

Distribuições de Frequências



Johann Carl Friedrich Gauss (ou *Gauß*) (Braunschweig, 30 de Abril de 1777 — Göttingen, 23 de Fevereiro de 1855) foi um matemático, astrônomo e físico alemão. Conhecido como o *príncipe dos matemáticos*. Muitos consideram Gauss o maior gênio da história da matemática. Seu QI foi estimado em cerca de 240

O encontro de Gauss com o teorema binômio inspirou-o para alguns de seus maiores trabalhos, se tornando Gauss, o primeiro "rigorista". Insatisfeito com o que ele e Bartels encontravam em seus livros, Gauss foi além, e iniciou a análise matemática.

Nenhum matemático anterior tinha a menor concepção do que é agora aceitável como prova, envolvendo o processo infinito. Ele foi o primeiro a ver que, a "prova" que pode levar a absurdos como "menos 1 é igual ao infinito", não é prova nenhuma. Mesmo que, em alguns casos, uma fórmula dê resultados consistentes, ela não tem lugar na matemática, até que a precisa condição sob a qual ela continuará a se submeter, tenha sido determinada consistentemente. O rigor imposto por Gauss à análise matemática a tornou totalmente diferente e superou toda a análise matemática feita por seus antecessores.

FONTE: <http://pt.wikipedia.org/wiki/Brasil>

- ⇒ Dados Brutos e Rol
- ⇒ Diagramas Ramo - e - Folhas
- ⇒ Distribuições de Frequências

3 - DISTRIBUIÇÕES DE FREQUÊNCIAS

No caso de dados provenientes de variáveis qualitativas as tabelas e os gráficos vistos até o momento são considerados boas ferramentas para análise de dados. Quando os dados são provenientes de variáveis qualitativas podem acontecer quatro situações:

1. Um volume muito grande de dados e uma variedade grande de valores;
2. Um volume muito grande de dados e uma variedade pequena de valores;
3. Um volume pequeno de dados e uma variedade grande de valores (comparável ao volume de dados); e
4. Um volume pequeno de dados e uma variedade pequena de valores (em comparação com o volume de dados).

Nestes casos as tabelas e os gráficos estudados até então não são considerados boas ferramentas de análise. Para dados provenientes de variáveis quantitativas a Distribuição de Frequências e os Gráficos de Distribuições são ferramentas mais adequadas para análise.

3.1 - DADOS BRUTOS E ROL

Quando um estatístico acaba de realizar uma coleta de dados ele tem em mãos um conjunto de valores totalmente desorganizado numericamente e, além disso, estes valores podem apresentar muitas repetições. Esta lista de valores é denominada **Dados Brutos** é a lista de valores obtidos imediatamente após a coleta. Nesta lista os valores estão desorganizados numericamente e apresentam-se repetidos.

- **Dados Brutos:** *Lista de valores obtidos imediatamente após a coleta, sendo que os valores desta lista estão desorganizados numericamente e apresentam repetições.*

Por exemplo, a lista da Figura 7 a seguir representa as idades de 65 pessoas entrevistadas numa pesquisa de satisfação do cliente realizada num shopping da Grande BH.

13	9	25	7	20	13	7	21	27	23	22	25	25
25	17	13	12	10	27	8	12	12	9	22	22	27
21	8	21	27	11	28	24	22	26	21	17	14	24
18	14	17	12	19	8	9	10	23	21	21	23	7
18	16	7	24	25	13	23	27	21	18	18	14	17

Figura 7.- Lista de Dados Brutos de uma pesquisa de satisfação do cliente

Trabalhando os dados desta lista, o estatístico pode simplesmente organizá-la numericamente, obtendo, assim, uma lista de dados denominada **Rol**.

- **Rol:** *Lista de valores organizados numericamente, ou em ordem crescente ou em ordem decrescente, porém ainda apresentando repetições.*

Na Figura 8 temos o Rol correspondente à lista de Dados Brutos da Figura 7.

7	7	7	8	8	8	8	9	9	9	9	10	10
10	11	11	12	12	12	12	13	14	14	14	15	15
16	17	17	18	18	18	18	19	19	19	20	20	20
21	21	21	21	21	22	22	23	23	24	24	25	25
25	26	26	26	27	27	27	27	27	27	28	28	28

Figura 8.- Rol de valores de uma pesquisa de satisfação do cliente

3.2 - DIAGRAMAS RAMO – E – FOLHAS

Um diagrama Ramo - e - Folhas é uma maneira simples de organizar os dados. Consiste na separação dos dígitos dos valores em duas partes, uma denominada Ramo (correspondente ao dígito mais à esquerda) e outra denominada Folhas (correspondente aos dígitos restantes mais à direita).

Para ilustrar a construção de um Diagrama Ramo – e – Folhas consideremos os dados a seguir.

24; 25; 27; 27; 29; 29; 29; 33; 34; 34; 34; 35; 37; 40; 40; 42; 42; 43; 44; 45; 45; 45; 47; 48; 48; 50; 50; 54; 54; 57 e 58

RAMO (dezenas)	FOLHAS (unidades)
2	4577999
3	344457
4	002234555788
5	004478

O Diagrama de Ramo – e – Folhas pode apresentar outras configurações com a seguir.

RAMO (dezenas)	FOLHAS (unidades)	
2	001344	← valores 20; 20; 21; 23; 24 e 24
2	55568	← valores 25; 25; 25; 26 e 28
3	224444	← valores 32; 32; 34; 34; 34 e 34
3	6778889	
4	111122222333344	
4	5566667777799	
5	0002244	
5	667777	
6	3344	
6	667	
7		
7	1122	

RAMO (dezenas)	FOLHAS (unidades)	
2 — 3	4557*23667	← valores 24; 25; 25; 27; 32; 33; 36; 36 e 37
4 — 5	3359*12223355568	← valores 43; 43; 45; 49; 51; 52....
6 — 7	33466677799*00358	
8 — 9	44677*002444	

3.3 - DISTRIBUIÇÕES DE FREQUÊNCIAS

A maneira mais utilizada para a representação de dados estatísticos são as Distribuições de Frequências. As Distribuições de Frequências são representações tabulares onde os valores coletados aparecem em correspondência com suas respectivas frequências (contagens).

As Distribuições de Frequências são utilizadas quando trabalhamos com um grande volume de dados ou quando esses dados apresentam uma grande variabilidade em seus valores. As Distribuições de Frequências podem apresentar os valores individualmente (Distribuição de Frequências de Valores Individuais) ou em grupos de intervalos (Distribuições de Frequências de Valores Agrupados em Classes). As frequências nada mais são do que a contagem dos valores, no caso destes se apresentarem individualmente, ou a contagem de valores pertencentes a um intervalo, no caso dos valores se apresentarem em grupos de intervalos (Classes).

- **Distribuições de Frequências:** Listam os valores dos dados (individualmente ou em grupos de intervalos) ordenadamente e em correspondência com suas frequências.

3.3.1 - DISTRIBUIÇÕES DE FREQUÊNCIAS DE VALORES DISTINTOS.

São representações onde, apesar do volume de dados ser grande, a variabilidade dos valores observados é pequena. Os valores são dispostos na forma tabular em correspondência com suas frequências.

Suponhamos que um controlador de uma usina hidrelétrica tenha coletados dados referentes ao número de descargas elétricas diárias na região onde está instalada a usina durante todo o ano, obtendo a lista de Dados Brutos a seguir. Apesar de que o número de valores observados seja grande (365), a variabilidade dos valores é pequena (variam de 0 a 8 ocorrências de descargas elétricas por dia).

4	1	2	1	2	2	0	2	3	3	1	2	2	2	2	3	2	1
1	3	0	1	1	2	4	4	2	3	0	1	3	3	4	3	2	2
3	2	3	3	3	3	1	5	1	2	4	0	2	1	4	1	3	2
2	3	1	0	0	2	4	1	3	2	2	1	2	3	5	1	2	1
1	3	3	2	4	2	0	4	0	0	0	2	0	1	2	3	3	3
0	0	1	1	2	1	2	1	2	1	3	2	1	1	4	1	0	3
1	1	2	2	2	4	1	3	1	1	2	0	0	1	2	1	0	1
3	3	2	2	2	3	2	3	3	1	2	5	2	3	1	1	0	4
2	2	3	3	2	6	2	1	0	1	3	3	2	1	0	3	0	0
2	0	3	1	1	1	4	2	3	1	3	2	2	4	2	3	2	0
0	3	2	3	2	4	1	2	2	1	2	3	2	3	1	3	2	3
1	0	1	1	1	1	2	3	1	3	2	1	2	3	0	4	2	5
0	3	4	2	1	2	1	2	3	1	0	2	3	1	0	0	3	3
1	3	2	1	1	2	0	3	1	0	1	1	2	0	2	1	3	2
1	2	3	0	2	2	3	1	2	2	2	2	0	0	5	1	2	1
3	2	2	2	1	2	3	1	1	1	0	1	4	3	3	1	2	1
3	1	2	1	3	2	2	2	3	2	2	1	1	1	3	2	0	0
0	3	2	3	2	3	4	1	1	1	3	3	1	1	2	2	1	3
2	2	0	2	0	2	2	1	2	2	1	3	3	0	3	4	1	3
2	2	3	3	1	4	1	3	1	1	1	4	5	2	2	0	0	2
2	2	4	0	1													

Figura 9.- Número de descargas elétricas no entorno da usina

Uma Distribuição de Frequências de Valores Distintos que representar estes dados é ilustrada na tabela a seguir.

TABELA 9.- DESCARGAS ELÉTRICAS DIÁRIAS NO ENTORNO DA USINA - 2008

DESCARGAS	FREQUÊNCIA
0	47
1	97
2	111
3	80
4	23
5	6
6	1
Σ	365

FONTE: Dados fictícios

Os valores referentes ao número de descargas elétricas são representados por x_i e as frequências com que estes valores ocorreram são representados por f_i , dita Frequência Absoluta Simples.

- **Frequência Absoluta Simples (f_i):** Número de repetições de um determinado valor x_i observado durante a coleta de dados.

3.3.2 - DISTRIBUIÇÕES DE FREQUÊNCIAS DE VALORES AGRUPADOS.

As Distribuições de Frequências de Valores Agrupados são utilizadas tanto quando o volume de valores observados, de uma variável discreta, é grande quanto à variabilidade destes valores também é grande ou quando a variável observada é uma variável contínua. A distribuição, no caso, representará intervalos (Classes) de valores e as frequências representarão as quantidades de valores pertencentes a cada um destes intervalos.

Para ilustrar, consideremos a lista de Dados Brutos da Figura 10 a seguir, que representa os valores referentes às variações nos pesos (em gramas) de 400 sacos de açúcar de 5 kg de certa marca.

2,87	6,94	23,45	2,52	17,43	10,20	1,10	17,20	1,88	24,61	1,28	13,51	2,29	3,45	3,02	3,31	8,16	11,98	0,24	4,86
1,39	10,86	6,56	16,97	13,06	2,48	37,82	3,16	20,59	2,27	35,54	2,20	7,60	7,11	8,73	0,55	0,47	18,58	6,80	36,33
10,79	5,56	31,58	35,60	0,34	6,69	9,11	6,63	11,59	5,82	33,58	2,38	2,74	3,12	0,46	1,52	9,34	2,67	1,30	14,79
26,94	5,17	0,83	0,39	21,38	3,68	3,93	7,44	8,78	5,24	4,69	18,06	9,38	0,48	4,09	4,00	13,03	20,23	7,99	1,43
0,73	2,80	6,15	5,21	6,05	0,63	30,03	20,95	7,86	9,44	1,13	6,58	3,40	4,62	0,36	3,87	21,32	8,96	0,12	2,28
3,63	4,43	0,87	0,53	3,34	3,69	8,77	6,85	1,42	4,57	5,49	31,28	2,03	17,39	9,52	2,87	0,58	12,83	4,63	2,23
4,89	46,71	12,73	8,38	3,98	2,87	4,88	9,22	8,61	4,52	2,76	1,50	5,41	0,14	4,41	7,27	0,76	1,88	19,38	1,42
2,55	14,75	2,85	9,29	11,18	1,69	24,37	36,99	1,34	7,67	14,84	13,24	11,49	8,55	6,94	12,79	0,13	8,51	3,29	11,85
3,82	9,89	33,53	15,22	1,86	0,50	8,44	2,13	29,96	11,81	0,38	1,83	24,87	1,45	5,68	15,23	14,54	1,52	3,49	18,29
9,56	10,66	3,78	3,34	4,31	18,54	1,05	20,99	29,68	0,47	9,28	2,47	17,72	0,15	38,88	9,05	3,09	30,96	8,17	6,66
18,88	12,13	6,19	4,62	54,31	26,41	11,23	19,32	1,20	13,79	1,24	2,20	2,08	24,63	35,29	2,82	4,62	6,45	12,87	12,34
0,79	4,68	12,95	35,60	0,53	0,75	13,55	5,00	3,37	3,38	18,43	2,86	6,89	2,45	0,93	0,44	0,34	12,01	8,25	24,21
2,51	46,01	0,63	2,98	0,74	0,30	2,57	3,25	14,65	1,19	0,00	1,88	3,47	6,41	4,91	2,83	14,02	4,57	5,42	2,42
0,84	2,61	16,67	9,28	3,08	7,82	9,20	46,33	0,28	4,32	2,34	5,49	11,78	1,57	3,01	3,08	10,70	26,27	2,63	1,05
5,21	0,81	20,96	6,58	21,75	6,16	9,15	15,15	10,83	0,47	19,59	3,62	1,73	7,29	5,69	28,34	39,88	0,58	2,97	14,67
3,96	13,35	10,46	12,59	1,67	0,85	0,17	12,57	9,64	19,08	7,48	2,57	6,92	11,32	16,69	4,36	0,04	0,02	3,60	5,75
4,95	2,61	3,53	4,09	6,30	3,49	5,69	2,42	8,29	6,71	4,67	2,40	2,60	12,19	0,48	4,83	3,41	31,14	0,26	9,40
2,46	5,72	0,30	11,73	25,59	41,01	13,27	3,10	2,60	1,90	5,09	24,05	17,24	1,32	11,43	2,89	14,45	34,79	6,08	12,77
0,14	7,05	4,82	17,94	5,38	2,97	2,68	1,90	5,34	46,28	13,44	6,89	14,38	18,11	23,26	19,88	1,53	0,75	1,88	0,36
26,89	7,35	3,43	19,44	29,11	9,27	2,94	8,54	1,19	7,67	5,52	0,23	7,66	0,97	0,90	4,37	0,01	1,53	18,79	7,91

Figura 10.- Variações de pesos em sacos de açúcar

A Tabela 10 mostra a Distribuição de Frequências que a representa estes dados.

TABELA 10.- VARIAÇÕES DE PESOS (EM GRAMAS) DE SACOS DE 5 KG DE AÇÚCAR DE CERTA MARCA - 2009

PESOS			SACOS
0,0	└───	6,1	212
6,1	└───	12,2	87
12,2	└───	18,3	42
18,3	└───	24,4	24
24,4	└───	30,5	13
30,5	└───	36,6	12
36,6	└───	42,7	5
42,7	└───	48,8	4
48,8	└───	54,9	1
			400

FONTE: Dados fictícios

Para compreendermos a tabela vejamos seus componentes.

3.3.2.1 - Elementos de uma Distribuição de Frequências

- Simbologia:

As simbologias encontradas nas diversas publicações variam muito quanto à forma e quanto ao seu significado. A Figura 11 a seguir ilustra as simbologias adotadas e seus significados.

SIMBOLOGIA				SIGNIFICADO
12,2	└───	18,3	⇒	O valor "12,2" pertence ao intervalo e o valor "18,3" não pertence
12,2	───┐	18,3	⇒	O valor "18,3" pertence ao intervalo e o valor "12,2" não pertence
12,2	───	18,3	⇒	Os valores "12,2" e "18,3" não pertencem ao intervalo
12,2	└───┐	18,3	⇒	Ambos os valores "12,2" e "18,3" pertencem ao intervalo

Figura 11.- Simbologias adotadas para as Distribuições de Frequências de Valores Agrupados em Classes

- **Em uma Distribuição de Frequências deve-se usar apenas um tipo de simbologia.**

- Amplitude Total (At)

É o comprimento total do intervalo de valores observados. A Amplitude Total é obtida pela diferença entre o maior valor e o menor valor da distribuição. No caso da Tabela 10 a Amplitude Total é de 54,9 gramas.

$$At = \text{MaiorValor} - \text{MenorValor} = 59.4 - 0 = 54.9$$

- Intervalo de Classe (c)

A Amplitude Total de uma Distribuição de Frequências é subdividida em intervalos menores, denominados Classes. Estas Classes, por sua vez, possuem um intervalo chamado Intervalo de Classe (c) que é determinado pela diferença entre seus valores extremos. Os valores extremos de uma classe são denominados *Limites de Classe*. Cada classe possui seu Limite Inferior (l_i) e seu Limite Superior (l_s). Assim podemos redefinir o Intervalo de Classe como sendo a diferença entre o Limite Superior e Limite Inferior de uma Classe.

$$c = l_s - l_i$$

➤ **Todos os Intervalos de Classes de uma Distribuição de Frequências devem ter o mesmo comprimento.**

- Número de Classes (k)

Após a obtenção dos dados um dos primeiros passos para a construção de uma Distribuição de Frequências é a determinação do número de classes que ela deva conter. Existem várias fórmulas e critérios diferentes para se determinar o número de classes de uma Distribuição de Frequências. Infelizmente não existe uma fórmula ou um critério que indique com exatidão o número de classes de uma Distribuição de Frequências. A seguir apresentamos duas maneiras de se determinar o número de classes próximo ao ideal. Devemos lembrar que os resultados obtidos por estas duas fórmulas apenas indicam um valor próximo do ideal, podendo, assim, ser necessários ajustes para se construir uma Distribuição de Frequências.

- Critério da Raiz

$$k = \sqrt{n}$$

onde: $k \equiv$ número de classes

$n \equiv$ número total de valores observados

- Fórmula de Sturges

$$k = 1 + 3,32 \log(n)$$

Exemplo 3.1 - Construa uma Distribuição de Frequências que represente a lista de Dados Brutos a seguir.

55,1	75,9	62,5	64,5	57,8	74,0	118,8	82,5	80,2	96,5	107,7	75,3	85,0	80,1	75,6	83,0
61,9	101,8	55,0	84,7	34,7	48,0	94,9	98,9	83,2	91,1	69,5	67,1	60,9	84,3	89,7	72,7
64,8	78,7	76,2	56,5	35,8	106,4	99,3	83,6	60,1	94,7	73,4	48,8	58,1	35,1	93,5	78,6
38,6	56,5	72,9	58,0	67,9	67,6	92,9	67,0	100,9	90,5	83,7	94,1	62,5	71,3	77,6	85,6
53,1	68,6	72,7	64,5	99,5	75,6	73,7	71,2	64,6	72,2	78,4	75,7	61,7	67,7	82,2	84,0
66,2	65,3	44,3	59,3	87,5	50,1	81,6	70,9	45,1	96,7	99,1	39,6	92,0	53,9	83,2	69,7
71,2	83,9	92,6	41,9	57,6	69,2	72,3	66,3	102,0	102,5	70,2	90,6	98,0	81,6	68,5	87,0
67,8	62,1	68,8	62,3	79,5	97,6	78,9	59,7	74,8	88,3	60,1	83,7	63,2	66,6	60,6	120,4
55,0	88,1	50,3	93,5	76,7	77,1	68,6	71,4	64,1	85,4	67,3	75,2	119,1	75,4	85,8	73,5
42,4	72,8	74,3	50,4	66,9	76,5	88,5	100,0	56,1	53,9	71,1	99,7	33,0	68,4	66,6	96,2
65,1	95,4	51,6	97,5	74,8	81,4	62,4	100,3	89,4	89,0	74,2	66,8	57,2	80,6	51,0	83,2
63,0	47,1	61,0	62,5	69,2	66,7	106,8	58,9	79,3	70,5	63,1	75,3	49,7	59,4	79,3	83,1
86,0	94,2	96,6	79,7	72,0	65,8	36,7	59,8	68,6	80,5	93,7	52,6	82,3	55,2	66,6	91,7
91,7	66,4	93,3	69,1	64,2	102,0	60,9	77,9	67,1	92,9	69,4	56,7	31,6	66,9	64,2	80,4
71,2	50,9	59,2	66,8	79,4	66,2	64,2	60,1	64,1	67,8	77,1	64,8	64,0	78,2	67,3	85,5
29,0	68,1	58,6	75,4	65,8	84,3	61,7	101,7	71,2	87,9	67,4	88,6	57,9	72,6	64,4	99,9
76,1	69,2	63,2	88,1	83,7	71,8	59,3	68,2	99,0	106,6	78,8	52,8	75,5	66,6	77,2	82,7
75,1	105,1	48,9	87,1	97,5	82,0	58,8	57,6	53,4	78,6	65,0	63,1	64,2	62,9	95,0	85,5
76,8	49,1	59,7	75,1	52,1	38,8	76,4	92,4	62,7	66,8	91,9	91,7	84,0	73,9	57,1	86,1
64,3	42,8	87,6	96,7	107,9	81,0	67,0	56,9	84,4	64,4	75,1	94,9	64,4	71,1	54,9	97,0

Figura 12.- Lista de Dados Brutos exemplo para construção de uma Distribuição de Frequências

Solução

O primeiro passo para se construir uma Distribuição de Frequências é determinar quantos valores foram observados (n). No caso da lista anterior temos 320 valores ($n = 320$).

A seguir determinamos qual o menor e qual o maior valor observado (Menor = 29,0 e Maior = 120,4)

A Amplitude Total será:

$$At = 120,4 - 29 = 91,4$$

Para determinar o Número de Classes usaremos a Regra de Sturges, assim

$$k = 1 + 3,32 \times \log(n)$$

$$k = 1 + 3,32 \times \log(320)$$

$$k = 1 + 3,32 \times 2,50515$$

$$k = 9,317098$$

O Número de Classes deve ser arredondado para um número inteiro, uma vez que não é possível ter Classes fracionárias, logo usaremos $k = 9$.

A futura Distribuição de Frequências terá entorno de 9 Classes. Para determinar a amplitude de cada uma destas Classes – Intervalo de Classe (c) – devemos dividir a Amplitude Total (At) pelo Número de Classes (k), isto é

$$c = \frac{At}{k} = \frac{91,4}{9} = 10,15567$$

Para ficar mais coerente e simples de trabalhar devemos arredondar o Intervalo de Classe (c) para o mesmo número de casas decimais dos valores observados, que no caso é de uma casa decimal, assim adotaremos $c = 10,2$.

Antes de começarmos a construção da Distribuição devemos verificar se os resultados obtidos, até então, estão de acordo com os valores observados. Para tal faremos uso da seguinte inequação:

$$k \times c + \text{Valor Inicial} > \text{Maior Valor}$$

Onde “Valor Inicial” é o valor com o qual iniciaremos a Distribuição e deve ser um valor não maior que o Menor Valor observado na lista de Dados Brutos e “Maior Valor” é o maior valor observado na lista de Dados Brutos.

Se a inequação for satisfeita podemos iniciar a construção da distribuição, caso contrário, devemos fazer ajustes nos valores do Intervalo de Classe (c) e/ou no Número de Classes (k) para que a futura distribuição contemple todos os valores observados.

Adotemos como Valor Inicial o valor 29. Procedendo a verificação, temos:

$$k \times c + Valor\ Inicial > Maior\ Valor$$












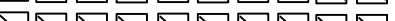

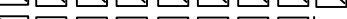

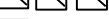


$$9 \times 10,2 + 29 > 120,4$$

$$120,8 > 120,4$$

Como a inequação é verdadeira construiremos a distribuição iniciando suas Classes com o valor 29. A partir deste valor iremos somar o Intervalo de Classe ($c=10,2$) sucessivamente até obtermos o Número de Classes ($k=9$) calculado.

CLASSES			f_i
29,0	┃————	39,2	
39,2	┃————	49,4	
49,4	┃————	59,6	
59,6	┃————	69,8	
69,8	┃————	80,0	
80,0	┃————	90,2	
90,2	┃————	100,4	
100,4	┃————	110,6	
110,6	┃————	120,8	

O próximo passo é marcar, um a um, em qual Classe cada valor da lista de Dados Brutos se encontra.

CLASSES			f_i
29,0		39,2	
39,2		49,4	
49,4		59,6	
59,6		69,8	
69,8		80,0	
80,0		90,2	
90,2		100,4	
100,4		110,6	
110,6		120,8	

A Distribuição de Frequências resultante será:

TABELA 11.- DISTRIBUIÇÃO DE FREQUÊNCIAS QUE REPRESENTA OS VALORES DA LISTA DE DADOS BRUTOS DA FIGURA 12

CLASSES			f_i
29,0	└───	39,2	9
39,2	└───	49,4	11
49,4	└───	59,6	39
59,6	└───	69,8	88
69,8	└───	80,0	65
80,0	└───	90,2	51
90,2	└───	100,4	42
100,4	└───	110,6	12
110,6	└───	120,8	3
			320

Observações: Ao se construir uma Distribuição de Frequências podemos encontrar várias soluções diferentes, dependendo do Valor Inicial escolhido, do Número de Classes Adotado ou do Intervalo de Classe utilizado. Porém devemos sempre ter em mente algumas regras.

- O Valor Inicial adotado nunca pode ser maior que o menor valor observado na lista de Dados Brutos;
- O Número de Classes (k) pode variar mas deve estar sempre entorno do valor calculado ou pela Regra de Sturges ou pelo Critério da Raiz;
- O Intervalo de Classe (c) deve ser um valor próximo ao valor calculado e nunca arredondado para um valor com número de casas decimais menor que o número de casas decimais dos valores da lista de Dados Brutos;
- Todos os valor observados na lista de Dados Brutos devem pertencer à algum intervalo da Distribuição de Frequências;
- Deve-se evitar ao máximo intervalos com frequência nula. Neste caso devemos alterar os valor do Número de Classes (k) e/ou o Intervalo de Classe (c) de maneira que a distribuição não tenha intervalos com frequência nula.

• Ponto Médio (x_i)

Durante a etapa de Análise dos Dados o estatístico realiza diversos cálculos necessários para o estudo do fenômeno em observação. No caso de Distribuições de Frequências de Valores Individuais os valores estão explícitos, porém quando o estatístico está trabalhando com Distribuições de Frequências de Valores Agrupados em Classes os valores não estão explícitos. Existem intervalos com infinitos pontos e o estatístico precisa de um valor que represente toda a classe durante os cálculos. Este valor é denominado Ponto Médio e é simbolizado por x_i . O Ponto Médio de uma classe pode ser determinado como sendo o ponto médio entre os limites da classe, isto é

$$x_i = \frac{l_i + l_s}{2}.$$

A escolha do Ponto Médio com sendo o valor médio entre os limites da classe não é arbitrária. Esta escolha está baseada num teorema conhecido com Teorema do limite Central que será enunciado mais adiante.

3.3.3 - TIPOS DE FREQUÊNCIAS

3.3.3.1 - Frequência Relativa Simples

A *Frequência Relativa Simples* (fr_i) de um valor (ou de uma classe) é a proporção deste valor em relação ao número total de valores observados. As Frequências Relativas Simples são facilmente obtidas dividindo-se cada Frequência Absoluta Simples pelo número total de valores da observados ($n = \sum f_i$).

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

As Frequências Relativas Simples podem ser expressas em porcentagem, bastando para tal multiplicando as proporções por 100%.

Considerando as frequências da Tabela 11 anterior, temos para a 1ª classe uma Frequência Relativa Simples igual a $9/320 = 0,028$ ou 2,8%, para a 2ª classe a Frequência Relativa Simples será $11/320 = 0,034$ ou 3,4%. Repetindo o processo até a última classe obtemos todas as Frequências Relativas Simples.

TABELA 12.- DISTRIBUIÇÃO DE FREQUÊNCIAS QUE REPRESENTA OS VALORES DA LISTA DE DADOS BRUTOS DA FIGURA 12

CLASSES			f_i	fr_i
29,0	└───	39,2	9	0,028
39,2	└───	49,4	11	0,034
49,4	└───	59,6	39	0,122
59,6	└───	69,8	88	0,275
69,8	└───	80,0	65	0,203
80,0	└───	90,2	51	0,159
90,2	└───	100,4	42	0,131
100,4	└───	110,6	12	0,038
110,6	└───	120,8	3	0,009
			320	0,999

Observando a Tabela 12 nota-se que a soma das Frequências relativas Simples não é igual à unidade. Este fato ocorre com muita frequência e contradiz uma das propriedades da proporção (“a soma de todas as proporções é igual a 1”) ilustrada na Figura 13 a seguir.

Nestes casos devemos realizar certas “correções” para que a soma das Frequências relativas Simples seja igual à unidade (ou 100% no caso em que as Frequências Relativas Simples estejam sendo explicitadas com porcentagens).

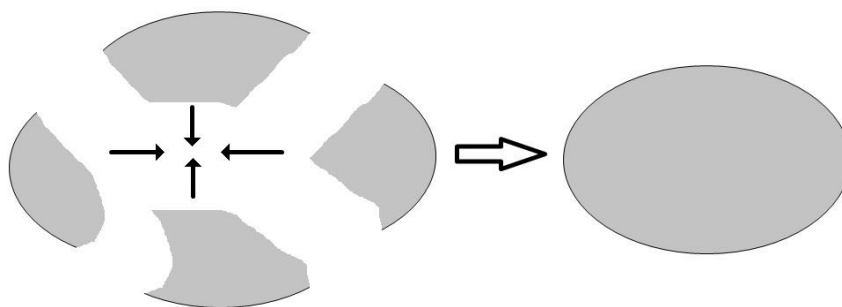


Figura 13.- Propriedade da União de proporções

Os procedimentos a seguir constituem uma maneira prática para se realizar estas correções.

- **Procedimento para se corrigir as Frequências relativas Simples.**

1. Determinar qual o número de casas decimais deve ser utilizado para representar as Frequências Relativas Simples;
2. Em um rascunho, calcular as Frequências Relativas Simples utilizando pelo menos duas casas decimais a mais do que estipulado no 1º passo, escrevendo os dígitos das casas destas casas decimais afastadas das demais para facilitar sua inspeção;
3. Somar as Frequências Relativas Simples considerando-se apenas as casas decimais estipuladas no 1º passo sem considerar nenhum arredondamento;
4. Verificar quantas unidades são necessárias para que a soma das Frequências Relativas Simples atinja a unidade (chamemos este valor de “a”);
5. Inspeccionando os dígitos escritos mais afastados e, do maior para o menor, marcar os números até atingirmos a quantidade “a” (caso ocorra um empate, a quantidade “a” seja ultrapassada e as respectivas Frequências Absolutas Simples sejam iguais devemos desconsiderar os números que estão empatando e marcar o próximo número considerando a sequência do maior para o menor. Caso haja um empate, mas as Frequências Absolutas Simples correspondentes sejam diferentes, marcamos normalmente, a partir do primeiro, até atingirmos a quantidade “a”);
6. Arredondar para o próximo valor acima apenas as Frequências Relativas Simples marcadas;
7. Somar todas as Frequências Relativas Simples e verificar se o resultado é igual à unidade;
8. Se a Soma for igual à unidade devemos transcrever as Frequências Relativas Simples corrigidas para a distribuição.

A Figura 14 a seguir ilustra este procedimento.

- Para o exemplo foi adotado 3 casas decimais para as Frequências Relativas Simples.
- De acordo com a soma das Frequências Relativas Simples sem arredondamento deve-se arredondar apenas 3 valores.

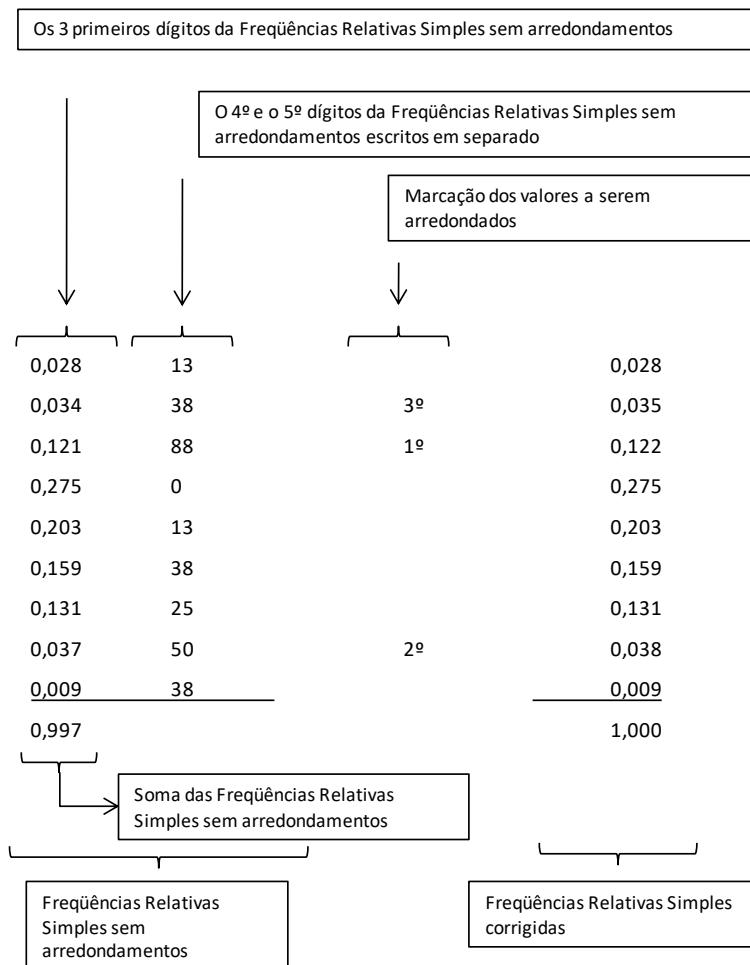


Figura 14.- Procedimento pra correção das Frequências relativas Simples

A Distribuição de Frequências após a correção fica

TABELA 13.- DISTRIBUIÇÃO DE FREQUÊNCIAS QUE REPRESENTA OS VALORES DA LISTA DE DADOS BRUTOS DA FIGURA 12

CLASSES	f_i	fr_i
29,0 — 39,2	9	0,028
39,2 — 49,4	11	0,035
49,4 — 59,6	39	0,122
59,6 — 69,8	88	0,275
69,8 — 80,0	65	0,203
80,0 — 90,2	51	0,159
90,2 — 100,4	42	0,131
100,4 — 110,6	12	0,038
110,6 — 120,8	3	0,009
	320	1,000

3.3.3.2 - Frequência Absoluta Acumulada “Abaixo de”

Em determinados momentos se faz necessário conhecer o número de valores são menores ou iguais a determinado valor observado. Esta informação pode ser conseguida através das *Frequências Absolutas Acumuladas “Abaixo de”* (F_i). Para tal devemos somar as Frequências Absolutas Simples dos valores (ou classes) anteriores ao valor (ou classe) desejado com a Frequência Absoluta Simples do valor (ou da classe) desejado. Podemos expressa a Frequência Absoluta Simples de um determinado valor (ou classe) como

$$F_i = f_1 + f_2 + f_3 + \dots + f_i$$

Assim:

$$F_1 = f_1$$

$$F_2 = f_1 + f_2$$

$$F_3 = f_1 + f_2 + f_3$$

$$F_4 = f_1 + f_2 + f_3 + f_4$$

Para a distribuição da Tabela 13 temos:

TABELA 14.- DISTRIBUIÇÃO DE FREQUÊNCIAS QUE REPRESENTA OS VALORES DA LISTA DE DADOS BRUTOS DA FIGURA 12

CLASSES			f_i	fr_i	F_i
29,0	└───	39,2	9	0,028	9
39,2	└───	49,4	11	0,035	20
49,4	└───	59,6	39	0,122	59
59,6	└───	69,8	88	0,275	147
69,8	└───	80,0	65	0,203	212
80,0	└───	90,2	51	0,159	263
90,2	└───	100,4	42	0,131	305
100,4	└───	110,6	12	0,038	317
110,6	└───	120,8	3	0,009	320
			320	1,000	

Da tabela podemos afirmar que 147 valores são menores ou iguais a 69,8 ou ainda que 305 valores estão “Abaixo de” 100,4.

3.3.3.3 - Frequência Relativa Acumulada “Abaixo de”

De maneira semelhante, a informação da proporção ou da porcentagem de valores inferiores a um determinado valor pode ser requerida. Este dado pode ser obtido pela *Frequência Relativa Acumulada “Abaixo de”* (Fr_i). Assim como as Frequências Absolutas Acumuladas “Abaixo de” as Frequências Relativas Acumuladas “Abaixo de” são obtidas pela soma sucessiva de frequências simples, porém da soma das Frequências Relativas Simples anteriores ao valor desejado com a Frequência Relativa Simples do valor desejado.

$$Fr_i = fr_1 + fr_2 + fr_3 + \dots + fr_i$$

Assim:

$$Fr_1 = fr_1$$

$$Fr_2 = fr_1 + fr_2$$

$$Fr_3 = fr_1 + fr_2 + fr_3$$

$$Fr_4 = fr_1 + fr_2 + fr_3 + fr_4$$

TABELA 15.- DISTRIBUIÇÃO DE FREQUÊNCIAS QUE REPRESENTA OS VALORES DA LISTA DE DADOS BRUTOS DA FIGURA 12

CLASSES			f_i	fr_i	F_i	Fr_i
29,0	└───	39,2	9	0,028	9	0,028
39,2	└───	49,4	11	0,035	20	0,063
49,4	└───	59,6	39	0,122	59	0,185
59,6	└───	69,8	88	0,275	147	0,460
69,8	└───	80,0	65	0,203	212	0,663
80,0	└───	90,2	51	0,159	263	0,822
90,2	└───	100,4	42	0,131	305	0,953
100,4	└───	110,6	12	0,038	317	0,991
110,6	└───	120,8	3	0,009	320	1,000
			320	1,000		

Da distribuição temos que 46% dos valores são menores ou iguais a 69,8 ou 95,3% dos valores estão “abaixo de” 100,4.

3.3.3.4 - Frequência Absoluta Acumulada “Acima de”

Não menos importante é o conhecimento de quantos dos valores observados são maiores ou iguais a um determinado valor. A *Frequência Absoluta Acumulada “Acima de”* (Fr_i) proporciona esta informação. A Frequência Absoluta Acumulada “Acima de” é determinada pela soma da Frequência Absoluta Simples deste valor (ou desta classe) com as Frequências Absolutas Simples é a soma da Frequência Absoluta Simples dos valores (ou classes) posteriores.

$$F_i = f_i + f_{i+1} + f_{i+2} + \dots + f_k$$

Assim

$$F_1 = f_1 + f_2 + f_3 + f_4 + \dots + f_k$$

$$F_2 = f_2 + f_3 + f_4 + \dots + f_k$$

$$F_3 = f_3 + f_4 + \dots + f_k$$

⋮

$$F_{k-1} = f_{k-1} + f_k$$

$$F_k = f_k$$

TABELA 16.- DISTRIBUIÇÃO DE FREQUÊNCIAS QUE REPRESENTA OS VALORES DA LISTA DE DADOS BRUTOS DA FIGURA 12

CLASSES			f_i	fr_i	F_i	Fr_i	F_i
29,0	└───	39,2	9	0,028	9	0,028	320
39,2	└───	49,4	11	0,035	20	0,063	311
49,4	└───	59,6	39	0,122	59	0,185	300
59,6	└───	69,8	88	0,275	147	0,460	261
69,8	└───	80,0	65	0,203	212	0,663	173
80,0	└───	90,2	51	0,159	263	0,822	108
90,2	└───	100,4	42	0,131	305	0,953	57
100,4	└───	110,6	12	0,038	317	0,991	15
110,6	└───	120,8	3	0,009	320	1,000	3
			320	1,000			

Inspecionando a distribuição podemos afirmar que dos 320 valores observados 173 valores são maiores ou iguais a 69,8, ou ainda, que apenas 57 valores estão “Acima de” 90,2.

3.3.3.5 - Frequência Relativa Acumulada “Acima de”

A *Frequência Relativa Acumulada “Acima de”* (Fr_i) nos fornece a informação de qual proporção dos valores observados é maior ou igual a determinado valor. Para calcular as Frequências Relativas Acumuladas “Acima de” para um determinado valor basta somarmos a Frequência Relativa Simples deste valor com as Frequências Relativas Simples posteriores.

$$Fr_i = fr_i + fr_{i+1} + fr_{i+2} + \cdots + fr_k$$

Assim

$$Fr_1 = fr_1 + fr_2 + fr_3 + fr_4 + \cdots + fr_k$$

$$Fr_2 = fr_2 + fr_3 + fr_4 + \cdots + fr_k$$

$$Fr_3 = fr_3 + fr_4 + \cdots + fr_k$$

⋮

$$Fr_{k-1} = fr_{k-1} + fr_k$$

$$Fr_k = fr_k$$

TABELA 17.- DISTRIBUIÇÃO DE FREQUÊNCIAS QUE REPRESENTA OS VALORES DA LISTA DE DADOS BRUTOS DA FIGURA 12

CLASSES			f_i	fr_i	F_i	Fr_i	F_i	Fr_i
29,0	└───	39,2	9	0,028	9	0,028	320	1,000
39,2	└───	49,4	11	0,035	20	0,063	311	0,972
49,4	└───	59,6	39	0,122	59	0,185	300	0,937
59,6	└───	69,8	88	0,275	147	0,460	261	0,815
69,8	└───	80,0	65	0,203	212	0,663	173	0,54
80,0	└───	90,2	51	0,159	263	0,822	108	0,337
90,2	└───	100,4	42	0,131	305	0,953	57	0,178
100,4	└───	110,6	12	0,038	317	0,991	15	0,047
110,6	└───	120,8	3	0,009	320	1,000	3	0,009
			320	1,000				

3.4 - EXERCÍCIOS PROPOSTOS

3.4.1) Construa uma Distribuição de Frequências para cada conjunto de dados a seguir:

a)

```

65  62  25  26  48  74  46  49  43  42
39  60  49  68  54  30  27  16  75  15
47  29  86  24  81  41  69  22  52  76
53  32  64  29  42  79  51  51  78  53
77  40  75  72  84  42  36  28  31  29

```

b)

```

65,5  82,5  86,4  64,6  92,0  71,3  54,2  74,1  49,0
86,9  84,4  82,6  78,8  75,9  80,1  43,4  55,1  48,4
57,6  72,0  57,2  87,4  64,8  67,1  74,8  55,6  55,6
57,2  82,8  84,6  75,1  76,3  69,3  49,9  96,0  84,4
67,6  60,1  82,1  93,7  67,1  55,7  96,0  51,4  51,8
89,8  39,4  84,1  95,2  61,1  95,1  95,5  49,6  95,3
54,1  83,0  49,8  76,0  67,8  62,4  57,9  81,4  66,3

```

c)

```

7  9  11  12  16  8  7  11  7  8  7
9  9  8  9  8  14  9  10  11  7  10
5  4  9  8  12  12  12  9  10  11  9
6  9  8  11  11  7  7  11  9  13  12
7  9  11  9  11  11  4  11  9  7  5
9  8  6  10  9  9  10  9  6  9  12
11 4  10  8  7  5  9

```

d)

25,5485	37,9616	36,3541	60,2476	39,8068	50,2556	46,0896
43,9847	33,7958	31,9999	46,7547	40,5977	36,8962	35,1224
38,1497	34,2331	55,8998	22,0038	10,2852	52,7873	47,9424
50,1225	34,4096	44,0583	36,9245	20,0920	42,7032	38,5012
41,6421	29,5321	36,6640	40,6729	40,7107	29,6552	22,0404
55,6242	40,5960	41,8157	24,7355	39,6182	30,9444	46,5001
36,0353	30,9810	45,5998	38,0279	32,5431	44,1257	15,3577
43,9971	46,4198	16,5430	29,4974	59,1437	23,2594	32,4509
47,7715	39,4266	42,9173	43,7734	43,7462	41,5556	25,1495
31,8963	36,7032	32,3372	45,5078	22,4181	6,7061	42,4060
37,6480	16,4651	18,1479	32,9063	29,6693	43,9527	24,0648
24,1057	38,6315	33,9553	52,1756	50,0591	60,4607	46,4139
57,5643	30,1195	11,6988	37,7043	52,7087	26,4989	35,0439
39,3847	18,1501	57,3476	24,2449	33,2255	0,6796	35,3955
29,4338	50,2101	20,8380	39,2780	53,2620	23,2147	33,4115
46,0868	30,7035	38,1588	37,5520	37,6972	16,4446	36,6152
26,1205	21,8375	35,4086	31,1469	29,5947	34,0800	16,7018
34,5478	30,0972	25,0769	65,5039	27,6852	49,7394	11,8467
51,0839	20,3689	63,5750	21,4368	49,2380	42,2288	32,1736

e)

34	58	66	29	37	48	28	51	36
32	54	44	46	47	36	40	50	35
65	62	55	41	32	50	66	37	45
60	51	50	33	34	28	64	27	45
58	69	68	29	40	51	28	36	31
29	31	28	37	55	58	28	49	51
40	39	57	71	36	35	63	31	44
51	55	64	60	65	45	36	42	56
48	70	59	61	53	61	49	56	43
27	32	68	36	70	56	39	46	52
52	69	53	67	69	55	46	36	40
30	45	42	28	54	36	34	55	29
34	32	47	70	49	31	55	71	32
56	40	39	33	31	50	38	56	44
31	37	64	59	51	54	51	64	60
59	63	66	34	31	55	59	49	58
35	54	38	37	57	64	40	44	30

f) Para a variável IDADE do Conjunto de Dados 2 do Apêndice B;

g) Para a variável PAPEL do Conjunto de Dados 16 do Apêndice B;

h) Para a variável COMIDA do Conjunto de Dados 16 do Apêndice B.

3.4.2) Determine TODAS as frequências das distribuições a seguir.

Distribuição A: (3 casas decimais)

CLASSES	f_i	fr_i	F_i	Fr_i	F_i	Fr_i
4 — 13	9					
13 — 22	5					
22 — 31	7					
31 — 40	4					
40 — 49	19					
49 — 58	11					
58 — 67	8					
67 — 76	2					
76 — 85	6					

Distribuição B: (3 casas decimais)

CLASSES	f_i	fr_i	F_i	Fr_i	F_i	Fr_i
2 — 13	6					
13 — 24	5					
24 — 35	12					
35 — 46	19					
46 — 57	11					
57 — 68	17					
68 — 79	3					
79 — 90	6					
90 — 101	8					

Distribuição C: (3 casas decimais)

CLASSES	f_i	fr_i	F_i	Fr_i	F_i	Fr_i
6 — 19	2					
19 — 32	7					
32 — 45	4					
45 — 58	9					
58 — 71	3					
71 — 84	9					
84 — 97	5					
97 — 110	1					
110 — 123	8					

Distribuição D: (4 casas decimais)

CLASSES			f_i	fr_i	F_i	Fr_i	F_i	Fr_i
2	└─	10	27					
10	└─	18	8					
18	└─	26	4					
26	└─	34	5					
34	└─	42	19					
42	└─	50	13					
50	└─	58	28					
58	└─	66	3					
66	└─	74	6					

Distribuição E: (3 casas decimais)

CLASSES			f_i	fr_i	F_i	Fr_i	F_i	Fr_i
2,0	└─	3,3	7					
3,3	└─	4,6	11					
4,6	└─	5,9	21					
5,9	└─	7,2	29					
7,2	└─	8,5	44					
8,5	└─	9,8	33					
9,8	└─	11,1	26					
11,1	└─	12,4	18					
12,4	└─	13,7	4					

Distribuição F: (2 casas decimais)

CLASSES			f_i	fr_i	F_i	Fr_i	F_i	Fr_i
1,00	└─	3,05	26					
3,05	└─	5,10	35					
5,10	└─	7,15	39					
7,15	└─	9,20	21					
9,20	└─	11,25	27					
11,25	└─	13,30	14					
13,30	└─	15,35	29					
15,35	└─	17,40	33					
17,40	└─	19,45	27					

Distribuição G: (3 casas decimais)

CLASSES	f_i	fr_i	F_i	Fr_i	F_i	Fr_i
0,20 — 0,27	5					
0,27 — 0,34	19					
0,34 — 0,41	38					
0,41 — 0,48	61					
0,48 — 0,55	44					
0,55 — 0,62	49					
0,62 — 0,69	3					

Distribuição H: (3 casas decimais)

CLASSES	f_i	F_i	F_i	fr_i	Fr_i	Fr_i
2,8 — 8,1	1					
8,1 — 13,4	11					
13,4 — 18,7	21					
18,7 — 24,0	31					
24,0 — 29,3	28					
29,3 — 34,6	18					
34,6 — 39,9	13					
39,9 — 45,2	6					
45,2 — 50,5	8					

Distribuição I: (3 casas decimais)

CLASSES	f_i	fr_i	F_i	Fr_i	F_i	Fr_i
6 — 14	2					
14 — 22	29					
22 — 30	13					
30 — 38	21					
38 — 46	26					
46 — 54	18					
54 — 62	9					
62 — 70	1					

Distribuição J: (3 casas decimais)

CLASSES	f_i	fr_i	F_i	Fr_i	F_i	Fr_i
3 — 22	3					
22 — 41	15					
41 — 60	23					
60 — 79	24					
79 — 98	19					
98 — 117	13					
117 — 136	8					
136 — 155	2					
155 — 174	6					

Distribuição K: (3 casas decimais)

CLASSES	f_i	fr_i	F_i	Fr_i	F_i	Fr_i
0,3 — 5,0	16					
5,0 — 9,7	5					
9,7 — 14,4	18					
14,4 — 19,1	19					
19,1 — 23,8	34					
23,8 — 28,5	27					
28,5 — 33,2	4					
33,2 — 37,9	8					
37,9 — 42,6	3					

Distribuição L: (4 casas decimais)

CLASSES	f_i	fr_i	F_i	Fr_i	F_i	Fr_i
13,3 — 20,2	6					
20,2 — 27,1	15					
27,1 — 34,0	13					
34,0 — 40,9	21					
40,9 — 47,8	17					
47,8 — 54,7	14					
54,7 — 61,6	9					
61,6 — 68,5	4					
68,5 — 75,4	8					

Distribuição M: (3 casas decimais)

CLASSES	f_i	fr_i	F_i	Fr_i	F_i	Fr_i
4 — 11	2					
11 — 18	12					
18 — 25	22					
25 — 32	62					
32 — 39	39					
39 — 46	43					
46 — 53	18					
53 — 60	11					
60 — 67	6					

Distribuição N: (3 casas decimais)

CLASSES	f_i	fr_i	F_i	Fr_i	F_i	Fr_i
1,73 — 3,78	1					
3,78 — 5,83	7					
5,83 — 7,88	9					
7,88 — 9,93	23					
9,93 — 11,98	26					
11,98 — 14,03	35					
14,03 — 16,08	13					
16,08 — 18,13	4					
18,13 — 20,18	15					

Distribuição O: (4 casas decimais)

CLASSES	f_i	fr_i	F_i	Fr_i	F_i	Fr_i
11,8 — 39,5	12					
39,5 — 67,2	15					
67,2 — 94,9	23					
94,9 — 122,6	21					
122,6 — 150,3	7					
150,3 — 178,0	4					
178,0 — 205,7	8					
205,7 — 233,4	14					
233,4 — 261,1	9					