-15]\cdot 10}{20}$$

$Md = 170 + 5$ $\therefore$ $\mathbf{Md = 175\ cm}$

Exemplo 31

Considerando os dados do ***exemplo 14*** temos:

Salários minímos (SM)	Funcionários
0 ├── 2	20
2 ├── 4	40
4 ├── 6	20
6 ├── 8	15
8 ├── 10	5
TOTAL	**100**

Primeiramente precisamos fazer a freqüência acumulada crescente

Salários minímos (SM)	Funcionários	Freq. Acum.
0 ├── 2	20	20
2 ├── 4	40	60
4 ├── 6	20	80
6 ├── 8	15	95
8 ├── 10	5	100
TOTAL	**100**	**X**

Daí, N =100 $\therefore$ $\frac{N}{2} = 50$, e então a classe que contém a mediana será a classe 2 ⊢— 4, logo, $\ell_i^* = 2$ $h^* = 2$ $f^* = 40$

$F_{AC.ANT} = 20$

$$Md = \ell_i^* + \frac{\left[\frac{N}{2} - F_{AC.ANT}\right] \cdot h^*}{f^*}$$

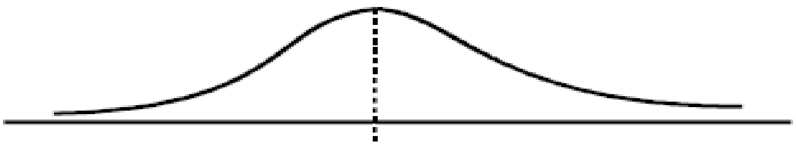
$$Md = 2 + \frac{[50 - 20] \cdot 2}{40}$$

$$Md = 2 + \frac{60}{40} \quad \therefore \quad \mathbf{Md = 3{,}5 \text{ salários mínimos}}$$

5.3. RELAÇÃO ENTRE A MÉDIA ARITMÉTICA, MEDIANA E MODA

Quando a distribuição for unimodal, isto é, a moda for única teremos a seguinte situação.

a) DISTRIBUIÇÃO SIMÉTRICA



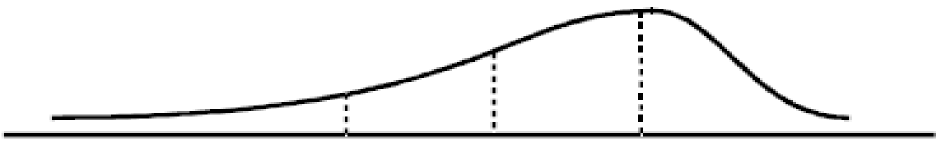
$$\overline{X} = Md = Mo$$

Isto é,

SE A DISTRIBUIÇÃO É SIMÉTRICA E UNIMODAL

ENTÃO: $\overline{X} = Md = Mo$

b) DISTRIBUIÇÃO ASSIMÉTRICA NEGATIVA



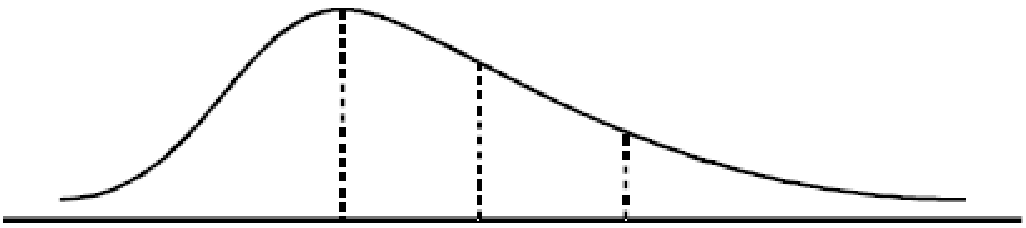
$$\overline{X} < Md < Mo$$

Isto é,

SE A DISTRIBUIÇÃO É ASSIMÉTRICA NEGATIVA

E UNIMODAL ENTÃO: $\overline{X} < Md < Mo$

c) DISTRIBUIÇÃO É ASSIMÉTRICA POSITIVA



$$Mo < Md < \overline{X}$$

Isto é,

SE A DISTRIBUIÇÃO É ASSIMÉTRICA POSITIVA E UNIMODAL ENTÃO: $Mo < Md < \overline{X}$

Capítulo 6

SEPARATRIZES

6.1. QUARTIS (Q)

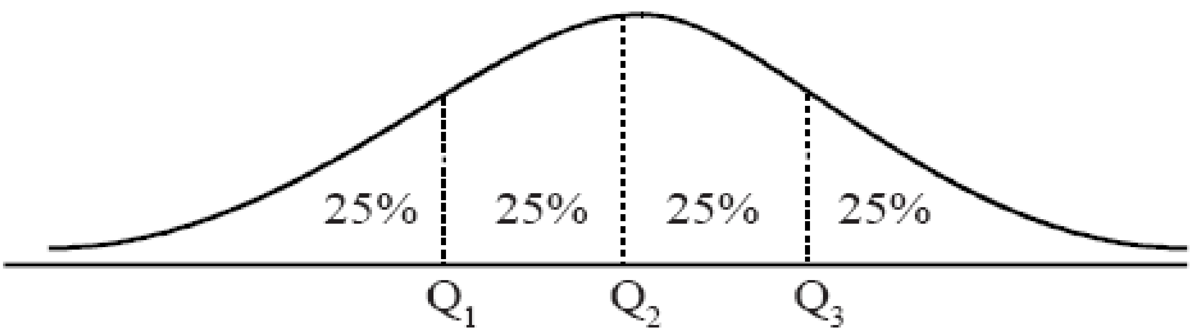
Chamamos de quartis os valores que dividem a distribuição em quatro partes iguais.

Logo, teremos três quartis:

Primeiro quartil (Q_1) será o valor que terá 25% dos dados à sua esquerda e 75% dos dados à sua direita.

Segundo quartil (Q_2) será o valor que terá 50% dos dados à sua esquerda e 50% dos dados à sua direita. Portanto o segundo quartil é a própria mediana.

Terceiro quartil (Q_3) será o valor que terá 75% dos dados à sua esquerda e 25% dos dados à sua direita logo.



Para calcular os quartil, basta, na fórmula da mediana, substituir o $\frac{N}{2}$ por $\frac{KN}{4}$

$$Q_K = \ell_i^* + \frac{\left[\frac{KN}{4} - F_{AC.ANT}\right] \cdot h^*}{f^*}$$

onde K = 1, 2, 3, daí teremos:

$$Q_1 = \ell_i^* + \frac{\left[\frac{N}{4} - F_{AC.ANT}\right] \cdot h^*}{f^*}$$

$$Q_2 = \ell_i^* + \frac{\left[\frac{N}{2} - F_{AC.ANT}\right] \cdot h^*}{f^*}$$

$$Q_3 = \ell_i^* + \frac{\left[\frac{3N}{4} - F_{AC.ANT}\right] \cdot h^*}{f^*}$$

Exemplo 32

Considerando os dados do ***exemplo 14***

Salários minímos (SM)	Funcionários	Freq. Acum.
0 ⊢ 2	20	20
2 ⊢ 4	40	60
4 ⊢ 6	20	80
6 ⊢ 8	15	95
8 ⊢ 10	5	100
TOTAL	**100**	**X**

Cálculo do primeiro quartil (Q_1):

$$K = 1 \quad \therefore \quad \frac{KN}{4} = \frac{100}{4} = 25.$$

Logo, a classe que contém o primeiro quartil será a classe 2 ⊢ 4. daí, $\ell_i^* = 2$ $h^* = 2$ $f^* = 40$ $F_{AC.ANT} = 20$ 46

$$Q_1 = \ell_i^* + \frac{\left[\frac{N}{4} - F_{AC.ANT}\right] \cdot h^*}{f^*}$$

$$Q_1 = 2 + \frac{[25 - 20] \cdot 2}{40}$$

$Q_1 = 2 + 0{,}25$ \ **Q_1 = 2,25 salários mínimos**

Cálculo do segundo quartil (Q_2)

$$K = 2 \quad \therefore \quad \frac{N}{2} = \frac{100}{2} = 50$$

Logo, a classe que contém o primeiro quartil será a classe

2 ⊢ 4. daí, $\ell_i^* = 2$ $h^* = 2$ $f^* = 40$ $F_{AC.ANT} = 20$

$$Q_2 = \ell_i^* + \frac{\left[\frac{N}{2} - F_{AC.ANT}\right] \cdot h^*}{f^*}$$

$$Q_2 = 2 + \frac{[50 - 20] \cdot 2}{40}$$

$Q_2 = 2 + 1{,}5$ \ **Q_2 = 3,5 salários mínimos**

Cálculo do terceiro quartil (Q_3)

$$K = 3 \quad \therefore \quad \frac{KN}{4} = \frac{3 \times 100}{4} = 75$$

logo, a classe que contém o terceiro quartil será a classe

4 ⊢ 6. daí, $\ell_i^* = 4$ $h^* = 2$ $f^* = 20$ $F_{AC.ANT} = 20$

$$Q_3 = \ell_i^* + \frac{\left[\frac{3N}{4} - F_{AC.ANT}\right] \cdot h^*}{f^*}$$

$$Q_3 = 4 + \frac{[75 - 60] \cdot 2}{20}$$

$$Q_3 = 4 + \frac{30}{20}$$

$Q_3 = 4 + 1,5$

Q_3 = 5,5 salários mínimos

Obs.: A Mediana será sempre igual ao 2º quartil.

6.2. DECIS (D)

Chamamos de decis os valores que dividem a distribuição em dez partes iguais.

Logo, teremos **nove decis**:

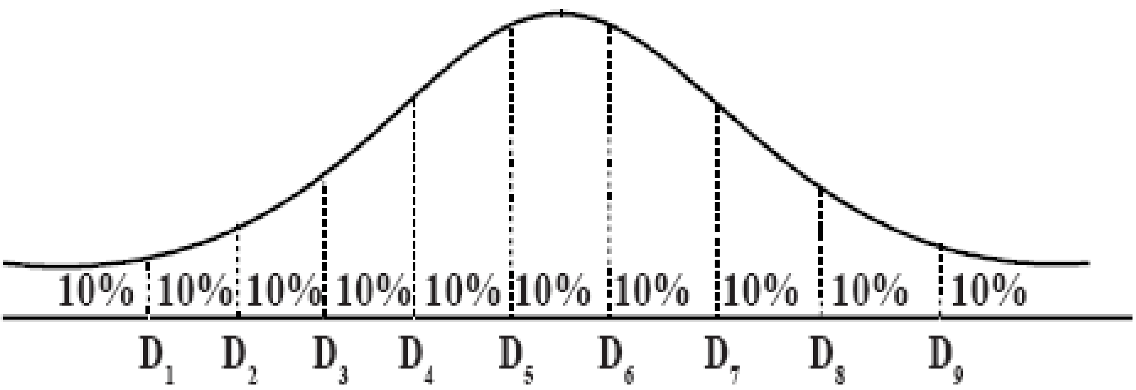
O primeiro decil (**D_1**) será o valor que terá 10% dos dados à sua esquerda e 90% dos dados à sua direita.

O segundo decil **(D_2)** será o valor que terá 20% dos dados à sua esquerda e 80% dos dados à sua direita.

O terceiro decil **(D_3)** será o valor que terá 30% dos dados à sua esquerda e 70% dos dados à sua direita e assim por diante até ...

Nono decil (D_9) será o valor que terá 90% dos dados à sua esquerda e 10% dos dados à sua direita.

.



obs: Note que o quinto decil (D_5), pela definição, coincide com a mediana.

logo:

$$D_K = \ell_i^* + \frac{\left[\dfrac{KN}{10} - F_{AC.ANT}\right] \cdot h^*}{f^*}$$

Para calcular os decis basta substituir na fórmula da mediana o $\frac{N}{2}$ por $\frac{KN}{10}$.

Exemplo 33

Salários minímos (SM)	Funcionários	Freq. Acum.
0 ⊢ 2	20	20
2 ⊢ 4	40	60
4 ⊢ 6	20	80
6 ⊢ 8	15	95
8 ⊢ 10	5	100
TOTAL	**100**	**X**

Calcular o nono decil ($\mathbf{D_9}$)

$$K = 9 \therefore \frac{KN}{10} = \frac{9 \times 100}{10} = 90$$

Logo, a classe que contém o nono decil será a classe

6 ⊢ 8, daí: $\ell_i^* = 6$ $h^* = 2$ $f^* = 15$ $F_{AC.ANT} = 80$

$$D_9 = \ell_i^* + \frac{\left[\frac{9N}{10} - F_{AC.ANT}\right] \cdot h^*}{f^*}$$

$$D_9 = 6 + \frac{[90 - 80] \cdot 2}{15} \quad \therefore \quad D_9 = 6 + \frac{20}{15}$$

$D_9 = 6 + 1{,}33$ \ **D_9 = 7,33 salários mínimos**

Obs.: A Mediana será sempre igual ao 5º decil.

6.3. PERCENTIS (P)

Chamamos de percentis os valores que dividem a distribuição em cem partes iguais.

Logo, teremos noventa e nove percentis:

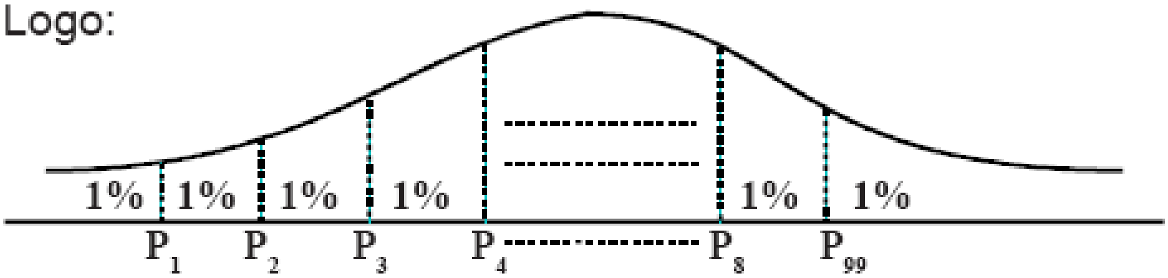
O primeiro percentil (**P_1**) será o valor que terá 1% dos dados à sua esquerda e 99% dos dados à sua direita.

O segundo percentil (**P_2**) será o valor que terá 2% dos dados à sua esquerda e 98% dos dados à sua direita.

E assim por diante até o Nonagésimo nono percentil (**P_{99}**) que será o valor que terá 99% dos dados à sua esquerda e 1% dos dados à sua direita.

Obs: Note que o quinquagésimo percentil (**P_{50}**), pela definição, coincide com a mediana.

Logo:



Para calcular o percentil basta substituir na fórmula da mediana o $\frac{N}{2}$ por $\frac{KN}{100}$

$$P_K = \ell_i^* + \frac{\left[\frac{KN}{100} - F_{AC.ANT}\right] \cdot h^*}{f^*}$$

Exemplo 34

Salários minímos (SM)	Funcionários	Freq. Acum.
0 ⊢ 2	20	20
2 ⊢ 4	40	60
4 ⊢ 6	20	80
6 ⊢ 8	15	95
8 ⊢ 10	5	100
TOTAL	**100**	**X**

Vamos calcular, por exemplo, o qüinquagésimo quinto percentil (P_{55})

$$K = 55 \quad \therefore \quad \frac{KN}{100} = \frac{55 \times 100}{100} = 55$$

Logo, a classe que contém o qüinquagésimo quinto percentil será a classe 2⊢4, daí: $\ell_i^* = 2 \quad h^* = 2 \quad f^* = 40$

$$F_{AC.ANT} = 20$$

$$P_{55} = \ell_i^* + \frac{\left[\frac{55N}{100} - F_{AC.ANT.}\right] \cdot h^*}{f^*}$$

$$P_{55} = 2 + \frac{[55 - 20] \cdot 2}{40}$$

$$P_{55} = 2 + \frac{70}{40} \quad \therefore \quad P_{55} = 2 + 1{,}75$$

P_{55}= 3,75 salários mínimos

Obs.: A Mediana será sempre igual 50º percentil.

Portanto:

$$Q_2 = D_5 = P_{50} = Md$$

Capítulo 7

OUTRAS MÉDIAS

7.1. MÉDIA GEOMÉTRICA (G)

7.1.1. MÉDIA GEOMÉTRICA PARA DADOS NÃO AGRUPADOS

Sejam $X_1,\ X_2,\ X_3, \ldots,\ X_n$ dados não agrupados, então a média geométrica será

$$G = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \cdots\cdots X_n}$$

Exemplo 1

Calcular a média geométrica dos dados: 2, 8, 4, 16, 1.

$$G = \sqrt[5]{2 \cdot 8 \cdot 4 \cdot 16 \cdot 1} \quad \therefore$$

$$G = \sqrt[5]{2^1 \cdot 2^3 \cdot 2^2 \cdot 2^4 \cdot 2^0} \quad \therefore$$

$$G = \sqrt[5]{2^{10}}$$

$$G = 2^2 \quad \therefore \quad \mathbf{G = 4}$$

Exemplo 2

Calcular a média geométrica de 2,6,8.

$$G = \sqrt[3]{2 \cdot 6 \cdot 8} \quad \rightarrow \quad G = \sqrt[3]{96} \quad \backslash \quad \mathbf{G = 4{,}5}$$

Observe que é necessário o uso de máquina de calcular.

7.1.2. MÉDIA GEOMÉTRICA PARA DADOS AGRUPADOS

$$G = \sqrt[N]{X_1^{f_1} \cdot X_2^{f_2} \cdots\cdots X_K^{f_K}}$$

Exemplo 3

Calcular a média geométrica

X_i	f_i	$X_i^{f_i}$
1	2	1
2	4	16
3	5	243
4	6	4.096
TOTAL	17	

$$G = \sqrt[17]{1 \cdot 16 \cdot 243 \cdot 4096}$$

$$G = \sqrt[17]{15925248} \quad \therefore \quad \mathbf{G = 2{,}65}$$

7.2. MÉDIA HARMÔNICA (H)

É o inverso da média aritmética dos inversos dos dados.

7.2.1. MÉDIA HARMÔNICA PARA DADOS NÃO AGRUPADOS

Sejam $X_1,\ X_2,\ X_3, \ldots\ldots,\ X_n$, dados não agrupados. Então a média hamônica será:

$$H = \frac{n}{\displaystyle\sum_{i=1}^{n} \frac{1}{X_i}}$$

Exemplo 4

Calcule a média harmônica dos dados: 2, 3, 4, 1.

$$H = \frac{4}{\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + 1} \quad \therefore \quad H = \frac{4}{\frac{6+4+3+12}{12}}$$

$$H = \frac{4 \times 12}{25} \quad \therefore \quad \mathbf{H = 1{,}92}$$

7.2.2. MÉDIA HARMÔNICA PARA DADOS AGRUPADOS

$$H = \frac{N}{\displaystyle\sum_{i=1}^{K} \frac{f_i}{X_i}}$$

Exemplo 5

X_i	f_i	$1/X_i$	f_i/X_i
1	2	1	2
2	4	0,5	2
3	6	0,33	2
4	5	0,25	1,25
TOTAL	17		7,25

$$H = \frac{N}{\sum \frac{f_i}{X_i}} \quad \rightarrow \quad H = \frac{17}{7{,}25} \quad \therefore \quad \mathbf{H = 2{,}34}$$

Observação importante: $\mathbf{H \le G \le \overline{X}}$

8.3. VARIÂNCIA E DESVIO PADRÃO

No tópico anterior vimos que a amplitude total sofre influência apenas dos valores extremos, por isso procuramos uma medida de dispersão que considere todos os dados.

A variância é uma medida de dispersão que considera o quadrado dos desvios em torno da média aritmética.
Assim teremos:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$$

Isto é,

A variância é a média aritmética dos quadrados dos desvios em torno da média aritmética.

Obs:

a) É fácil mostrar que:

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n}x_i^2 - n\overline{x}^2$$

Portanto podemos escrever a variância como:

$$\sigma^2 = \frac{\sum_{i=1}^{n}x_i^2}{n} - \overline{x}^2$$

Fórmula prática

b) Quando queremos estimar a variância através de uma amostra, consideramos como estimador não tendencioso da variância a estatística

$$s^2 = \frac{\sum_{i=1}^{n}(x_1 - \overline{x})^2}{n-1}$$

É fácil observar que se a variância considera o quadrado dos desvios em torno da média, a sua unidade é o quadrado da unidade original.

Por isso, se criou a medida chamada de desvio padrão (σ), como sendo a raiz quadrada do desvio padrão.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \quad \text{ou} \quad \sigma = \sqrt{\frac{\sum_{i=1}^{n}x^2}{n} - \overline{x}^2}$$

PROPRIEDADES

a) Quando somamos (ou subtraímos) a todos os nossos dados uma constante (**k**), a variância não se altera, continua a mesma.

b) Quando somamos (ou subtraímos) a todos os nossos dados uma constante (**K**), o desvio padrão não se altera, continua o mesmo.

c) Quando multiplicamos (ou dividimos) todos os nossos dados por uma constante (**k**), a nova variância se altera, fica multiplicada (ou dividida) pelo quadrado da constante.

d) Quando multiplicamos (ou dividimos) todos os nossos dados por uma constante (**K**), o novo desvio padrão se altera, fica multiplicado (ou dividido) pelo valor absoluto da constante.

8.3.1. PARA DADOS NÃO AGRUPADOS

VARIÂNCIA $\left(\sigma^2\right)$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} \qquad (1)$$

ou

$$\sigma^2 = \frac{\sum_{i=1}^{n}x_i^2}{n} - \overline{x}^2 \qquad (2)$$

DESVIO PADRÃO (σ)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \qquad (3)$$

ou

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}x_i^2}{n} - \overline{x}^2} \qquad (4)$$

Exemplo:

Vamos aplicar as duas fórmulas para exemplificar.
Seja a variável X: 0, 2, 4, 6, 8

Pela fórmula 1

X: 0, 2, 4, 6, 8

desvios: -4, -2, 0, 2, 4

desvio quadrado: 16, 4, 0, 4, 16

$$\rightarrow \begin{cases} \overline{X} = \dfrac{0+2+4+6+8}{5} \\ \overline{X} = \dfrac{20}{5} \\ \overline{X} = 4 \end{cases}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n} = \frac{16+4+0+4+16}{5}$$

$$\sigma^2 = \frac{40}{5} \quad \therefore \quad \boldsymbol{\sigma^2 = 8}$$

Pela fórmula 2

X_i	X_i^2
0	0
2	4
4	16
6	36
8	64
20	**120**

$$\overline{X} = \frac{\sum X_1}{n} = \frac{20}{5} = 4$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2$$

$$\sigma^2 = \frac{120}{5} - 4^2$$

$$\sigma^2 = 24 - 16 \quad \therefore \quad \boldsymbol{\sigma^2 = 8}$$

Logo, vimos que a variância (σ^2) é 8.

Para calcular o desvio padrão basta saber que o desvio padrão é a raiz quadrada da variância, daí

$$\sigma = \sqrt{8} \quad \therefore \quad \boldsymbol{\sigma = 2{,}83}$$

8.3.2. PARA DADOS AGRUPADOS EM UMA DISTRIBUIÇÃO DE FREQÜÊNCIA

VARIÂNCIA (σ^2)

$$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \cdot f_i}{N}}$$

ou

$$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{n} X_i^2 f_i}{N} - \overline{X}^2}$$

Onde $N = \sum f_i$ (Freqüência Total)

Exemplo:

X_i	f_i
0	10
2	20
4	40
6	20
8	10
Total	**100**

Vamos calcular a variância usando as duas fórmulas, para exemplificar

1ª solução: – Primeiramente calculamos a média aritmética.

X_i	f_i	f_iX_i
0	10	0
2	20	40
4	40	160
6	20	120
8	10	80
Total	**100**	**400**

$$\overline{X} = \frac{\sum X_i f_i}{N} = \frac{400}{100} \quad \therefore \quad \boldsymbol{\overline{X} = 4}$$

Prosseguimos, no cálculo dos desvios e seus quadrados

X_i	f_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$(X_i - \overline{X})^2 f_i$
0	10	–4	16	160
2	20	–2	4	80
4	40	0	0	0
6	20	2	4	80
8	10	4	16	160
Total	**100**			**480**

$$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2 \cdot f_i}{N} \rightarrow \sigma^2 = \frac{480}{100} \therefore \boldsymbol{\sigma^2 = 4{,}8}$$

2ª solução: Como já calculamos a média aritmética, temos:

X_i	f_i	X_i^2	$X_i^2 f_i$
0	10	0	0
2	20	4	80
4	40	16	640
6	20	36	720
8	10	64	640
Total	**100**		**2080**

$$\sigma^2 = \frac{\sum_{i=1}^{n} X_i^2 f_i}{N} - \overline{X}^2 \rightarrow \sigma^2 = \frac{2080}{100} - (4)^2$$

$$\sigma^2 = 20{,}8 - 16 \quad \therefore \quad \boldsymbol{\sigma^2 = 4{,}8}$$

Para o cálculo do desvio padrão temos:

$$\boldsymbol{\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2 \cdot f_i}{N}}}$$

ou

$$\boldsymbol{\sigma = \sqrt{\frac{\sum_{i=1}^{n} X_i^2 f_i}{N} - \overline{X}^2}}$$

Para calcular o desvio padrão basta achar a raiz quadrada da variância

$$\sigma = \sqrt{4{,}8} \quad \therefore \quad \boldsymbol{\sigma = 2{,}19}$$

8.3.3. PARA DADOS AGRUPADOS EM UMA DISTRIBUIÇÃO DE FREQÜÊNCIA POR CLASSE

VARIÂNCIA (σ^2)

$$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2 \cdot f_i}{N}}$$

ou

$$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{n} X_i^2 f_i}{N} - \overline{X}^2}$$

onde: $N = \sum f_i$ — (Freqüência Total)

X_i — ponto médio da i-ésima classe

Exemplo:

Considere as alturas de 50 alunos de uma turma, conforme a distribuição abaixo:

Alturas (cm)	Alunos
150 ⊢ 160	5
160 ⊢ 170	10
170 ⊢ 180	20
180 ⊢ 190	10
190 ⊢ 200	5
Total	**50**

Para exemplificar vamos resolver pelas duas fórmulas. Primeiramente vamos calcular a média aritmética.

Alturas	Alunos	X_i	$X_i f_i$
150 ⊢ 160	5	155	775
160 ⊢ 170	10	165	1650
170 ⊢ 180	20	175	3500
180 ⊢ 190	10	185	1850
190 ⊢ 200	5	195	975
Total	**50**		**8750**

$$\overline{X} = \frac{\sum X_i f_i}{N} = \frac{8750}{50} \quad \therefore \quad \boldsymbol{\overline{X} = 175 \text{ cm}}$$

1ª solução:

Prosseguimos no cálculo dos desvios e seus quadrados.

Altura (cm)	Alunos	X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$(X_i - \overline{X})^2 f_i$
150 ⊢ 160	5	155	–20	400	2000
160 ⊢ 170	10	165	–10	100	1000
170 ⊢ 180	20	175	0	0	0
180 ⊢ 190	10	185	10	100	1000
190 ⊢ 200	5	195	20	400	2000
Total	**50**				**6000**

$$\sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2 f_i}{N} = \frac{6000}{50} \therefore \boldsymbol{\sigma^2 = 120\ \text{cm}}$$

2ª solução:

Altura (cm)	Alunos	X_i	X_i^2	$X_i^2 f_i$
150 ⊢ 160	5	155	24.025	120.125
160 ⊢ 170	10	165	27.225	272.250
170 ⊢ 180	20	175	30.625	612.500
180 ⊢ 190	10	185	34.225	342.250
190 ⊢ 200	5	195	38.025	190.125
Total	**50**			**1.537.250**

$$\sigma^2 = \frac{\sum X_i^2 f_i}{N} - \overline{X}^2 \rightarrow \sigma^2 = \frac{1.537.250}{50} - (175)^2$$

$$\sigma^2 = 30.745 - 30.625 \therefore \boldsymbol{\sigma^2 = 120\ \text{cm}}$$

Para o cálculo do desvio padrão temos:

$$\boldsymbol{\sigma = \sqrt{\frac{\sum_{1=1}^{n} (X_i - \overline{X})^2 \cdot f_i}{N}}}$$

ou

$$\boldsymbol{\sigma = \sqrt{\frac{\sum_{i=1}^{n} X_i^2 f_i}{N} - \overline{X}^2}}$$

Mas para calcular o desvio padrão basta achar a raiz quadrada da variância.

$$\sigma = \sqrt{120} \therefore \boldsymbol{\sigma = 10{,}95\ \text{cm}}$$

8.3.4. PROCESSO BREVE PARA CÁLCULOS DO DESVIO PADRÃO

Analogamente ao cálculo do processo breve para a média aritmética, mudamos a variáveis **x** por **y** tal que

$$\boldsymbol{y = \frac{X_i - X_o}{h}}$$

onde: X_o – é um valor arbitrário
h – intervalo de classe

No exemplo anterior teríamos:

$h = 10$ $X_o = 175\text{cm}$

Alturas (cm)	Alunos	X_i	Y_i	$f_i Y_i$	Y_i^2	$f_i Y_i^2$
150 ⊢ 160	5	155	–2	–10	4	20
160 ⊢ 170	10	165	–1	–10	1	10
170 ⊢ 180	20	175	0	0	0	0
180 ⊢ 190	10	185	1	10	1	10
190 ⊢ 200	5	195	2	10	4	20
Total	**50**		**0**			**60**

$$\sigma = h\sqrt{\frac{\sum_{i=1}^{n} f_i Y_i^2}{N} - \left(\frac{\sum_{i=1}^{n} f_i Y_i}{N}\right)^2}$$

$$\sigma = 10\sqrt{\frac{60}{50} - \left(\frac{0}{50}\right)^2}$$

$$\sigma = 10\sqrt{\frac{60}{50}}$$

$$\sigma = 10\sqrt{1{,}20} \therefore \boldsymbol{\sigma = 10{,}95\ \text{cm}}$$

Exercícios

1) Calcule o desvio médio dos dados estatísticos:
12, 2, 10, 8, 4, 6, 0.

2) Calcule a variância dos dados estatísticos:
12, 2, 10, 8, 4, 6, 0.

3) As notas de dez alunos num exame estão dadas a seguir:

2 , 5, 8, 3, 6 , 5, 8, 7, 6, 10

O desvio médio e a variância dessas notas podem ser expressos, respectivamente, por:
(A) 1,8 e 4,6;
(B) 2,0 e 2,2;
(C) 1,8 e 5,2;
(D) 2,0 e 4,6;
(E) 2,0 e 1,9.

4) Entre os funcionários de um órgão do governo, foi retirada uma amostra de dez indivíduos. Os números que representam as ausências ao trabalho registradas para cada um deles, no último ano, são: **0, 0, 0, 2, 2, 2, 4, 4, 6 e 10**. Sendo assim, o valor do desvio padrão desta amostra é:
a) $\sqrt{3}$
b) $\sqrt{9}$
c) $\sqrt{10}$
d) $\sqrt{30}$

5) Os dados seguintes, ordenados do menor para o maior, foram obtidos de uma amostra aleatória, de 50 preços (X_i) de ações, tomada numa bolsa de valores internacional. A unidade monetária é o dólar americano.
4, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10, 10, 10, 10, 11, 11, 12, 12, 13, 13,14, 15, 15, 15, 16, 16, 18, 23
Os valores seguintes foram calculados para a amostra:

$$\sum_{i=1}^{50} X_i = 490 \quad e \quad \sum_{i=1}^{50} X_i^2 - (\sum_{i=1}^{50} X_i)^2 / 50 = 668$$

Assinale a opção que corresponde à mediana e à variância amostral, respectivamente (com aproximação de uma casa decimal)
a) (9,0 14,0)
b) (9,5 14,0)
c) (9,0 13,6)
d) (8,0 13,6)
e) (8,0 15,0)

6) A média da distribuição dos salários da Cia. Monte Branco corresponde a R$ 3.000,00, com desvio padrão igual a R$ 30,00. Se, por hipótese, os salários de todos os funcionários forem multiplicados por 1,2, comparando-se a nova distribuição com a antiga, é correto afirmar que
(A) a média aumentou para R$ 3.800,00.

(B) a variância fica inalterada, continuando a ser 900 (R$)^2.
(C) a mediana fica multiplicada por 1,44.
(D). a soma algébrica dos desvios em relação à nova média aumenta de valor.
(E) o desvio padrão passou a ser R$ 36,00.

7) **(SEFAZ-SP – FCC 2011)** Seja uma amostra aleatória simples extraída de uma população, com tamanho 10 e representada por X_i; i = 1, 2, ... , 10. Sabe-se que

$$\sum_{i=1}^{10} X_i = 270 \quad e \quad \sum_{i=1}^{10} X_i^2 = 7.803$$

A variância desta amostra apresenta o valor de
(A) 67,3
(B) 63,0
(C) 61,0
(D) 59,7
(E) 57,0

8) (TRT-9ºREG - FCC - 2010) Uma população com 16 valores estritamente positivos X_1, X_2, X_3, . . . , X_{16} , correspondente a um determinado atributo, apresenta as seguintes informações:

$$\sum_{i=1}^{16} X_i = 192 \quad e \quad \sum_{i=1}^{16} X_i^2 = 4.464$$

O elemento X_{10}, tal que $X_{10} = 12$, é retirado da população. Os valores da variância da primeira população e da nova população formada são, respectivamente, iguais a
(A) 144 e 134,40.
(B) 144 e 144.
(C) 135 e 144.
(D) 135 e 135.
(E) 135 e 126.

9) Uma administradora de locação de imóveis, com o objetivo de analisar o mercado em sua região, procedeu às seguintes operações:
I. Multiplicou por dois os valores de todos os aluguéis de sua carteira.
II. Subtraiu R$ 1 200,00 de cada valor encontrado no item I.
III. Dividiu por R$ 1 000,00 cada valor encontrado no item II.
IV. Calculou a média aritmética de todos os valores apurados no item III.
Se o valor encontrado no item IV foi 3/10, então, a média aritmética dos valores dos aluguéis, em reais, é
(A) 2 300
(B) 1 700
(C) 1 500
(D) 1 300
(E) 750

10) Considerando as respectivas definições e propriedades relacionadas às medidas de posição e de variabilidade, é correto afirmar:
(A) Concedendo um reajuste de 10% em todos os salários dos empregados de uma empresa, tem-se também que a respectiva variância fica multiplicada por 1,10.
(B) Definindo coeficiente de variação (CV) como sendo o quociente da divisão do desvio padrão pela respectiva média aritmética (diferente de zero) de uma seqüência de valores, tem-se então que CV também poderá ser obtido dividindo a correspondente variância pelo quadrado da média aritmética.
(C) Subtraindo um valor fixo de cada salário dos funcionários de uma empresa, tem-se que o respectivo desvio padrão dos novos valores é igual ao valor do desvio padrão dos valores anteriores.
(D) Dividindo todos os valores de uma seqüência de números estritamente positivos por 4, tem-se que o respectivo desvio padrão fica dividido por 2.
(E) Em qualquer distribuição de valores em estudo, a diferença entre a mediana e a moda é sempre diferente de zero.

11) (ESAF-AFRF-2002-1) Um atributo W tem média amostral $a \neq 0$ e desvio padrão positivo $b \neq 1$. Considere a transformação Z=(W-*a*)/*b*. Assinale a opção correta.
a) A média amostral de Z coincide com a de W.
b) O coeficiente de variação amostral de Z é unitário.
c) O coeficiente de variação amostral de Z não está definido.
d) A média de Z é *a/b*.
e) O coeficiente de variação amostral de W e o de Z coincidem.

12) A tabela abaixo dá os valores dos preços P_{ti} e quantidades Q_{ti} de quatro itens de consumo A, B, C e D nos tempos $t_1 < t_2$. Os preços estão em reais e as quantidades em unidades apropriadas.

Item	P_{t1}	P_{t2}	Q_{t1}	Q_{t2}
A	10	15	5	4
B	9	11,5	5	4
C	4	5	3	2
D	5	6,5	3	2

Assinale a opção que dá o valor mais próximo do índice de preços de Paasche no tempo t_2 com base em t_1.
a)136
b)137
c)138
d)139
e)136,5

13) A tabela abaixo dá a evolução nos tempos t1 e t2 dos preços, em reais e das quantidades, em unidades apropriadas, de três produtos A, B e C. Assinale a opção que corresponde ao índice de preços de Paasche com base em t1, com duas casas decimais.

Produtos	Preços		Quantidades	
	t1	t2	t1	t2
A	2,20	3,00	50	40
B	2,00	2,00	2	3
C	0,50	0,60	80	100

a)131%
b)202%
c)129%
d)186%
e)154%

14) A tabela abaixo apresenta a evolução de preços e quantidades de cinco produtos.

Ano	1960 (ano base)		1970	1979
	Preço (p_o)	Quant. (q_o)	Preço (p_1)	Preço (p_2)
Produto A	6,5	53	11,2	29,3
Produto B	12,2	169	15,3	47,2
Produto C	7,9	27	22,7	42,6
Produto D	4,0	55	4,9	21,0
Produto E	15,7	393	26,2	64,7
Totais	$\Sigma p_o q_o = 9009{,}7$		$\Sigma p_1 q_o = 14358{,}3$	$\Sigma p_2 q_o = 37262{,}0$

Assinale a opção que corresponde aproximadamente ao índice de Laspeyres para 1979 com base em 1960.
a) 415,1
b) 414,4
c) 398,6
d) 416,6
e) 413,6