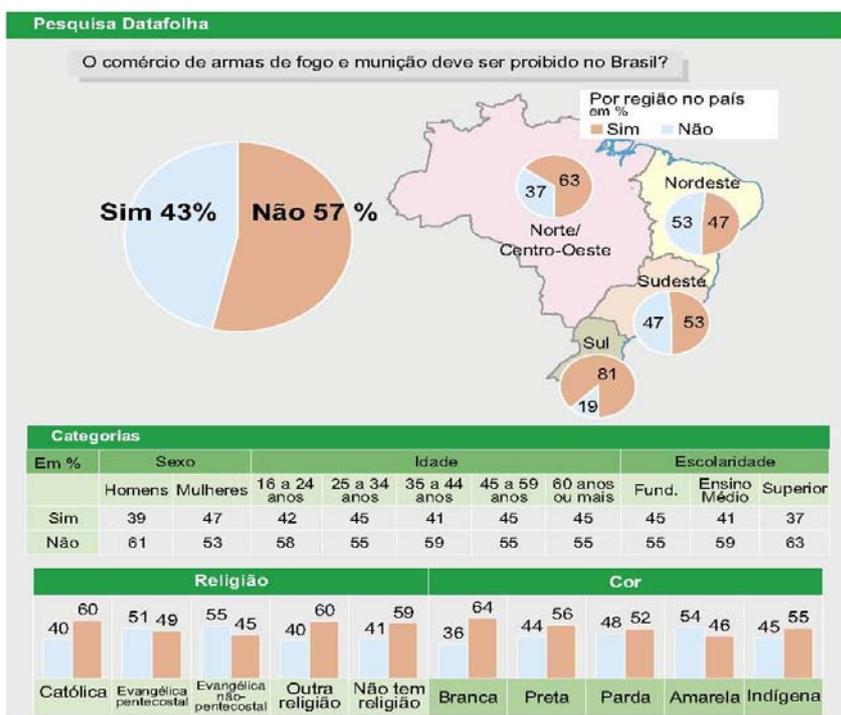


Probabilidade e Estatística

Estatística Aplicada no Excel

Matemática

Bertolo, L. A.



Versão BETA

Abril 2010

Capítulo 2 – Medidas Estatísticas

A distribuição de freqüências permite-nos descrever, de modo geral, os grupos de valores assumidos por uma variável. Dessa forma, podemos localizar a maior concentração de valores de uma dada distribuição. Esta concentração se encontra onde? No início, no meio, ou no final dos valores?

Quando confrontamos distribuições e queremos destacar as tendências de cada uma, isoladamente, necessitamos de conceitos que expressem através de números estas tendências. Esses conceitos são denominados **elementos típicos** da distribuição (ou estatísticas) e são:

- Medidas de Posição (locação ou tendência central)
- Medidas de Dispersão (variabilidade)
- Medidas de Assimetria
- Medidas de Curtose

2.1 – Medidas de Posição (ou tendência central)

Mostram o valor representativo em torno do qual os dados tendem a agrupar-se com maior ou menor freqüência.

A medida de tendência central é um número que está representando todo o conjunto de dados; nas pesquisas tal número pode ser encontrado a partir das medidas:

- média aritmética;
- da moda ou
- da mediana

O uso de cada uma delas é mais conveniente de acordo com o nível de mensuração, o aspecto ou forma da distribuição de dados e o objetivo da pesquisa.

Outras medidas de posição são as **separatrizes**, que englobam:

- a própria mediana;
- os quartis;
- os percentis.

2.1.1 – Média Aritmética Simples (\bar{x})

É a medida de centralidade mais comum, porém deve ser usada em dados representados por intervalos, pois não haveria sentido utilizá-la em uma distribuição em que a variável fosse, por exemplo, time de futebol ou sexo. A média representa, ainda, o ponto de distribuição no qual se equilibram as discrepâncias (diferenças) positivas e negativas de cada dado, ou seja, as discrepâncias positivas somadas se anulam com as negativas somadas.

2.1.1.1 – Dados não-agrupados

Definida da seguinte forma:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

é a soma de todos os números dividida pelo número de parcelas. É uma das medidas de tendência central de maior emprego.

Usada em dados não agrupados.

EX: 4 15 20 20 24 27 30

$$\bar{x} = \frac{4 + 15 + 20 + 20 + 24 + 27 + 30}{7} = 20$$

Observe que: $(20-4) + (20-15) + (20-20) + (20-24) + (20-27) + (20-30) = 0$

2.1.1.2 – Dados agrupados

Vamos dividi-los em duas categorias: sem intervalo de classe e com intervalo de classe.

2.1.1.2.1 – Sem intervalos de classe

Seja a distribuição relativa a 34 famílias de quatro filhos, tomando para a variável o número de filhos do sexo masculino:

Nº de filhos	n_i
0	2
1	6
2	10
3	12
4	4
$\Sigma = 34$	

Neste caso, como as freqüências são números indicadores da intensidade de cada valor da variável, elas funcionam como fatores de ponderação, o que nos leva à **Média Aritmética Ponderada**.

Média Aritmética Ponderada \bar{x}

É um tipo de média aritmética de vários valores com pesos diferentes, dada por:

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \dots + f_nx_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}$$

f_i = freqüência do valor x_i na amostra.

Um modo rápido de obtermos a média ponderada é abrir, na tabela, uma coluna correspondente aos produtos $x_i f_i$:

Nº de filhos	n_i	$x_i f_i$
0	2	0
1	6	6
2	10	20
3	12	36
4	4	16
$\Sigma = 34$		$\Sigma = 78$

$$\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i} \Rightarrow \bar{x} = \frac{78}{34} = 2,29$$

2.1.1.2.1 – Com intervalos de classe

Aqui, convencionamos que *todos os valores incluídos em um determinado intervalo de classe coincidem com o seu ponto médio*, e determinamos a média aritmética ponderada por meio da fórmula:

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

Onde x_i é o ponto médio da classe.

Consideremos a distribuição:

i	X	f_i	x_i	$x_i f_i$
1	150 — 154	4	152	608
2	154 — 158	9	156	1.404
3	158 — 162	11	160	1.760
4	162 — 166	8	164	1.312
5	166 — 170	5	168	840
6	170 — 174	3	172	516
		$\Sigma = 40$		$\Sigma = 6.440$

$$\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i} \Rightarrow \bar{x} = \frac{6.440}{40} = 161$$

Ver como fazer nas calculadoras científicas, financeira e no Excel (MÉDIA, MÉDIAA, MÉDIASE, MÉDIASES)

A média aritmética simples pode ser vista como a média ponderada com todos os pesos iguais. Para efeito de nomenclatura sempre trataremos a média aritmética simples ou ponderada simplesmente por média representada por (\bar{x}).

2.1.2 – Média Geométrica (\bar{x}_G) e Média Harmônica (\bar{x}_H)

A Média Geométrica é definida como a raiz de ordem n do produto desses números.

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots x_n} = \sqrt[n]{\prod_{i=1}^n x_i}$$

A Média Harmônica é definida assim:

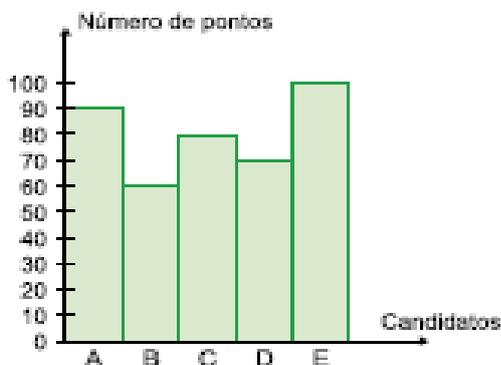
$$\bar{x}_H = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}} = \frac{1}{\sum \frac{1}{x}}$$

O Excel apresenta três funções para calcular a média geométrica, harmônica e interna:
 MÉDIA.GEOMÉTRICA,
 MÉDIA.HARMÔNICA,
 MÉDIA.INTERNA

Exercícios de Aplicação

01. Temos um gráfico que nos mostra o desempenho dos 5 melhores classificados em um determinado concurso, no qual a pontuação varia de zero a cem pontos.

- Qual é a soma dos pontos dos candidatos A, B, C, D e E?
- Determine a média aritmética dos pontos dos candidatos discriminados no gráfico.
- Mostre qual o candidato que fez mais e o que fez menos pontos.



Resposta: a. $90 + 60 + 80 + 70 + 100 = 400$

$$b. \bar{x} = \frac{90+60+80+70+100}{5} = 80 \text{ pontos}$$

c. O candidato que fez mais pontos foi o candidato E (100 pontos), e o candidato que fez menos pontos foi o candidato B

02. Um professor de uma determinada disciplina resolveu que suas provas bimestrais terão pesos diferentes em cada bimestre e que seus alunos, só no final do 4º bimestre, receberão a média final. Escolhendo aleatoriamente um aluno desse professor, vamos, de acordo com suas notas e respectivos pesos, verificar sua média final.

O aluno no primeiro bimestre tirou 6 e a prova tinha peso 2, no 2º bimestre tirou 5 e o peso era 4, no 3º bimestre o aluno tirou 3 e o peso era 2 e, finalmente, no 4º bimestre tirou 10 e o peso era 4. Calcule sua média final.

Resposta

Para resolver esse exercício, desconsidere a média ponderada.

$$\bar{x}_p = \frac{6 \cdot 2 + 5 \cdot 4 + 3 \cdot 2 + 10 \cdot 4}{2 + 4 + 2 + 4} = \frac{78}{12} = 6,5 \quad \text{a média final do aluno foi 6,5.}$$

03. A tabela a seguir apresenta a distribuição de freqüências dos salários de um grupo de 50 empregados de uma empresa, num certo mês.

Número da classe	Salário do mês em reais	Número de empregados
1	1.000 — 2.000	20
2	2.000 — 3.000	18
3	3.000 — 4.000	9
4	4.000 — 5.000	3

O salário médio desses empregados, nesse mês, foi de:

- R\$ 2.637,00
- R\$ 2.500,00
- R\$ 2.420,00
- R\$ 2.400,00

Resposta

Os valores centrais (ou médios) das classes 1, 2, 3 e 4 são, respectivamente, 1.500, 2.500, 3.500 e 4.500 reais. Da seguinte maneira:

$$\text{classe 1} \Rightarrow \frac{1.000 + 2.000}{2} = \frac{3.000}{2} = 1.500$$

$$\text{classe 2} \Rightarrow \frac{2.000 + 3.000}{2} = \frac{5.000}{2} = 2.500$$

$$\text{classe 3} \Rightarrow \frac{3.000 + 4.000}{2} = \frac{7.000}{2} = 3.500$$

$$\text{classe 4} \Rightarrow \frac{4.000 + 5.000}{2} = \frac{9.000}{2} = 4.500$$

Encontramos a média aritmética simples dos limites das classes, para cada classe

Para determinar o salário médio, precisamos encontrar a média aritmética **ponderada** (os pesos serão as freqüências).

$$\bar{x}_p = \frac{1.500 \cdot 20 + 2.500 \cdot 18 + 3.500 \cdot 9 + 4.500 \cdot 3}{20 + 18 + 9 + 3}$$

$$\bar{x}_p = \frac{30.000 + 45.000 + 31.500 + 13.500}{50} = \frac{120.000}{50} = 2.400$$

Portanto, o salário médio é de R\$ 2.400

04. Calcule a média geométrica da série (2, 4, 8) Resposta: 4

2.1.3 – Médiana (\tilde{x})

É o valor “do meio” de um conjunto de dados, quando os dados estão dispostos em ordem crescente ou decrescente, cortando, assim, a distribuição em duas partes com o mesmo número de elementos.

É uma medida separatriz definida e exata, de fácil compreensão. Ela serve para análise comparativa e é representada por \tilde{x}

2.1.3.1 – Dados não-agrupados

Para dados não agrupados em classes:

Se n é **ímpar** $\rightarrow \tilde{x} = \left(\frac{n+1}{2}\right)^o$ termo

Com os números:

20 27 30 24 20 15 4

devemos, colocá-los em ordem:

4 15 20 20 24 27 30

Mediana igual a 20

Se n é **par** $\rightarrow \tilde{x} = \frac{\left(\frac{n}{2}\right)^o \text{ termo} + \left(\frac{n}{2} + 1\right)^o \text{ termo}}{2}$ temos que a mediana será a média aritmética dos dois elementos centrais, após todos os elementos serem colocados em ordem.

Com os números:

20 27 30 24 20 15

Deve-se colocá-los em ordem:

15 20 20 24 27 30

Mediana igual a $\tilde{x} = \frac{20+24}{2} = 22$

2.1.3.2 – Dados agrupados

Se os dados se agrupam em uma distribuição de freqüência, o cálculo da mediana se processa de modo muito semelhante àquele dos dados não-agrupados, implicando, porém, a determinação prévia das freqüências acumuladas. Ainda aqui, temos que determinar um valor tal que divida a distribuição em dois grupos que contenham o mesmo número de elementos.

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

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

2.1.3.2.1 – Sem intervalos de classe

Neste caso, é o bastante identificar a freqüência acumulada **imediatamente superior** à metade da soma das freqüências. A mediana será aquele valor da variável que corresponde a tal freqüência acumulada.

Por exemplo,

Nº de filhos	f_i	FA_i
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
	$\Sigma = 34$	

Sendo

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

A menor freqüência acumulada que supera esse valor é 18, que corresponde ao valor 2 da variável nº de filhos, sendo este o valor mediano. Logo, Md = 2 filhos.

2.1.3.2.2 – Com intervalos de classe

Neste caso, o problema consiste em determinar o **ponto do intervalo** em que está compreendida a mediana. Para tanto, temos inicialmente que determinar a classe na qual se acha a mediana – **classe mediana**. Tal classe será, evidentemente, aquela correspondente à freqüência acumulada imediatamente superior a $\frac{\sum f_i}{2}$.

Seja a distribuição:

i	X	f _i	F _i
1	150 — 154	4	4
2	154 — 158	9	13
3	158 — 162	11	24
4	162 — 166	8	32
5	166 — 170	5	37
6	170 — 174	3	40

Temos:

$$\frac{\sum f_i}{2} = \frac{40}{2} = 20$$

→ Classe mediana

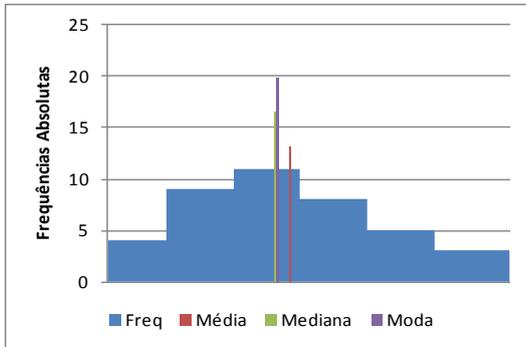
#01. Como há 24 valores incluídos nas três primeiras classes da distribuição e como pretendemos determinar o valor que ocupa o 20º lugar, a partir do início da série, vemos que este deve estar localizado na terceira classe (i = 3), supondo que as freqüências dessas classes estejam uniformemente distribuídas.

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

$$\left. \begin{array}{l} 11 \dots 4 \\ 7 \dots x \end{array} \right\} \Rightarrow x = \frac{7}{11} \cdot 4 = \frac{20-13}{11} \cdot 4$$

E a mediana será dada por:

$$\tilde{x} = Md = 158 + \frac{7}{11} \cdot 4 = 158 + 2,54 = 160,54$$



#02. Poderíamos num histograma determinar graficamente a mediana como sendo aquele ponto do eixo das abscissas por onde passa a vertical que divide o histograma em duas áreas iguais:

$$4 \cdot 4 + 4 \cdot 9 + x \cdot 11 = (4-x) \cdot 11 + 4 \cdot 8 + 4 \cdot 5 + 4 \cdot 3 \Rightarrow 52 + 11x = 44 - 11x + 64 \text{ ou } 22x = 56 \Rightarrow x = 2,5454$$

$$Md = 158 + 2,5454 = 160,54$$

#03. Existe, também, uma fórmula para calcularmos a mediana diretamente da tabela de distribuição de freqüências:

$$Md = l_i^* + \frac{\left[\frac{\sum f_i}{2} - FA_{anterior} \right] h}{f^*}$$

Onde: l_i^{*} é o limite inferior da classe mediana;

FA_{anterior} é a frequência acumulada da classe anterior à classe mediana;

f^{*} é a frequência absoluta da classe mediana;

h é a amplitude do intervalo da classe mediana.

No exemplo anterior:

$$Md = 158 + \frac{[20 - 13] \cdot 4}{11} = 160,54$$

Exercícios de Aplicação

01. {35, 36, 37, 38, 40, 40, 41, 43, 46} ⇒ $\tilde{x} = 40$
 {12, 14, 14, 15, 16, 16, 17, 20} ⇒ $\tilde{x} = \frac{15+16}{2} = 15,5$

02. Em um colégio, estão matriculados numa determinada classe 21 alunos. Durante o 1º bimestre foi feito um levantamento da frequência destes alunos e foram observadas as seguintes faltas: 0, 0, 3, 5, 7, 9, 0, 1, 2, 3, 11, 2, 3, 5, 6, 4, 10, 12, 0, 1, 2. Qual a mediana \tilde{x} das faltas? **Resposta: 3**

03. As idades dos atletas amadores de uma determinada modalidade esportiva são 14, 12, 16, 13, 17, 16 anos. Encontre a mediana da série. **Resposta: 15 anos**

04. Calcule a mediana da seguinte distribuição de freqüências:

Custos (R\$)	450 -- 550	550 -- 650	650 -- 750	750 -- 850	750 -- 850	850 -- 950	950 -- 1.050	1.050 -- 1.150
f _i	8	10	11	16	13	13	5	1

2.1.4 – Média x Médiana

A média é muito sensível a valores extremos de um conjunto de observações, enquanto a mediana não sofre muito com a presença de alguns valores muito altos ou muito baixos. A mediana é mais “robusta” do que a média.

Devemos preferir a mediana como medida sintetizadora quando o histograma do conjunto de valores é assimétrico, isto é, quando há predominância de valores elevados em uma das caudas.

Ex.: { 200, 250, 250, 300, 450, 460, 510 }
 $\bar{x}=345,7$ $\tilde{x}=300$
 Tanto \bar{x} como \tilde{x} , são boas medidas de posição.

Ex.: { 200, 250, 250, 300, 450, 460, 2300 }
 $\bar{x} = 601$ $\tilde{x} = 300$
 Devido ao valor 2300, \tilde{x} é preferível a \bar{x} .

2.1.5 – Moda

É o valor que ocorre com maior frequência em um conjunto de observações individuais. Para dados agrupados temos a **classe modal**. Em alguns casos pode haver mais de uma moda. Assim temos uma distribuição bimodal, trimodal, etc..

A moda é o valor em torno do qual os dados estatísticos tendem a estar mais pesadamente concentrados e é representada por M_o , também conhecida pelo nome de norma ou modo.

O termo moda foi introduzido por Pearson.

Exercícios de Aplicação

2.1.5.1 – Dados não-agrupados

01. Em um grupo de pessoas cujas idades são: 3, 2, 5, 2, 6, 2, 4, 4, 2, 7, 2 anos, a moda é 2 anos ($M_o = 2$). Portanto, denomina-se *unimodal*.

02. Algumas pessoas freqüentaram a escola por estes números de anos: 5, 3, 7, 5, 5, 8, 5, 3, 1, 1, 3, 3, 10, 3, 5.

Nesta série de números, podem-se ter duas modas: $M_o = \begin{cases} 3 \\ 5 \end{cases}$ Portanto bimodal

Para exemplificar tomamos dados observados e colocamos em uma tabela.

2.1.5.2 – Dados agrupados

2.1.5.2.1 – Sem intervalos de classe

03. Temos um grupo de pessoas cujas idades são: 3, 2, 5, 2, 6, 2, 4, 4, 2, 7, 2 anos:

Idade	2	3	4	5	6	7
Freqüência	5	1	2	1	1	1

Fica claro que a moda é 2 anos.

04. Tempo, em anos, que um grupo de pessoas freqüentou a escola.

Tempo de Escolaridade	
Tempo em anos de permanência na escola	Freqüência
1	2
3	5
5	5
7	1
8	1
10	1

Nesse exemplo, afirmamos que há duas modas, 3 anos e 5 anos, portanto o conjunto de dados é bimodal.

Nota importante

Quando não houver repetição de números, não haverá moda (o conjunto de dados é **amodal**).

2.1.5.2.2 – Com intervalos de classe

Quando os dados estão agrupados em classes,

X	x_i	n_i
10 — 20	15	2
20 — 30	25	4
30 — 40	35	10
40 — 50	45	6
50 — 60	55	2

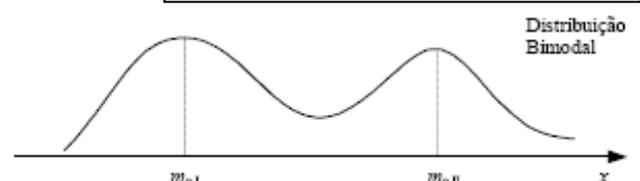
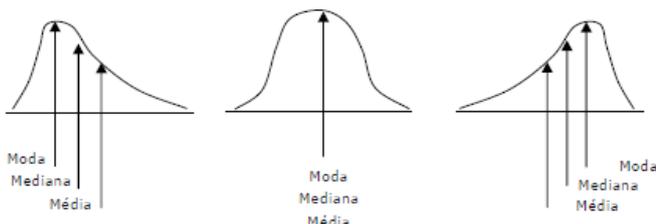
⇒ Classe Modal

Se precisarmos de um número representativo, tomamos o ponto médio do intervalo de classe.

Entretanto, temos a fórmula de **Czuber**:

$$M_o = l_i^* + \frac{D_1}{D_1 + D_2} h$$

$$D_1 = f^* - f_{\text{anterior}} \quad \text{e} \quad D_2 = f^* - f_{\text{posterior}}$$



Exercícios de Aplicação

- Nesta série, 1, 7, 9, 12, 17, não há moda, pois não há repetição de número. Observe a resolução deste exemplo.
- Considere os números 621, 310, 621, 201 e calcule:
 - a média aritmética (\bar{x});
 - a média aritmética ponderada (\bar{x}_p), com pesos 2, 3, 1 e 2, respectivamente;
 - a moda (M_o).

Resposta

Primeiramente, monta-se a tabela:

Números	621	310	201
Freqüência	2	1	1

$$a. \bar{x} = \frac{621+310+621+201}{4} = \frac{1.753}{4} = 438,25$$

$$b. \bar{x} = \frac{621 \cdot 2 + 310 \cdot 3 + 621 \cdot 1 + 201 \cdot 2}{2+3+1+2} = \frac{3.195}{8} = 399,375$$

c. Observando a tabela com os dados do exercício, verificamos que o número 621 aparece 2 vezes. Essa é a maior freqüência de acordo com a tabela, portanto $M_o = 621$.

2.1.6 – Separatrizes

Como vimos, a mediana caracteriza uma série de valores devido à sua posição central. No entanto, ela apresenta outra característica, tão importante quanto a primeira: ela separa a série em dois grupos que apresenta o mesmo número de valores.

Assim há outras medidas que não são de tendência central, mas que estão ligadas à mediana. Essas medidas, juntamente com a mediana são chamadas separatrizes. São elas: os **quartis**, os **percentis** e os **decis**.

2.1.6.1 – Quartis

São os valores de uma série que a dividem em quatro partes iguais.

Primeiro quartil – é o valor situado de tal modo na série que uma quarta parte (25%) dos dados é menor que ele e três quartos dos dados (75%) são maiores que ele.

Segundo quartil – coincide com a mediana

Terceiro quartil - é o valor situado de tal modo na série que três quartos (75%) dos dados é menor que ele e um quarto dos dados (25%) são maiores que ele.

2.1.6.1.1 – Dados agrupados

Neste caso usamos a mesma técnica do cálculo da mediana, bastando substituir a fórmula da mediana, $\frac{\sum f_i}{2}$ por:

$$\frac{k \sum f_i}{4}$$

Sendo k o número de ordem do quartil.

Podemos, também, aplicar uma fórmula para calcularmos os quartis diretamente da tabela de distribuição de freqüências:

$$Q_k = l_i^* + \frac{\left[\frac{k \sum f_i}{4} - FA_{anterior} \right] h}{f^*}$$

Onde: l_i^* é o limite inferior da classe do quartil k;

$FA_{anterior}$ é a freqüência acumulada da classe anterior à classe quartil k;

f^* é a freqüência absoluta da classe do quartil k;

h é a amplitude do intervalo da classe do quartil k.

Exercícios de Aplicação

01. Seja a tabela de distribuição:

i	X	f_i	F_i
1	150 — 154	4	4
2	154 — 158	9	13
3	158 — 162	11	24
4	162 — 166	8	32
5	166 — 170	5	37
6	170 — 174	3	40

Temos para o Primeiro Quartil:

$$\frac{\sum f_i}{4} = \frac{40}{4} = 10$$

$$\rightarrow Q_1 = 154 + \frac{(10 - 4)x4}{9} = 154 + \frac{24}{9} = \mathbf{156,66}$$

Para o 3º Quartil, temos:

$$\frac{3 \sum f_i}{4} = \frac{3x40}{4} = 30$$

$$\rightarrow Q_3 = 162 + \frac{(30 - 24)x4}{8} = 162 + \frac{24}{8} = \mathbf{165}$$

2.1.6.2 – Percentis

“ O percentil de ordem p , $0 \leq p \leq 100$, de um conjunto de valores dispostos em ordem crescente é um valor tal que $p\%$ das observações estão nele ou abaixo dele e $(1 - p)\%$ estão nele ou acima dele.”

2.1.6.2.1 – Dados não-agrupados

Ex: Para valores de 51 a 100, ordenados crescentemente:

P_{25} deixa 25% dos dados ($12,5 \Rightarrow 13$ valores) nele ou abaixo dele e 75% dos dados ($37,5 \Rightarrow 38$ valores) nele ou acima dele. Assim: $P_{25} = 63$.

Similarmente, P_{80} deixa 80% dos dados (40 valores) nele ou abaixo dele e 20% dos dados (10 valores) nele ou acima dele. Assim: $P_{80} = \frac{90+91}{2} = 90,5$

2.1.6.2.1 – Dados agrupados

Para dados agrupados em classes, os percentis podem ser obtidos por interpolação linear (regra de três simples).

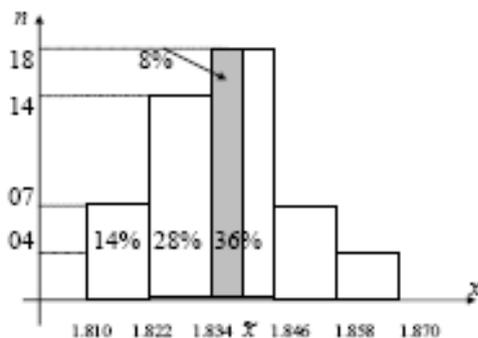
Ex.: Dada a distribuição de frequência de uma variável X qualquer:

X	x_i	N_i	N_i
1,810 — 1,822	1,816	7	7
1,822 — 1,834	1,828	14	21
1,834 — 1,846	1,840	18	39
1,846 — 1,858	1,852	7	46
1,858 — 1,870	1,864	4	50

Temos que, para P_{50} (50% de 50) será o 25º elemento, está na terceira classe. Isto porque a segunda classe contém 21 elementos e a terceira, 39 elementos. Logo, o 25º elemento estará na 3ª classe.

$$P_k = l_i^* + \frac{\left[\frac{k \sum f_i}{100} - FA_{anterior} \right] h}{f^*}$$

$$P_{50} = 1834 + \frac{[50x0,5 - 21] 12}{18} = 1836,67$$



$$\frac{1.846 - 1.834}{36\%} = \frac{\tilde{x} - 1.834}{8\%} \Rightarrow \tilde{x} = 1.837$$

Outro processo gráfico pode ser usado para o cálculo desses percentis. (Veja Ogiva de Galton). Tal processo exige rigor no traçado e deve-se preferir papel milimetrado.

Obs.: As calculadoras geralmente não fornecem mediana e percentis.

Exercícios de Aplicação

01 UFRN-RN Uma prova foi aplicada em duas turmas distintas. Na primeira, com 30 alunos, a média aritmética das notas foi 6,40. Na segunda, com 50 alunos, foi 5,20. A média aritmética dos 80 alunos foi:

- a) 5,65 c) 5,75
b) 5,70 d) 5,80

Resolução $\frac{30 \cdot 6,40 - 50 \cdot 5,20}{80} = 5,65$ Resp A

02 Fuvest-SP Uma prova continha cinco questões, cada uma valendo dois pontos. Em sua correção, foram atribuídas a cada questão apenas as notas 0 ou 2, caso a resposta estivesse, respectivamente, errada ou certa. A soma dos pontos obtidos em cada questão forneceu a nota do aluno. Ao final da correção, produziu-se a seguinte tabela, contendo a porcentagem de acertos em cada questão.

Questão	Porcentagem de acerto
1	30%
2	10%
3	60%
4	80%
5	40%

Logo, a média das notas da prova foi:

- a) 3,8 d) 4,4
b) 4,0 e) 4,6
c) 4,2

Resolução $30\% \cdot 2 + 10\% \cdot 2 + 60\% \cdot 2 + 80\% \cdot 2 + 40\% \cdot 2 = 4,4$

Resposta: D

03 TCU - Considere a distribuição de frequências dos tempos de auditoria:

Tempo de auditoria (min)	Frequência
10 ... 19	10
20 ... 29	20
30 ... 39	40
40 ... 49	20
50 ... 59	10

Assinale a opção **incorreta**.

- a) O intervalo de classe modal é dado por [30; 39].
b) O tempo médio de auditoria é dado por 34,5 min.
c) A mediana, a moda e a média da distribuição são coincidentes.
d) A distribuição é assimétrica.
e) 30% das auditorias demoram menos de trinta minutos.

Resolução

- a) Verdadeiro: A classe modal é aquela que possui a maior frequência.
b) Verdadeiro: O tempo médio será encontrado pela média dos tempos, e os tempos de cada classe serão representados pelos seus valores médios.

$$\bar{T} = \frac{10 \cdot 14,5 + 20 \cdot 24,5 + 40 \cdot 34,5 + 20 \cdot 44,5 + 10 \cdot 54,5}{10 + 20 + 40 + 20 + 10} = 34,5 \text{ min}$$

- c) Verdadeiro: As três medidas de centralidade são iguais a 34,5 minutos, já que estamos diante de uma seqüência simétrica.
d) Falso: A seqüência é simétrica, já que, se "dobrássemos" a seqüência ao meio, as metades iriam coincidir.
e) Verdadeiro: A soma das frequências menores que 30 minutos é igual a 30 em um universo de 100.

Resposta: D

Exercícios Propostos

- Responda as questões abaixo:
 - Média, Mediana e Moda são medidas de:
 - () Dispersão
 - () Posição
 - () Assimetria
 - () Curtose
 - Na série 10, 20, 40, 50, 70, 80 a mediana será:
 - () 30
 - () 35
 - () 40
 - () 45
 - 50% dos dados da distribuição situa-se:
 - () abaixo da média
 - () acima da mediana
 - () abaixo da moda
 - () acima da média

- Calcule para cada caso abaixo a respectiva média.

- a. 7, 8, 9, 12, 14
b.

X_i	3	4	7	8	12
n_i	2	5	8	4	3

c.

Classes	68 -- 72	72 -- 76	76 -- 80	80 -- 84
n_i	8	20	35	40

3. Calcule o valor da mediana:

- a. 82, 86, 88, 84, 91, 93
b.

X_i	73	75	77	79	81
n_i	2	10	12	5	2

c.

Classes	1 -- 3	3 -- 5	5 -- 7	7 -- 9	9 -- 11	11 -- 13
n_i	2	5	8	6	4	3

4. Calcule a moda:

- a. 3, 4, 7, 7, 7, 8, 9, 10
b.

X_i	2,5	3,5	4,5	6,5
n_i	7	17	10	5

c.

Classes	10 -- 20	20 -- 30	30 -- 40	40 -- 50
n_i	7	19	28	32

5. Para a distribuição abaixo calcular o 2º decil, o 4º percentil e o 3º quartil:

Classes	10 -- 20	20 -- 30	30 -- 40	40 -- 50
n_i	7	19	28	32

6. Em 15 dias, um restaurante serve almoço para 40-52-55-38-40-48-56-56-60-37-58-63-46-50-61 fregueses. Determine a mediana

7. (TTN) Considere a distribuição de frequência transcrita a seguir para responder às próximas três questões.

Peso (kg)	f_i
2 → 4	9
4 → 6	12
6 → 8	6
8 → 10	2
10 → 12	1

8. A média da distribuição é igual a:

- a) 5,27 kg d) 5,19 kg
b) 5,24 kg e) 5,30 kg
c) 5,21 kg

9. A mediana da distribuição é igual a:

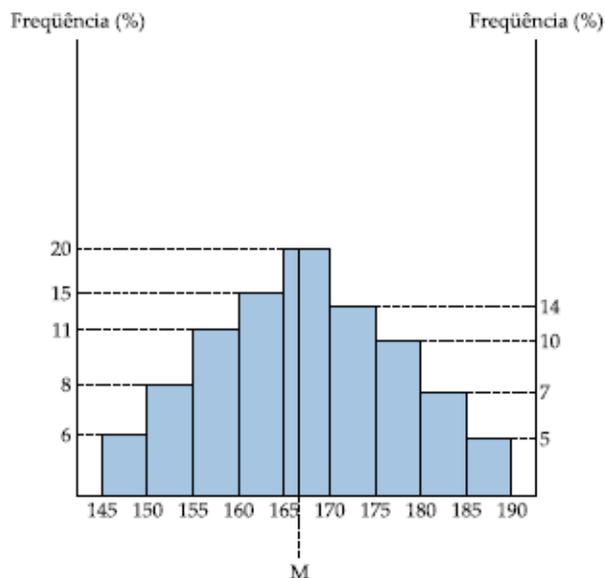
- a) 5,30 kg
b) 5,00 kg
c) um valor inferior a 5 kg.
d) 5,10 kg
e) 5,20 kg

10. A moda da distribuição:

- a) coincide com o limite superior de um intervalo de classe.
b) coincide com o ponto médio de um intervalo de classe.
c) é maior do que a mediana e do que a média geométrica.
d) é um valor inferior à média aritmética e à mediana.

e) pertence a um intervalo de classe distinto do da média aritmética.

11. **PUC-SP** O histograma seguinte representa a distribuição das estaturas de 100 pessoas e as respectivas freqüências. Por exemplo, na terceira classe (155 – 160) estão situados 11% das pessoas com estaturas de 1,55 m a 1,59 m. A quinta classe (165 – 170) chama-se classe mediana. Pelo ponto M situado na classe mediana, traça-se uma reta paralela ao eixo das freqüências, de modo a dividir a área da figura formada pelos nove retângulos das freqüências em duas regiões de mesma área. Determine a abscissa do ponto M (mediana das observações).



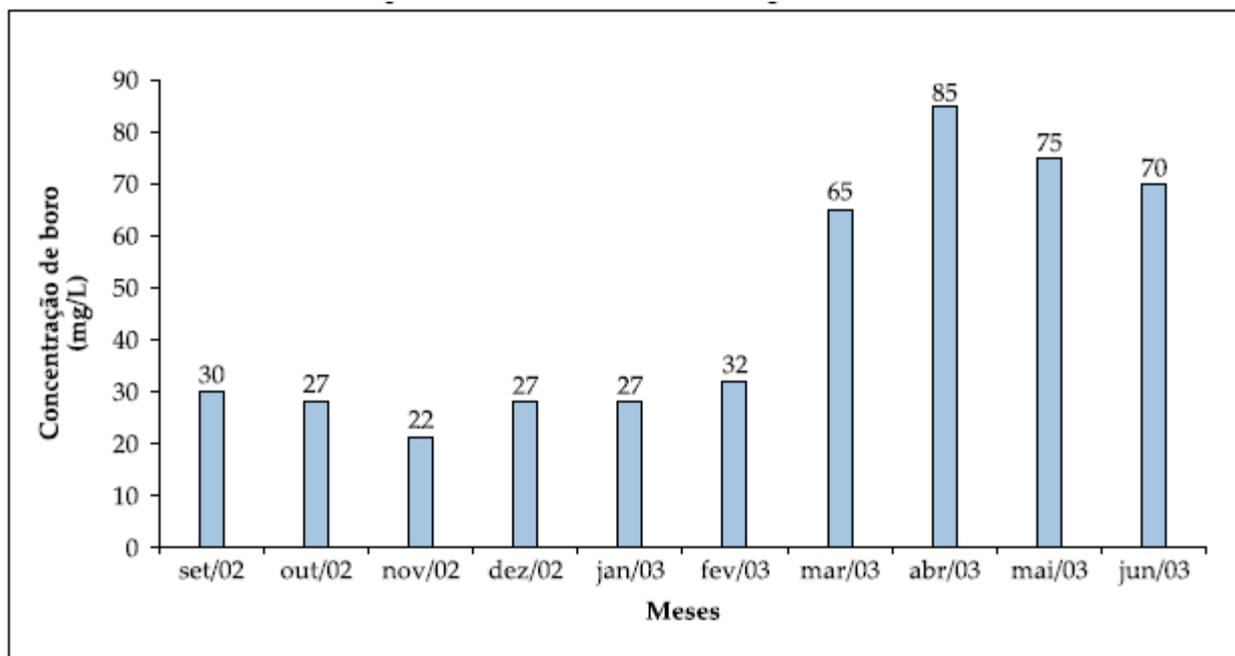
12. **UnB-DF** A tabela abaixo mostra os diferentes tipos sangüíneos, com os correspondentes antígenos, e a sua distribuição em uma população de 10.000 indivíduos.

Antígenos presentes			Tipo sangüíneo	Número de indivíduos
A	B	Rh		
Não	Não	Não	O-	660
Não	Não	Sim	O+	3.740
Sim	Não	Não	A-	630
Sim	Não	Sim	A+	3.570
Não	Sim	Não	B-	150
Não	Sim	Sim	B+	850
Sim	Sim	Não	AB-	60
Sim	Sim	Sim	AB+	340

À população anteriormente estudada foi acrescentado um grupo de 1.000 indivíduos, registrando-se nesse grupo apenas os tipos sangüíneos O-, A-, B- e AB-. Considerando a série numérica formada pelos números da tabela anterior e a nova série que se obtém ao se acrescentar à população anterior esse novo grupo de indivíduos, julgue os itens abaixo.

- (1) A moda da nova série é maior que a da série anterior.
- (2) A mediana da nova série é maior que a da série anterior.
- (3) Na representação dos dados em gráficos de setores, os indivíduos que apresentam o antígeno Rh corresponderiam a percentuais iguais nas duas séries.

13. **PAS-UnB-DF** Um laudo da companhia de saneamento da cidade de Padre Cícero denunciou que os níveis de boro no ribeirão Vermelho, que abastece a população daquela cidade, atingiram valores muito superiores aos permitidos por lei (0,75 mg/L). O laudo revela que a possível origem do boro é uma substância chamada hidroboracita, matéria-prima utilizada na fabricação de fibras de vidro. Sabe-se que uma indústria de fibras de vidro tem depositado rejeitos industriais em uma voçoroca localizada no aterro Pedra Azul, nas proximidades da nascente do córrego Cristal, afluente do ribeirão Vermelho.



A figura anterior mostra a concentração de boro no córrego Cristal, no período de setembro de 2002 a junho de 2003, medida no dia 15 de cada mês. Considerando o texto III e a seqüência numérica dos valores dessas concentrações, julgue os itens seguintes.

- (1) Caso tenha sido retirada uma amostra de 3 L de água do córrego Cristal em abril de 2003, seria necessário adicionar mais de 300 L de água destilada a essa amostra, para que os níveis de boro ficassem dentro do permitido por lei.
- (2) Para a seqüência numérica citada, a moda é superior à mediana.
- (3) Se, em 15 de julho de 2003, a concentração medida foi igual à média aritmética da seqüência numérica das 10 concentrações medidas anteriormente, então o desvio-padrão da nova seqüência numérica, com 11 medições, é superior ao desvio-padrão da seqüência com 10 medições.
- (4) Na seqüência numérica de concentrações, existe pelo menos uma medição superior à soma da média aritmética com o desvio-padrão dessa seqüência.

14. **UnB-DF** Utilizando dois instrumentos distintos, A e B, foi feita, com cada um deles, uma série de vinte medições de um mesmo ângulo, e os resultados obtidos estão listados na tabela abaixo, em que a frequência A e a frequência B indicam a quantidade de vezes em que o resultado foi encontrado com os instrumentos A e B, respectivamente.

Frequência	Resultado das medições							
	67°30'10''	67°30'12''	67°30'13''	67°30'14''	67°30'15''	67°30'16''	67°30'17''	67°30'18''
A	1	1	2	4	4	3	2	3
B	1	1	2	3	6	2	2	3

Com base nessas informações, julgue os itens que se seguem.

- (1) A média da série dos resultados das medições feitas com o instrumento A é menor que 67° 30' 14''.
- (2) As séries dos resultados das medições feitas com os instrumentos A e B têm o mesmo desvio-padrão.
- (3) A moda e a média da série dos resultados das medições feitas com o instrumento B são iguais.
- (4) A mediana da série dos resultados das medições feitas com o instrumento B é maior que a da série dos resultados das medições feitas com o instrumento A.

15. **UnB-DF** - Em uma usina de álcool, foi selecionada uma certa variedade de cana do seu canavial. Tomando-se várias unidades, ao acaso, em diversos pontos da lavoura, obtiveram-se, em quilogramas, os pesos seguintes:

1,58	1,32	1,76	1,51	1,50
1,38	1,55	1,71	1,54	1,67

Nessas condições, julgue os itens seguintes.

- (1) A média aritmética desses 10 dados é 1,552 kg.
- (2) Podemos afirmar que 1,552 kg é o peso médio de uma cana para toda a lavoura.

16. **UnB-DF** - Para comparar dois métodos de alfabetização, A e B, um professor tomou um conjunto de alunos, dividiu-os ao acaso em dois grupos e alfabetizou um dos grupos pelo método A e o outro, pelo método B. Terminado o período de alfabetização, o professor submeteu os dois grupos de alunos à mesma prova. Os alunos obtiveram, nessa prova, as notas apresentadas na tabela a seguir.

Notas dos alunos segundo os métodos de alfabetização								
A	6	5	7	3	5	2	4	8
B	7	9	6	7	6	6	9	6

Nessas condições, julgue os itens a seguir:

- (1) As médias das notas dos métodos A e B são, respectivamente, 5,0 e 7,0.
 (2) Na amostra observada, a nota média dos alunos alfabetizados pelo método B é 40% maior do que a nota média dos alunos alfabetizados pelo método A.
17. **UnB-DF** - A tabela abaixo apresenta o levantamento das quantidades de peças defeituosas para cada lote de 100 unidades fabricadas em uma linha de produção de autopeças, durante um período de 30 dias úteis.

Dia	Nº de peças defeituosas	Dia	Nº de peças defeituosas
1	6	16	7
2	4	17	5
3	3	18	6
4	4	19	4
5	2	20	3
6	4	21	2
7	3	22	6
8	5	23	3
9	1	24	5
10	2	25	2
11	1	26	1
12	5	27	3
13	4	28	2
14	1	29	5
15	3	30	7

Considerando S a série numérica de distribuição de freqüências de peças defeituosas por lote de 100 unidades, julgue os itens abaixo.

- (1) A moda da série S é 5.
 (2) Durante o período de levantamento desses dados, o percentual de peças defeituosas ficou, em média, abaixo de 3,7%.
 (3) Os dados obtidos nos 10 primeiros dias do levantamento geram uma série numérica de distribuição de freqüências com a mesma mediana da série S.
18. **UnB-DF**

Respirando veneno

O inverno de 1998 vai ter uma péssima qualidade do ar – uma das piores da história. Quem garante são os especialistas. A estiagem provocada pelo El Niño deve tornar este período mais seco, dificultando a dispersão de gases e fumaças. Os técnicos acreditam que este inverno será ainda pior que o de 1997. Saiba em quantos dias do ano passado a poluição ficou acima dos níveis aceitáveis nas cidades que têm controle da qualidade do ar. Observe a tabela.

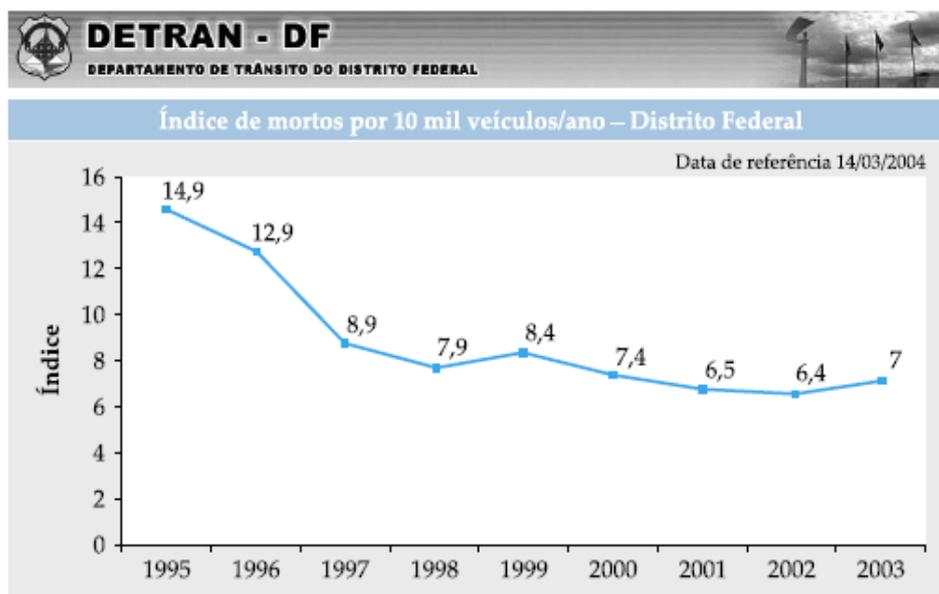
Cidade	Dias
Alta Floresta	90
Belo Horizonte	0
Contagem	12
Cubatão	48
Cuiabá	90
Curitiba	14
Porto Alegre	6
Rio de Janeiro	80
São João de Meriti	22
São Paulo	132
Volta Redonda	18

Veja, 24/06/98 (com adaptações)

Considerando a série numérica formada pelos números da tabela, julgue os seguintes itens.

- (1) O número de dias em que Volta Redonda apresentou poluição acima dos níveis aceitáveis é a mediana da série, indicando que, das 11 cidades, cinco apresentaram índices de poluição menores que o de Volta Redonda.
- (2) A média aritmética da série indica o número de dias com poluição acima dos níveis aceitáveis que cada cidade analisada teria se o total de dias fosse igualmente distribuído por todas as cidades analisadas.
- (3) Tanto um gráfico de setores como um gráfico de barras são representações adequadas para a série estudada.
- (4) A moda da série é de 132, indicando que São Paulo foi a cidade que mais tempo ficou com índices de poluição acima dos níveis aceitáveis.

19. UnB-DF



A frota foi ajustada excluindo-se os veículos com placas antigas (2 letras), em virtude da adequação com o sistema Renavam (Registro Nacional de Veículos Automotores) do Departamento de Trânsito.
GDF/SSPDF/DETRAN

Dados do Departamento Nacional de Trânsito (Denatran) revelam que, por dia, os acidentes de trânsito no Brasil matam cerca de 100 pessoas e ferem outras 1.000, muitas vezes deixando seqüelas irreversíveis. Os gastos decorrentes da violência no trânsito chegam a mais de R\$ 10 bilhões por ano.

Segundo o diretor do Denatran, entre os principais fatores que colaboram para o aumento de acidentes nas vias urbanas e rodoviárias estão dois velhos conhecidos: o uso de álcool e o excesso de velocidade.

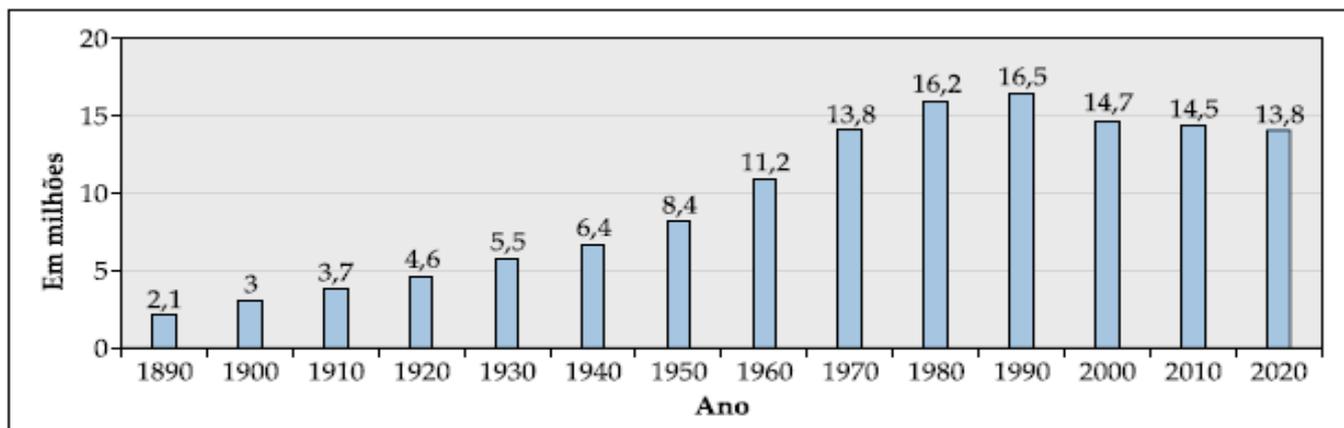
Com relação a essas informações, julgue os itens seguintes.

- (1) As informações contidas no gráfico são suficientes para que se possa concluir que o número de vítimas fatais de acidentes de trânsito no DF foi maior em 1999 que em 2002.
- (2) No DF, se a frota de veículos em 1996 fosse 10% menor que a frota de veículos em 2000, então o número de mortos em acidentes de trânsito em 2000 teria sido inferior a 60% do número de mortos em acidentes de trânsito em 1996.
- (3) A média aritmética da seqüência numérica formada pelos índices correspondentes aos anos de 1995, 1996, 1997, 1998 e 1999 é superior a 10,7.

20. (UnB-DF)



Sebastião Salgado, In: Êxodos



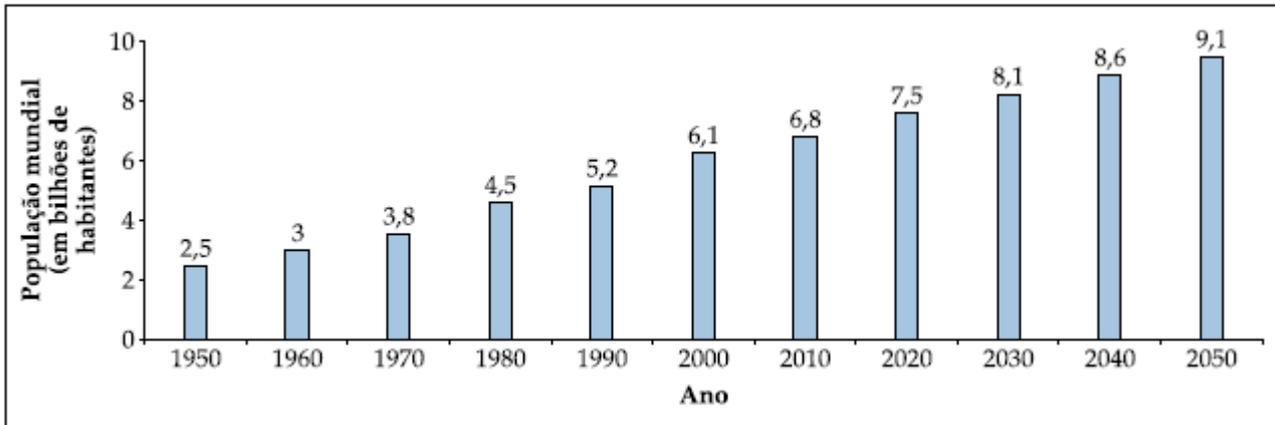
IBGE (com adaptações)

O gráfico acima ilustra o número de indivíduos de 0 a 4 anos de idade na população brasileira, incluindo previsões para os anos de 2010 e 2020. Com base nas informações do gráfico, julgue os itens que se seguem.

- (1) As informações do gráfico são suficientes para se concluir que, no período de 1990 a 2000, a mortalidade infantil no Brasil aumentou.
- (2) Em 2020, a população brasileira com idade entre 20 e 24 anos, desconsiderando-se mortes e migração, será superior a 14 milhões.
- (3) Em 1930, desconsiderando-se mortes e migração, a população brasileira com idade não superior a 44 anos era inferior a 19 milhões.
- (4) A mediana da seqüência numérica relativa à população brasileira de indivíduos entre 0 e 4 anos de idade, no período de 1890 a 2020, é superior a 10.

21. **UnB-DF** Estima-se que, em 2050, a população mundial será de, aproximadamente, 9,1 bilhões de habitantes, como mostra o gráfico a seguir. Uma boa qualidade de vida no futuro está associada ao desenvolvimento de

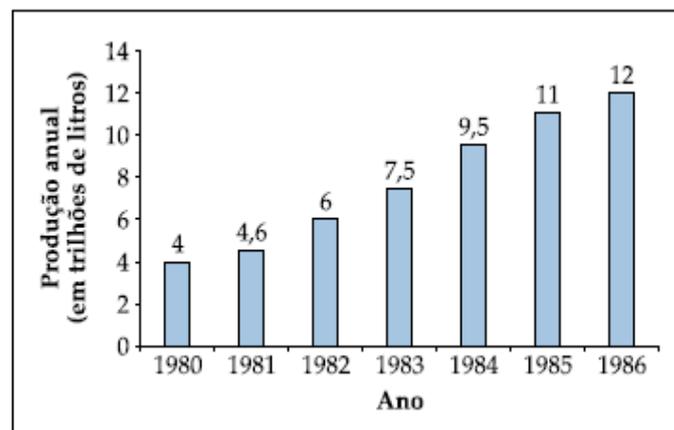
profunda consciência socioambiental, aliada à utilização de tecnologias para solucionar problemas atuais, como a poluição atmosférica, o acúmulo de lixo e a falta de tratamento de esgoto nas zonas urbanas.



Considerando as informações acima e os vários aspectos que o tema suscita, julgue os itens seguintes.

- (1) De acordo com o gráfico, em 2050 a população mundial será inferior a $9,1 \cdot 10^9$ habitantes.
- (2) De acordo com o gráfico, no período 2010-2050, o crescimento da população mundial, em bilhões de habitantes, será superior ao ocorrido no período 1950-1990.
- (3) A moda da seqüência numérica formada pelos valores mostrados no gráfico correspondentes à população mundial no período 1950-2050 é superior a 6,3 bilhões.
- (4) A média da seqüência numérica formada pelos valores mostrados no gráfico correspondentes à população mundial no período 1950-2050 é superior à mediana dessa seqüência.

22. **PAS-UnB-DF** O Proálcool — Programa Nacional do Álcool —, criado em 1975 para reduzir a importação de petróleo, foi uma importante iniciativa para substituir combustíveis fósseis por um combustível alternativo e renovável: o álcool etílico. O programa foi fortemente subsidiado e, a partir de 1978, o Brasil passou a exportar etanol, sobretudo para os Estados Unidos da América e para o Japão. O gráfico ao lado mostra a produção anual brasileira de álcool etílico de 1980 a 1986. Representando por p_n a produção brasileira de álcool etílico no ano $1980 + n$, $n = 0, 1, \dots, 6$, julgue os itens seguintes.



- (1) A média aritmética da seqüência numérica $\{p_n\}$, $n = 0, 1, \dots, 6$, é menor que a sua mediana.
- (2) Para cada $n = 0, 1, \dots, 6$, $p_n \in [8 - \sigma, 8 + \sigma]$, em que σ é o desvio-padrão da seqüência numérica $\{p_n\}$.
- (3) Se p_7 representa a produção de álcool etílico brasileira em 1987 e p_7 é menor que a mediana da seqüência $\{p_n\}$, $n = 0, 1, \dots, 6$, então a média aritmética da seqüência $\{p_n\}$, $n = 0, 1, \dots, 6$ é maior que a da seqüência $\{p_n\}$, $n = 0, 1, \dots, 7$.
- (4) Se, a partir de 1983, a produção anual brasileira de álcool etílico tivesse crescido segundo uma progressão aritmética de razão $p_3 - p_2$, então, em 1986, essa produção teria sido superior àquela apresentada no gráfico para esse ano.
- (5) Existe uma função quadrática, da forma $f(x) = ax^2 + b$, em que a e b são constantes reais, tal que o gráfico de f contém os pontos da forma (n, p_n) , $n = 0, 1, \dots, 6$.

2.2 – Medidas de Dispersão ou Variabilidade

Vimos que a **moda**, a **mediana** e a **média aritmética** possuem a função de representar, a partir de um único número, a seqüência a ser analisada. Porém, tal método ainda é muito incompleto para que nós possamos tirar alguma conclusão sobre o trabalho. É necessário que possamos enxergar algo mais nessa seqüência que estamos analisando, como, por exemplo, certa “personalidade” da seqüência. Observe a seguinte situação: quatro turmas, uma de cada um dos cursos Ciência da Computação, Matemática, Ciências Contábeis e Fisioterapia, fizeram uma prova de estatística e quando o professor verificou a média das notas de cada turma, constatou que, em cada uma das quatro turmas, a média dos alunos foi igual a 6,0. E aí? Será que podemos concluir que o desempenho das quatro turmas foi o mesmo? Será que todos os alunos, de todas as turmas, tiraram nota 6,0 na prova? É óbvio que, nesse momento, o bom senso fala mais alto e podemos, no mínimo, desconfiar de que não. Pois é exatamente aí que reside a tal “personalidade” que podemos atribuir a cada turma em relação ao comportamento das notas. O que quero dizer é que, com as **medidas de dispersão**, seremos capazes de verificar que, por mais que a média das turmas na prova de estatística tenha sido 6,0, poderemos com tais medidas determinar as turmas que tiveram um comportamento homogêneo, em que os alunos tiraram notas próximas de 6,0, como também determinar as turmas que tiveram um comportamento heterogêneo em relação à nota 6,0, ou seja, por mais que a média tenha sido 6,0, as notas não foram próximas de 6,0. Em outras palavras, torna-se necessário estabelecer medidas que indiquem o grau de dispersão em relação ao valor central. Algumas medidas de dispersão que sintetizam essa variabilidade são:

2.2.1 – Amplitude (H)

É uma medida de dispersão muito rápida e, ao mesmo tempo, **muito imprecisa**, pois consiste simplesmente em verificar a diferença entre o maior valor e o menor valor obtido na coleta de dados. Essa é nossa velha conhecida. Mesmo assim um exemplo.

Exercícios de Aplicação

Sejam os pesos dos alunos de uma série escolar:

Pessoas	Peso (kg)
Robson	30
Adriano	15
Vinicius	55
S ^{ta} Adélia	52
Ranzani	60
Camila	53
Bertolo	75
Daniel	20
Kátia	40

Na tabela ao lado, temos o peso das pessoas de um determinado grupo analisado e podemos verificar que a amplitude total foi de: $AT = 75 - 15 = 60$

Quando os dados estiverem agrupados com intervalo de classe, tomamos a diferença entre o L_i (limite superior) da última classe e l_i (limite inferior) da primeira classe.

2.2.2 – Desvio Médio

Como a palavra desvio está associada à diferença, temos que, no contexto da nossa matéria, o desvio deve ser empregado com a diferença do elemento analisado em relação à média, ou seja, o quanto o elemento se afasta da média da seqüência. Daí é importante perceber que essa diferença deve ser necessariamente trabalhada em módulo, pois não tem sentido a distância negativa¹.

2.2.2.1 – Dados Não Agrupados

O desvio médio, então, passa a ser encontrado como a média aritmética de todos os desvios em valor absoluto.

$$\text{Desvio Médio} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}| + \dots + |x_N - \bar{x}|}{N} = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{N}$$

Exercícios de Aplicação

Com os dados do exercício anterior, temos:

$$\bar{x} = \frac{30 + 15 + 55 + 52 + 60 + 53 + 75 + 20 + 40}{9} = 44,4$$

¹ E também porque é fácil ver que a soma dos desvios, é identicamente nula e que, portanto, não serve como medida de dispersão:

$\sum_{j=1}^N (x_j - \bar{x}) = \sum_{j=1}^N (x_j) - \sum_{j=1}^N (\bar{x}) = N \bar{x} - N \bar{x} = 0$. Por isso temos duas opções: a) considerar os desvios em valor absolutos ou b) considerar os quadrados dos desvios.

Desvio Médio

$$= \frac{|30 - 44,4| + |15 - 44,4| + |55 - 44,4| + |52 - 44,4| + |60 - 44,4| + |53 - 44,4| + |75 - 44,4| + |20 - 44,4| + |40 - 44,4|}{9}$$

$$= 16,17$$

2.2.2.2 –Dados Agrupados

Dados Agrupados Sem Intervalo de Classe

Neste caso devemos acrescentar uma coluna a mais na nossa tabela de distribuição de frequências para calcularmos o desvio médio. Seja então o nosso exemplo das vendas:

ROL	x_i	Frequência f_i	Freq. Acum. FA_i	Freq. Acum. FAR_i	$ (x_i - x_{med}) $
280	280	2	2	0,08	78,48
280	305	1	3	0,12	53,48
305	310	1	4	0,16	48,48
310	320	1	5	0,2	38,48
320	330	2	7	0,28	28,48
330	340	1	8	0,32	18,48
330	341	1	9	0,36	17,48
340	355	1	10	0,4	3,48
341	360	1	11	0,44	1,52
355	365	1	12	0,48	6,52
360	369	1	13	0,52	10,52
365	370	3	16	0,64	11,52
369	371	1	17	0,68	12,52
370	375	1	18	0,72	16,52
370	380	1	19	0,76	21,52
370	390	1	20	0,8	31,52
371	400	2	22	0,88	41,52
375	401	1	23	0,92	42,52
380	420	1	24	0,96	61,52
390	430	1	25	1	71,52
400		0			
400	Σ	25		Σ	616,08
401					
420					
430					

Agora o desvio médio passa a ser calculado a partir da média ponderada com f_i como ponderação:

$$\text{Desvio Médio} = = \frac{\sum_{i=1}^N |x_i - \bar{x}| \cdot f_i}{\sum_{i=1}^N f_i} = 31,504$$

Dados Agrupados Com Intervalo de Classe

Aqui, convencionamos que *todos os valores incluídos em um determinado intervalo de classe coincidem com o seu ponto médio*, e determinamos o desvio médio.

2.2.3 – Variância

A variância é uma medida de dispersão muito parecida com o desvio médio, a única diferença em relação a este é que, na variância, ao invés de trabalharmos em módulo as diferenças entre cada elemento e a média, tomamos os quadrados das diferenças. Isso se dá pelo fato de que, elevando cada diferença ao quadrado, continuamos trabalhando com números não negativos, como também pelo fato de que, em procedimentos estatísticos mais avançados, tal método facilita futuras manipulações algébricas.

2.2.3.1 – Dados Não Agrupados

$$\text{Variância } \sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N} = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}$$

Exercícios de Aplicação

Com os dados do exercício anterior do peso dos alunos da 3º ano de Matemática, temos:

$$\begin{aligned} \text{Variância} &= \frac{(30 - 44,4)^2 + (15 - 44,4)^2 + (55 - 44,4)^2 + (52 - 44,4)^2 + (60 - 44,4)^2 + (53 - 44,4)^2 + (75 - 44,4)^2 + (20 - 44,4)^2 + (40 - 44,4)^2}{9} \\ &= 345,57 \end{aligned}$$

2.2.3.2 – Dados Agrupados

Dados Agrupados Sem Intervalo de Classe

Neste caso devemos acrescentar uma coluna a mais na nossa tabela de distribuição de freqüências para calcularmos o desvio médio. Seja então o nosso exemplo das vendas:

ROL	x_i	Freqüência f_i	Freq. Acum. FA_i	Freq. Acum. FAR_i	$ (x_i - x_{med}) $	$(x_i - x_{med})^2$
280	280	2	2	0,08	78,48	6159,1104
280	305	1	3	0,12	53,48	2860,1104
305	310	1	4	0,16	48,48	2350,3104
310	320	1	5	0,2	38,48	1480,7104
320	330	2	7	0,28	28,48	811,1104
330	340	1	8	0,32	18,48	341,5104
330	341	1	9	0,36	17,48	305,5504
340	355	1	10	0,4	3,48	12,1104
341	360	1	11	0,44	1,52	2,3104
355	365	1	12	0,48	6,52	42,5104
360	369	1	13	0,52	10,52	110,6704
365	370	3	16	0,64	11,52	132,7104
369	371	1	17	0,68	12,52	156,7504
370	375	1	18	0,72	16,52	272,9104
370	380	1	19	0,76	21,52	463,1104
370	390	1	20	0,8	31,52	993,5104
371	400	2	22	0,88	41,52	1723,9104
375	401	1	23	0,92	42,52	1807,9504
380	420	1	24	0,96	61,52	3784,7104
390	430	1	25	1	71,52	5115,1104
400		0				
400	Σ	25		Σ	616,08	
401						
420						
430						

Uma fórmula alternativa para a *variância* é dada por:

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

Relacionados à inferência estatística, alguns autores usam $(n - 1)$ como divisor para a variância:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}, \text{ e isto será visto adiante (tendenciosidade)}$$

Obs.: Muitas calculadoras científicas possuem duas medidas para desvio padrão. Uma associada à divisão por n (simbolizada geralmente por σ ou σ_n) e outra associada à divisão por $n - 1$ (chamada também de não-polarizada, simbolizada geralmente por S ou σ_{n-1}). Verifique a simbologia usada pela sua calculadora, caso você possua uma!

A variância da **população** é dada por:

$$\text{Variância} = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot f_i}{\sum_{i=1}^N f_i} = 1515,45$$

A variância da **amostra** é dada por:

$$\text{Variância} = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot f_i}{(\sum_{i=1}^N f_i) - 1} = 1578,593$$

Dados Agrupados Com Intervalo de Classe

Aqui, convencionamos que *todos os valores incluídos em um determinado intervalo de classe coincidem com o seu ponto médio* x_i , e determinamos a variância da mesma maneira que fizemos com os dados agrupados sem intervalo de classe.

A fórmula original como a média dos quadrados dos desvios é: $s^2 = \frac{\sum (x_i - \bar{x})^2}{\sum f_i}$. Porém, isto é trabalhoso manualmente.

Classes	Freq. f_i	fr_i	FA	FAR	x_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
< 280	0	0,00%	0	0,00%	265	0	0
[280,310)	3	12,00%	3	12,00%	305	915	279.075
[310,340)	4	16,00%	7	28,00%	325	1.300	422.500
[340,370)	6	24,00%	13	52,00%	355	2.130	27.221.400
[370,400)	7	28,00%	20	80,00%	385	2.695	1.037.575
[400,430)	4	16,00%	24	96,00%	415	1.660	688.900
>=430	1	4,00%	25	100,00%	445	445	198.025
Total	25	100%				9.145	29.847.475

Apliquemos a fórmula obtida da definição acima, levando em conta as frequências

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N} = \frac{\sum_{i=1}^N (x_i)^2}{N} - (\bar{x})^2 = \frac{29.847.475}{25} - \left(\frac{9.145}{25}\right)^2 = 1.193.899 - 133.809,64 = 1.060.089,36 =$$

2.2.4 – Desvio-padrão

Para entendermos o procedimento para o cálculo do desvio-padrão, é interessante percebermos que, no cálculo da variância, tal como vimos no tópico anterior, cometemos um “erro técnico” que será corrigido pelo desvio-padrão, ou seja, no momento em que elevamos ao quadrado as dispersões (diferenças) de cada elemento em relação à média, automaticamente alteramos a **unidade** de trabalho. Por exemplo: se estivermos trabalhando com a coleta das alturas, em metro, das pessoas de uma determinada comunidade, a unidade da variância encontrada será o m^2 (metro quadrado), que representa áreas. E é aí que entra o desvio-padrão, ou seja, extraindo a raiz quadrada da variância.

$$\text{Desvio – padrão } (\sigma) = \sqrt{\text{Variância}}$$

2.2.4.1 – Dados Não Agrupados

Exercícios de Aplicação

1. No exemplo do item anterior a variância encontrada foi 345,57, temos que o desvio-padrão foi de $\sqrt{345,57} = 18,58$

Observação: O uso do Desvio Médio pode causar dificuldades quando comparamos conjuntos de dados com números diferentes de observações:

2. Em $A = \{3,4,5,6,7\}$ temos o Desvio Médio (DM) como $6/5 = 1,2$ e $\sigma^2 = 10/5 = 2$

3. Em $D = \{3,5,5,7\}$ temos o Desvio Médio (DM) = 1,0 e $\sigma^2 = 2$

Assim, podemos dizer que, segundo o Desvio Médio, o grupo D é mais homogêneo (tem menor dispersão) do que A , enquanto que ambos têm a mesma homogeneidade segundo a variância. O desvio médio possui pequena utilização em estatística e em geral vale 0,8 vezes o desvio padrão

2.2.4.2 – Dados Agrupados

Dados Agrupados Sem Intervalo de Classe

Fazemos como no caso da variância e extraímos a raiz quadrada.

Dados Agrupados Com Intervalo de Classe

Fazemos como no caso da variância e extraímos a raiz quadrada.

2.2.5 – Momentos de uma distribuição de frequências

2.2.5.1 – Dados Não Agrupados

Definimos o **momento de ordem t** de um conjunto de dados como:

$$M_t = \frac{\sum_{i=1}^N (x_i)^t}{N}$$

Definimos o momento de ordem t **centrado em relação a uma constante a** como

$$M_t = \frac{\sum_{i=1}^N (x_i - a)^t}{N}$$

Especial interesse tem o caso do momento centrado em relação a \bar{x} , dado por:

$$m_t = \frac{\sum_{i=1}^N (x_i - \bar{x})^t}{N}$$

2.2.5.2 – Dados Agrupados

Dados Agrupados Sem Intervalo de Classe

Conforme já vimos nos casos da média e da variância, as expressões precedentes podem ser reescritas levando-se em consideração as frequências dos diferentes valores existentes. Temos então respectivamente,

$$M_t = \frac{\sum_{i=1}^N (x_i)^t \cdot f_i}{N}$$

$$M_t = \frac{\sum_{i=1}^N (x_i - a)^t \cdot f_i}{N}$$

$$m_t = \frac{\sum_{i=1}^N (x_i - \bar{x})^t \cdot f_i}{N}$$

É fácil ver que $M_1 = \bar{x}$;
 $m_1 = 0$; $m_2 = \sigma^2$.

Dados Agrupados Com Intervalo de Classe

Aqui, convencionamos que *todos os valores incluídos em um determinado **intervalo de classe** coincidem com o seu ponto médio x_i* , e determinamos a variância da mesma maneira que fizemos com os dados agrupados sem intervalo de classe.

2.2.6 – Coeficiente de variação (CV)

O coeficiente de variação exprime a variabilidade em termos relativos. É uma medida adimensional e sua grande utilidade é permitir a comparação das variabilidades em diferentes conjuntos de dados.

$$CV = \frac{\sigma}{\bar{x}}$$

Exercícios de Aplicação

1. Testes de resistência à tração, aplicados a dois tipos diferentes de aço:

	\bar{x} (kg/mm ²)	σ (kg/mm ²)
Tipo I	27,45	2,0
Tipo II	147,00	17,25

$$CV_I = 2/27,45 = 7,29\%$$

$$CV_{II} = 17,25/145 = 11,73\%$$

Assim, apesar do Tipo I ser menos resistente, é ele mais estável, mais consistente.

O uso do coeficiente de variação pode ser pensado considerando a questão: Um desvio padrão de 10 se a média é 10.000 é bem diferente se a média é 100!

2. Numa empresa o salário médio dos funcionários do sexo masculino é de R\$ 4.000,00, com desvio padrão de R\$ 1.500,00, os funcionários do sexo feminino é em média de R\$ 3.000,00, com desvio padrão de R\$ 1.200,00. Quais os coeficientes de variação.

Sexo masculino: $CV = 1500/4000 = 0,375$ ou 37,5%

Sexo feminino: $CV = 1200/3000 = 0,40$ ou 40%

Logo, podemos concluir que o salário das mulheres apresenta maior dispersão relativa que a dos homens.

Exercícios Propostos

1. Mostre que:

a. $\sum_{i=1}^n (x_i - \bar{x}) = 0$

b. $\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum x_i)^2}{n}$

2. O conjunto abaixo representa as notas do exame final de uma determinada turma:

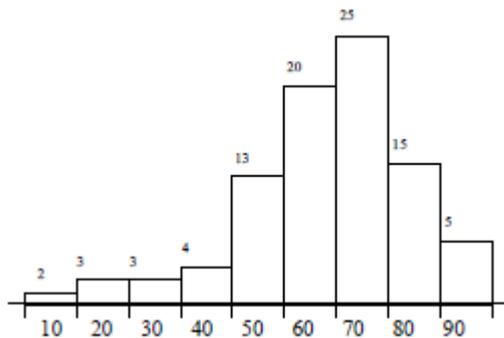
54	50	15	65	70	42	77	33	55	92
61	50	35	54	75	71	64	10	51	86
70	66	60	60	71	64	63	77	75	70
81	48	34	73	65	62	66	75	60	85
64	57	74	60	63	85	47	67	79	37
66	45	58	67	71	53	23	61	66	88
58	48	73	76	81	83	62	75	69	68
66	71	66	67	50	76	60	45	61	74

a. Construir uma distribuição de freqüências, adotando um intervalo de classe conveniente, o histograma e o polígono de freqüências. (sugerem-se classes de tamanho 10, a partir de 10).

b. Calcular a média, o desvio padrão, a mediana, o 1º quartil e o 65º percentil.

3. Resolver (b) graficamente no que couber, com auxílio da ogiva de freqüência acumuladas.

4. Dado o histograma abaixo, calcular a média, a variância, a mediana e o 3º quartil. Sabe-se que o número total de observações é 90.



5. Dado o conjunto de observações

120 107 95 118 150 130 132 109 136

a. Determinar os quartis.

b. Calcular a média

6. Construir dois conjuntos de dados de mesma amplitude mas com variabilidade diferentes.

7. Por engano, um professor omitiu uma nota no conjunto de notas de 10 alunos. Se as nove notas restantes são 48, 71, 79, 95, 45, 57, 75, 83, 97 e a média das 10 notas é 72, qual o valor da nota omitida.

8. Em um certo ano, uma universidade pagou a cada um de seus 45 instrutores um salário médio mensal de R\$ 1.500,00, a cada um de seus 67 assistentes R\$ 2.000,00, a cada um de seus 58 adjuntos R\$ 2.600,00 e a cada um de seus 32 titulares R\$ 3.100,00. Qual o salário médio mensal dos 202 docentes?

9. São os seguintes os números de alunos e respectivos QI médios em três estabelecimentos de ensino:

Estabelecimento	Nº de alunos	QI Médio
A	790	104
B	155	110
C	530	106

Qual é o QI médio global dos três estabelecimentos? Que tipo de média foi utilizada?

10. Alega-se que, para amostras de tamanho $n = 10$, a amplitude amostral deve ser aproximadamente três vezes o desvio padrão. Verifique a alegação com relação aos seguintes dados que representam a velocidade de 10 carros cronometrados em um posto de controle:

97 80 88 62 93 105 74 86 83 95

11. Calcule \bar{x} , σ^2 e σ para os dados abaixo, onde os valores de x são pontos médios de intervalos.

x_i 15 25 35 45 55 65 75 85 95 105

f_i 1 5 12 18 21 19 10 7 6 1

12. Tempos, em minutos, de espera numa fila de ônibus, durante 13 dias, de um cidadão que se dirige diariamente ao seu emprego:

15 10 2 17 6 8 3 10 2 9 5 9 13

Calcular a média, a mediana e a moda

13. Dada a seguinte distribuição de idades dos membros de uma sociedade,

Idade	x_i	f_i	Idade	x_i	f_i
15 --- 20	17	16	35 --- 40	37	17
20 --- 25	22	35	40 --- 45	42	8
25 --- 30	27	44	45 --- 50	47	2
30 --- 35	32	27	50 --- 55	52	1

Calcular:

- \bar{x} e σ
- \tilde{x} , Q_1 e Q_3
- O quartil médio dado por $\frac{1}{2}(Q_1 + Q_2)$
- O intervalo interquartil $Q_3 - Q_1$.
- O intervalo semi-interquartil dado por $\frac{1}{2}(Q_3 - Q_1)$

Nota: o quartil médio é às vezes usado como medida de tendência central em lugar da média. O intervalo interquartil e o intervalo semi-interquartil são usados como medidas de variabilidade.

14. A tabela abaixo mostra o tempo gasto por empregados numa determinada operação em uma fábrica:

Tempo (min)	Frequência
10 --- 15	5
15 --- 20	57
20 --- 25	42
25 --- 30	28
30 --- 40	18

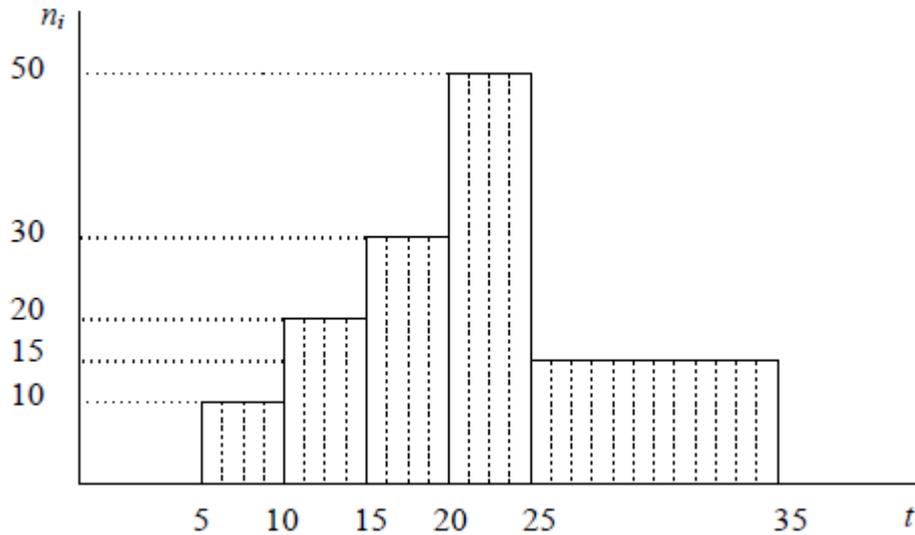
- Esboce o histograma correspondente
- Calcule a média e o desvio padrão
- O presidente decide separar os funcionários mais rápidos (com tempo inferior a um desvio padrão abaixo da média) para receberem promoção. Qual a porcentagem desses funcionários?

15. Em testes de resistência à tração aplicados a um tipo de aço, obteve-se um coeficiente de variação de 8%. O escore padronizado² obtido para uma resistência de 28 kg/mm² foi de -0,4.

² Define-se o escore padronizado (muito usado na distribuição normal - 3º bimestre) como:

$$z_i = \frac{x_i - \bar{x}}{\sigma}$$

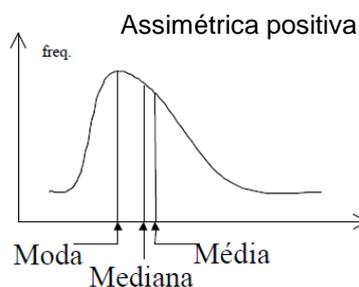
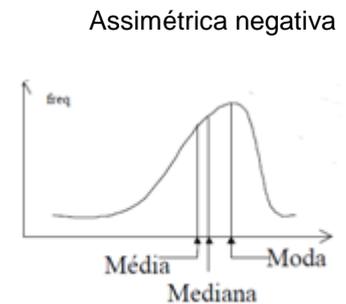
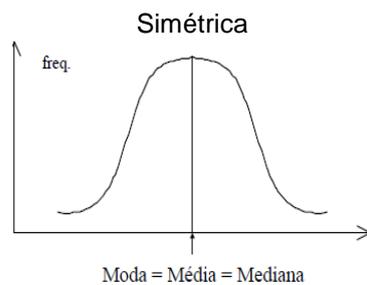
- Qual a resistência média?
 - Qual o desvio padrão?
 - O que significa o resultado negativo do escore padronizado?
16. O histograma abaixo descreve o tempo (em min) gasto por funcionários de uma fábrica em uma determinada operação:



- Construa a distribuição para o histograma;
- Calcule a média e o desvio padrão;
- O presidente decide demitir os funcionários mais lentos (com tempo superior a meio desvio padrão acima da média). Qual a porcentagem desses funcionários?

2.3 – Medidas de Assimetria

É o grau de deformação de uma distribuição em relação ao eixo de simetria. Podemos observar os tipos de assimetria abaixo:



Existem vários coeficientes com o objetivo de quantificar tais assimetrias. Estudaremos dois destes coeficientes que veremos a seguir:

COEFICIENTE DE PEARSON

O coeficiente de Pearson é apresentado pela seguinte fórmula:

$$As = \frac{\bar{x} - Mo}{\sigma}$$

Classificação do coeficiente de Pearson:

$As = 0$	Distribuição Simétrica
$0 < As < 1$	Distribuição Assimétrica Positiva Fraca
$As \geq 1$	Distribuição Assimétrica Positiva Forte
$-1 < As < 0$	Distribuição Assimétrica Negativa Fraca
$As \leq -1$	Distribuição Assimétrica Negativa Forte

COEFICIENTE DE BOWLEY

O coeficiente de Bowley é apresentado pela seguinte fórmula:

$$As = \frac{Q_3 + Q_1 - 2Md}{Q_3 - Q_1}$$

Classificação do coeficiente de Bowley:

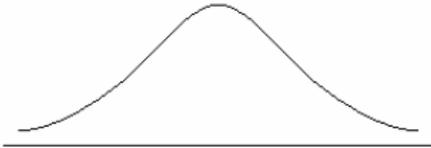
$As = 0$	Distribuição Simétrica
$0 < As \leq 0,1$	Distribuição Assimétrica Positiva Fraca
$0,1 < As < 0,3$	Distribuição Assimétrica Positiva Moderada
$0,3 \leq As \leq 1$	Distribuição Assimétrica Positiva Forte
$-0,1 \leq As < 0$	Distribuição Assimétrica Negativa Fraca
$-0,3 < As < -0,1$	Distribuição Assimétrica Negativa Moderada
$-1 \leq As \leq -0,3$	Distribuição Assimétrica Negativa Forte

2.4 – Medidas de Curtose

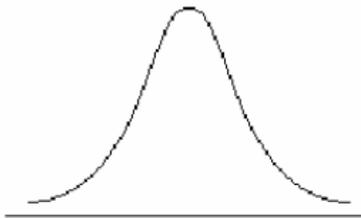
Entende-se por curtose o grau de achatamento de uma distribuição. Podemos ter:



⇒ CURVA PLATICÚRTICA



⇒ CURVA MESOCÚRTICA



⇒ CURVA LEPTOCÚRTICA

Para medir o grau de curtose utilizaremos o coeficiente:

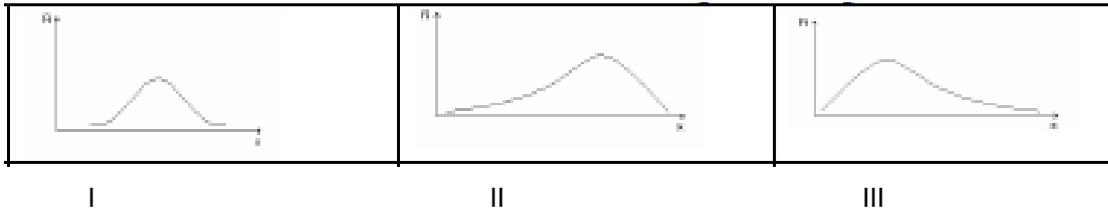
$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

Classificação do coeficiente de Curtose:

K = 0,263	CURVA MESOCÚRTICA
K > 0,263	CURVA PLATICÚRTICA
K < 0,263	CURVA LEPTOCURTICA

Exercícios Propostos

1. Analisando as curvas abaixo marque a resposta correta.



- a. A curva I é simétrica - $\bar{x} > med > mo$;
 b. A curva II é assimétrica positiva - $mo > \sigma^2 > \bar{x}$;
 c. A curva I é simétrica $\bar{x} = med = mo$;
 d. A curva III é simétrica positiva $\bar{x} = med = mo$.
2. Para as distribuições abaixo foram calculados:

Distrib. A		Distrib. B		Distrib. C	
Classes	Fi	Classes	Fi	Classes	Fi
02 - 06	6	02 - 06	6	02 - 06	6
06 - 10	12	06 - 10	12	06 - 10	30
10 - 14	24	10 - 14	24	10 - 14	24
14 - 18	12	14 - 18	30	14 - 18	12
18 - 22	6	18 - 22	6	18 - 22	6
$\bar{x} = 12\text{Kg}$		$\bar{x} = 12,9\text{Kg}$		$\bar{x} = 11,1\text{Kg}$	
Med = 12Kg		Med = 13,5Kg		Med = 10,5Kg	
Mo = 12Kg		Mo = 16Kg		Mo = 8Kg	
S = 4,42Kg		S = 4,20Kg		S = 4,20Kg	

Marque a alternativa correta:

- a. A distribuição I é assimétrica negativa;
 b. A distribuição II é assimétrica positiva;
 c. A distribuição III é assimétrica negativa moderada;
 d. A distribuição I é simétrica.
3. Sabe-se que uma distribuição apresentou as seguintes medidas:
 $Q_1 = 24,4 \text{ cm}$ $Q_3 = 41,2 \text{ cm}$
 $P_{10} = 20,2 \text{ cm}$ $P_{90} = 49,5 \text{ cm}$
 Com tais medidas a curtose é:
- a. () Leptocúrtica c. () Mesocúrtica
 b. () Platicúrtica d. () Assimétrica

2.5 – Medidas Estatísticas no Excel

2.5.1 – Dados não agrupados

Abra aquela pasta onde você fez, anteriormente, a distribuição de frequências com intervalo de classe.

	A	B	C	D	E	F	G
3	Dados Brutos	ROL		Funções			
4	280	280	Maior Valor	430	=MÁXIMO(B4:B28)		
5	365	280	Menor Valor	280	=MÍNIMO(B4:B28)		
6	305	305	Tamanho da Amostra	25	=CONT.VALORES(A4:A28)		
7	280	310	Número de classes	5	=RAIZ(D6)		
8	320	320	Amplitude Total	150	=MÁXIMO(B4:B28)-MÍNIMO(B4:B28)		
9	375	330	Amplitude das classes	30	=D8/D7		
10	330	330					
11	380	340	Classe	Limite Inferior	Limite Superior	Frequências Absolutas	
12	310	341	280 - 310	280	309,99	3	
13	400	355	310 - 340	310	339,99	4	
14	340	360	340 - 370	340	369,99	6	
15	371	365	370 - 400	370	399,99	7	
16	330	369	400 -	400	430	5	
17	390	370					
18	341	370					
19	400	370					
20	369	371					
21	370	375					
22	355	380					
23	401	390					
24	370	400					
25	420	400					
26	360	401					
27	430	420					
28	370	430					

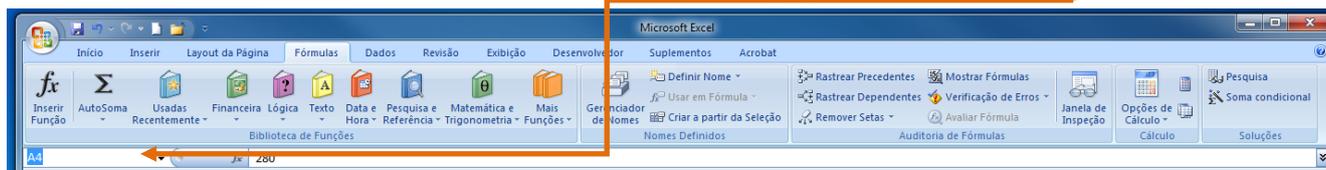
Marque os intervalos: A1:F28. Pressione a tecla CTRL + C.

Abra uma nova pasta de trabalho e a nomeie de MEDIDASestatísticas_DadosNaoAgrupados

Na célula A1, digite o título: Calculando Estatísticas de Dados NÃO Agrupados com Funções Embutidas do Excel

Marque o intervalo (nesta nova pasta) C11:F16 e delete-o.

Marque o intervalo A4:A28 e o nomeie como: Dados_Brutos . Para fazer isto, clique na caixa de nomes:



Digite: Dados_Brutos.

Agora toda vez que você quiser se referir a este intervalo de células, pode informar Dados_Brutos ao invés de A4:A28

No intervalo de células C12:C17, digite: Média Aritmética, Média Geométrica, Média Harmônica, Mediana, Moda, Percentil-30, Quartil(3).

No intervalo de células C20:C25, digite: Desvio Médio, Desvio Padrão, Variância da Amostra, Variância da População, Curtose, Distorção³.

A coisa ficou assim:

³ Isto significa o grau de assimetria.

	A	B	C	D	E	F
1	Calculando Estatísticas de Dados NÃO Agrupados com Funções Embutidas do Excel					
2						
3	Dados Brutos	ROL		Funções		
4	280	280	Maior Valor	430	=MÁXIMO(B4:B28)	
5	365	280	Menor Valor	280	=MÍNIMO(B4:B28)	
6	305	305	Tamanho da Amostra	25	=CONT.VALORES(A4:A28)	
7	280	310	Número de classes	5	=RAIZ(D6)	
8	320	320	Amplitude Total	140	=MÁXIMO(B4:B28)-MÍNIMO(B4:B28)	
9	375	330	Amplitude das classes	28	=D8/D7	
10	330	330				
11	380	340				
12	310	341	Média Aritmética			
13	400	355	Média Geométrica			
14	340	360	Média Harmônica			
15	371	365	Mediana			
16	330	369	Moda			
17	390	370	Percentil-30			
18	341	370	Quartil (3)			
19	400	370				
20	369	371	Desvio Médio			
21	370	375	Desvio Padrão			
22	355	380	Variância da Amostra			
23	401	390	Variância da População			
24	370	400	Curtose			
25	420	400	Distorsão			
26	360	401				
27	430	420				
28	370	430				
29						

Introduza as seguintes fórmulas nas células correspondentes:

Célula	Fórmula
D12	=MÉDIA(Dados_Brutos)
D13	=MÉDIA.GEOMÉTRICA(Dados_Brutos)
D14	=MÉDIA.HARMÔNICA(Dados_Brutos)
D15	=MED(Dados_Brutos)
D16	=MODO(Dados_Brutos)
D17	=PERCENTIL(Dados_Brutos;0,3)
D18	=QUARTIL(Dados_Brutos;3)
D20	=DESV.MÉDIO(Dados_Brutos)
D21	=DESVPAD(Dados_Brutos)
D22	=VAR(Dados_Brutos)
D23	=VARP(Dados_Brutos)
D24	=CURT(Dados_Brutos)
D25	=DISTRORÇÃO(Dados_Brutos)

Agora estamos assim:

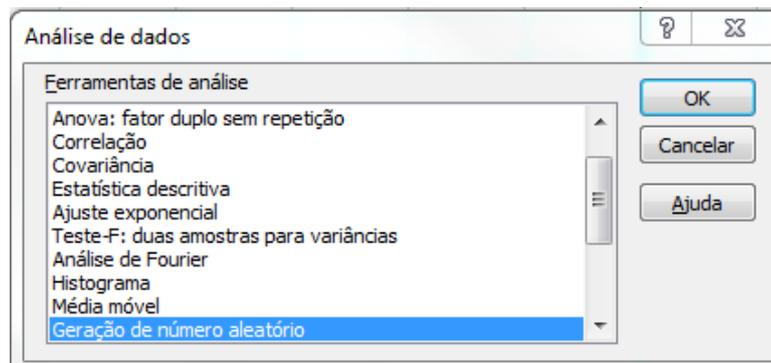
	A	B	C	D	E	F
1	Calculando Estatísticas de Dados NÃO Agrupados com Funções Embutidas do Excel					
2						
3	Dados Brutos	ROL		Funções		
4	280	280	Maior Valor	430	=MÁXIMO(B4:B28)	
5	365	280	Menor Valor	280	=MÍNIMO(B4:B28)	
6	305	305	Tamanho da Amostra	25	=CONT.VALORES(A4:A28)	
7	280	310	Número de classes	5	=RAIZ(D6)	
8	320	320	Amplitude Total	150	=MÁXIMO(B4:B28)-MÍNIMO(B4:B28)	
9	375	330	Amplitude das classes	30	=D8/D7	
10	330	330				
11	380	340				
12	310	341	Média Aritmética	358,480	=MÉDIA(Dados_Brutos)	
13	400	355	Média Geométrica	356,290	=MÉDIA.GEOMÉTRICA(Dados_Brutos)	
14	340	360	Média Harmônica	354,021	=MÉDIA.HARMÔNICA(Dados_Brutos)	
15	371	365	Mediana	369,000	=MED(Dados_Brutos)	
16	330	369	Moda	370,000	=MODO(Dados_Brutos)	
17	390	370	Percentil-30	340,200	=PERCENTIL(Dados_Brutos;0,3)	
18	341	370	Quartil (3)	380,000	=QUARTIL(Dados_Brutos;3)	
19	400	370				
20	369	371	Desvio Médio	31,504	=DESV.MÉDIO(Dados_Brutos)	
21	370	375	Desvio Padrão	39,73151562	=DESVPAD(Dados_Brutos)	
22	355	380	Variância da Amostra	1578,593333	=VAR(Dados_Brutos)	
23	401	390	Variância da População	1515,4496	=VARP(Dados_Brutos)	
24	370	400	Curtose	-0,314328358	=CURT(Dados_Brutos)	
25	420	400	Distorsão	-0,343739072	=DISTRORÇÃO(Dados_Brutos)	
26	360	401				
27	430	420				
28	370	430				

Beleza

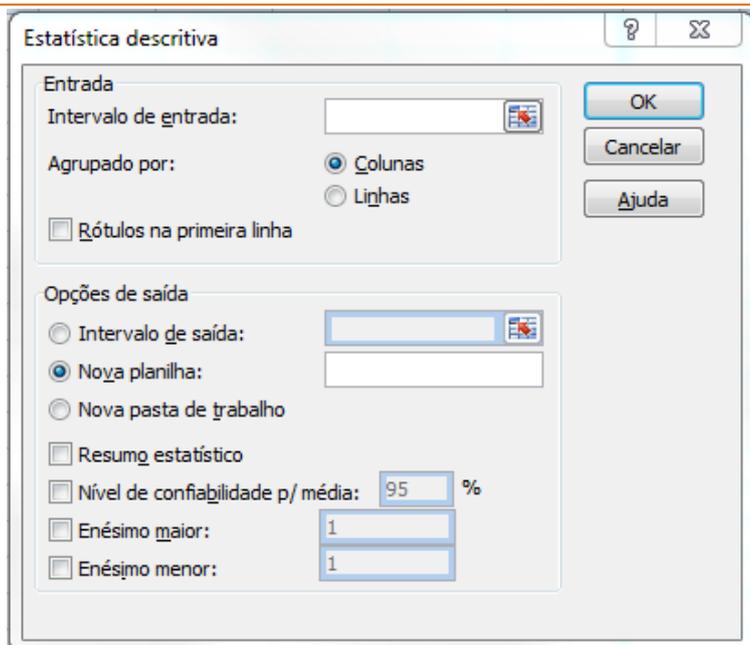
Quanta coisa este encantador Excel é capaz de fazer por nós. Faça, então, muita coisa pelo seu irmão em troca desta benção!.

Mais bonito ainda é que Bill Gates, ajudando o próximo, criou uma ferramenta de análise de dados, chamada Estatística Descritiva que faz tudo isto “com dois palitos”. Vamos aprendê-la.

Clique na célula C30 e na guia Dados, grupo de ferramentas Análise, selecione e pressione Análise de Dados e vai aparecer a janela **Análise de dados**:



Selecione Estatística Descritiva como a ferramenta de análise. Vai aparecer a janela **Estatística descritiva**:



Na caixa Intervalo de entrada: digite: Dados_brutos.

Marque a caixa de verificação Rótulos na primeira linha.

Na seção Opções de saída, selecione o botão Intervalo de saída e na caixa insira C30.

Marque Resumo Estatístico.

A seguir pressione OK.

	A	B	C	D	E
29					
30			280		
31					
32			Média	361,75	
33			Erro padrão	7,550614237	
34			Mediana	369,5	
35			Modo	370	
36			Desvio padrão	36,99030425	
37			Variância da amostra	1368,282609	
38			Curtose	-0,167032944	
39			Assimetria	-0,283223911	
40			Intervalo	150	
41			Mínimo	280	
42			Máximo	430	
43			Soma	8682	
44			Contagem	24	
45			Nível de confiança(99,0%)	21,19710874	
46					

2.5.2 – Dados agrupados sem intervalo de classe

Agora vá ao intervalo de células G3:J3 e introduza em cada célula os títulos: ROL, x_i , x_i agrupado e freqüência f_i . Copie o intervalo de células B4:B28 para G4:G28.

	G	H	I	J
3	ROL	xi	xi agrupado	frequência fi
4	280			
5	280			
6	305			
7	310			
8	320			
9	330			
10	330			
11	340			
12	341			
13	355			
14	360			
15	365			
16	369			
17	370			
18	370			
19	370			
20	371			
21	375			
22	380			
23	390			
24	400			
25	400			
26	401			
27	420			
28	430			
29				

Na célula H4, digite: =G4.

Na célula H5, introduza a fórmula: =SE(G5:\$G\$28<>G4;G5;"")

Arraste pela alça esta fórmula até a célula H28. E a coisa fica assim:

	G	H	I	J
3	ROL	xi	xi agrupado	frequência fi
4	280	280		
5	280			
6	305	305		
7	310	310		
8	320	320		
9	330	330		
10	330			
11	340	340		
12	341	341		
13	355	355		
14	360	360		
15	365	365		
16	369	369		
17	370	370		
18	370			
19	370			
20	371	371		
21	375	375		
22	380	380		
23	390	390		
24	400	400		
25	400			
26	401	401		
27	420	420		
28	430	430		

Se você ainda não descobriu como juntar os dados da coluna H (excluir as células vazias dos dados) faça isto manualmente no intervalo I4:I23.

Selecione o intervalo J4:J24 e introduza a fórmula matricial: **=FREQUÊNCIA(G4:G28;I4:I23)**. Não pressione ENTER ainda. Lembra-se!

Pressione F2 e a seguir, e ao mesmo tempo, CTRL+SHIFT+ENTER.

Assim você construiu uma **tabela de frequências dos dados agrupados sem intervalo de classe**.

	G	H	I	J
3	ROL	xi	xi agrupado	frequência fi
4	280	280	280	2
5	280		305	1
6	305	305	310	1
7	310	310	320	1
8	320	320	330	2
9	330	330	340	1
10	330		341	1
11	340	340	355	1
12	341	341	360	1
13	355	355	365	1
14	360	360	369	1
15	365	365	370	3
16	369	369	371	1
17	370	370	375	1
18	370		380	1
19	370		390	1
20	371	371	400	2
21	375	375	401	1
22	380	380	420	1
23	390	390	430	1
24	400	400		0
25	400			
26	401	401		
27	420	420		
28	430	430		

Completa a tabela de distribuição de frequências para torná-la assim:

	G	H	I	J	K	L	M
3	ROL	xi	xi agrupado	frequência fi	Freq. Rel. fri	Freq. Acum FAi	Freq. Acum FARi
4	280	280	280	2	0,08	2	0,08
5	280		305	1	0,04	3	0,12
6	305	305	310	1	0,04	4	0,16
7	310	310	320	1	0,04	5	0,2
8	320	320	330	2	0,08	7	0,28
9	330	330	340	1	0,04	8	0,32
10	330		341	1	0,04	9	0,36
11	340	340	355	1	0,04	10	0,4
12	341	341	360	1	0,04	11	0,44
13	355	355	365	1	0,04	12	0,48
14	360	360	369	1	0,04	13	0,52
15	365	365	370	3	0,12	16	0,64
16	369	369	371	1	0,04	17	0,68
17	370	370	375	1	0,04	18	0,72
18	370		380	1	0,04	19	0,76
19	370		390	1	0,04	20	0,8
20	371	371	400	2	0,08	22	0,88
21	375	375	401	1	0,04	23	0,92
22	380	380	420	1	0,04	24	0,96
23	390	390	430	1	0,04	25	1
24	400	400		0	0		
25	400		Total	25	1		
26	401	401					
27	420	420					
28	430	430					

Média Aritmética Ponderada

Neste caso, dados agrupados sem intervalo de classe, como as frequências f_i são números indicadores da intensidade de cada valor da variável, elas funcionam como fatores de ponderação, o que nos leva à **Média Aritmética Ponderada**.

Para o seu cálculo no Excel lançamos mão da função embutida SOMARPRODUTO cuja sintaxe é:

SOMARPRODUTO(matriz1;matriz2;matriz3;...)

Matriz1, matriz2, matriz3, ... são matrizes de 2 a 255 cujos componentes se deseja multiplicar e depois somar.

Vá à célula P5 e digite: Média

Na célula Q5, introduza a fórmula: **=SOMARPRODUTO(I4:I23;J4:J23)/SOMA(J4:J23)**

O resultado será **358,48**.

Mediana

A mediana neste caso, sem intervalos de classe, é obtida identificando a frequência acumulada imediatamente superior à metade da soma das frequências. A mediana será o menor valor da variável que supera aquele valor obtido.

Teremos então de procurar na coluna L, frequência acumulada FA_i, o valor imediatamente superior a $J25/2$ que, no nosso exemplo é **13**.

Como procurar valores numa matriz de dados com o Excel?

Teremos que aprender a função PROCV que se encontra na categoria PROCURA e REFERÊNCIA.

Esta função localiza um valor na primeira coluna de uma matriz de tabela e retorna um valor na mesma de outra coluna desta matriz.

A letra V em PROCV significa que a procura será na vertical (coluna)

A sintaxe da função PROCV é:

PROCV(valor_procurado;matriz_tabela;num_índice_coluna;procurar_intervalo)

Os argumentos em negrito são obrigatórios.

Valor_procurado: é o valor a ser procurado na primeira coluna da matriz-tabela. Pode ser um valor ou uma referência. Se o valor_procurado for menor que o menor valor da primeira coluna da matriz-tabela, o PROCV retornará o valor de erro #N/D.

Matriz_tabela: são duas ou mais colunas de dados. Use uma referência para um intervalo ou um nome de intervalo. Os valores na primeira coluna desta matriz são os valores a serem procurados e que podem ser texto, números ou valores lógicos. Textos em maiúsculas e minúsculas são equivalentes.

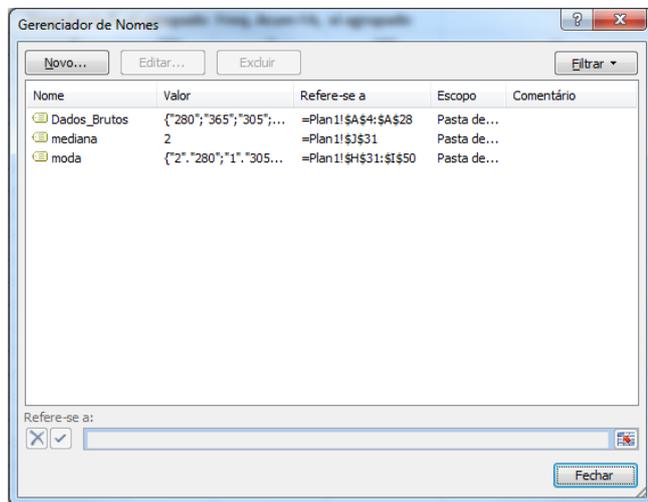
Núm_índice_coluna: é o número da coluna da matriz-tabela a partir do qual o valor correspondente deve ser retornado. Um núm_índice_coluna igual a 1 retornará o valor da primeira coluna da matriz-tabela; um núm_índice_coluna igual a 2 retornará o valor da segunda coluna da matriz-tabela, e assim por diante.

Procurar_intervalo: é um valor lógico que especifica se você quer que PROCV localize uma correspondência exata ou aproximada.

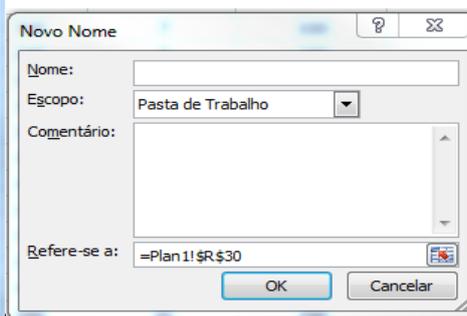
A nossa tabela de distribuição de frequências não está apropriada para a utilização da função PROCV. Usando CTRL+C e CTRL+V, copie e cole as colunas *frequência f_i* , *x_i agrupado*, *Freq. Acum FA_i* , e *x_i agrupado* nas células, H30, I30, J30, L30.

	H	I	J	K
30	frequência f_i	x_i agrupado	Freq. Acum FA_i	x_i agrupado
31	2	280	2	280
32	1	305	3	305
33	1	310	4	310
34	1	320	5	320
35	2	330	7	330
36	1	340	8	340
37	1	341	9	341
38	1	355	10	355
39	1	360	11	360
40	1	365	12	365
41	1	369	13	369
42	3	370	16	370
43	1	371	17	371
44	1	375	18	375
45	1	380	19	380
46	1	390	20	390
47	2	400	22	400
48	1	401	23	401
49	1	420	24	420
50	1	430	25	430

Crie agora os nomes, Mediana e Moda, para o intervalo de células: H31:I50 e J31:K50. Para fazer isto vá à guia Fórmulas, no grupo Nomes Definidos, clique em Gerenciador de Nomes. Aparecerá a janela:



Clique em Novo... para aparecer a janela Novo Nome.



Em Nome digite: Mediana e na caixa Refere-se a: digite o intervalo de células H31:I50.

Repita os passos anteriores e crie o Nome Moda para o intervalo de células J31:K50.

Usamos a função CORRESP em vez de PROCV quando precisamos da posição de um item em um intervalo ao invés do item propriamente dito

A função CORRESP retorna a posição relativa de um item em uma matriz que coincide com um valor especificado em uma ordem específica. U

Sintaxe da função CORRESP:

CORRESP(valor_procurado;matriz_procurada;tipo_correspondência)

Moda

Aqui basta procurarmos a máxima frequência absoluta f_i . Para fazer isso no Excel, digite em P7: Moda.

Em Q7, introduza a fórmula: =PROCV(MÁXIMO(J4:J23);moda;2;FALSO)

O resultado será: **370**.

Percentil

O percentil para dados agrupados, sem intervalo de classe, pode ser obtido por interpolação linear (regra de três simples).

O P_{30} deixa (30% de 25 dados = 7,5 \Rightarrow 8 dados) nele ou abaixo dele e 70% dos dados acima dele. Como fizemos com a mediana, procuremos a $FA_i = 8$ e o x_i correspondente.

Quartil

Desvio Médio

Para encontrarmos o desvio médio precisamos criar mais uma coluna de dados na nossa tabela de frequências.

Vá à Célula N3 e digite: $|(x_i - x_{med})|$.

Em N4, introduza a fórmula: =ABS(I4-\$Q\$5)

Arraste a fórmula até a N23.

Em P11 digite: Desvio Médio

Em Q11, introduza a fórmula: =SOMARPRODUTO(N4:N23;J4:J23)/\$J\$25

O resultado será **31,504**.

Variância

Para encontrarmos a variância precisamos criar ainda outra coluna de dados na nossa tabela de frequências.

Vá à Célula O3 e digite: $(x_i - x_{med})^2$.

Em O4, introduza a fórmula: =N4^2

Arraste a fórmula até a O23.

Em P13 digite: Variância da amostra

Em P14 digite: Variância da população

Em Q13, introduza a fórmula: =SOMARPRODUTO(O4:O23;J4:J23)/(\$J\$25-1). O resultado será **1578,593333**.

Em Q14, introduza a fórmula: =SOMARPRODUTO(O4:O23;J4:J23)/\$J\$25. O resultado será **1515,450**.

Desvio Padrão

O desvio padrão é a raiz quadrada da variância, logo em P12 digite: Desvio Padrão.

Em Q12, introduza a fórmula: =RAIZ(SOMARPRODUTO(O4:O23;J4:J23)/(\$J\$25-1)). O resultado será **39,73151562**.

Curtose

Assimetria

	G	H	I	J	K	L	M	N	O	P	Q	R	S
3	ROL	x_i	x_i agrupado	frequência f_i	Freq. Rel. f_{ri}	Freq. Acum FA_i	Freq. Acum FAR_i	$ (x_i - x_{med}) $	$(x_i - x_{med})^2$				
4	280	280	280	2	0,08	2	0,08	78,48	6159,1104				
5	280		305	1	0,04	3	0,12	53,48	2860,1104	Média	358,48	<--=SOMARPRODUTO	
6	305	305	310	1	0,04	4	0,16	48,48	2350,3104	Mediana	369	<--=PROCV(CORRESP(
7	310	310	320	1	0,04	5	0,2	38,48	1480,7104	Moda	370	<--=PROCV(MÁXIMO(
8	320	320	330	2	0,08	7	0,28	28,48	811,1104	Percentil-30	340	<--=PROCV(CORRESP(
9	330	330	340	1	0,04	8	0,32	18,48	341,5104	Quartil(3)	380	<--=PROCV(CORRESP(
10	330		341	1	0,04	9	0,36	17,48	305,5504				
11	340	340	355	1	0,04	10	0,4	3,48	12,1104	Desvio Médio	31,504	<--=SOMARPRODUTO	
12	341	341	360	1	0,04	11	0,44	1,52	2,3104	Desvio Padrão	39,73151562	<--=RAIZ(SOMARPROD	
13	355	355	365	1	0,04	12	0,48	6,52	42,5104	Variância da Amostra	1578,593333	<--=SOMARPRODUTO	
14	360	360	369	1	0,04	13	0,52	10,52	110,6704	Variância da População	1515,450	<--=SOMARPRODUTO	
15	365	365	370	3	0,12	16	0,64	11,52	132,7104	Curtose			
16	369	369	371	1	0,04	17	0,68	12,52	156,7504	Assimetria			
17	370	370	375	1	0,04	18	0,72	16,52	272,9104				
18	370		380	1	0,04	19	0,76	21,52	463,1104				
19	370		390	1	0,04	20	0,8	31,52	993,5104				
20	371	371	400	2	0,08	22	0,88	41,52	1723,9104				
21	375	375	401	1	0,04	23	0,92	42,52	1807,9504				
22	380	380	420	1	0,04	24	0,96	61,52	3784,7104				
23	390	390	430	1	0,04	25	1	71,52	5115,1104				
24	400	400		0	0								
25	400		Total	25	1			616,08					
26	401	401											
27	420	420											
28	430	430											
29													
30		frequência f_i	x_i agrupado	Freq. Acum FA_i	x_i agrupado			i	x_i agrupado				
31		2	280	2	280			1	280				
32		1	305	3	305			2	305				
33		1	310	4	310			3	310				
34		1	320	5	320			4	320				
35		2	330	7	330			5	330				
36		1	340	8	340			6	340				
37		1	341	9	341			7	341				
38		1	355	10	355			8	355				
39		1	360	11	360			9	360				
40		1	365	12	365			10	365				
41		1	369	13	369			11	369				
42		3	370	16	370			12	370				
43		1	371	17	371			13	371				
44		1	375	18	375			14	375				
45		1	380	19	380			15	380				
46		1	390	20	390			16	390				
47		2	400	22	400			17	400				
48		1	401	23	401			18	401				
49		1	420	24	420			19	420				
50		1	430	25	430			20	430				

2.5.3 – Dados agrupados com intervalo de classe

Vamos aqui dar um *toque* profissional às nossas planilhas. Iremos construir uma planilha para o cálculo das estatísticas descritivas como faria um profissional consultor de empresas⁴.

Primeiramente, abra uma pasta e dê o nome MEDIDASestatísticas_DadosAgrupadosComIntervalo.

Nomeie a PLAN1 como Dados Brutos e a PLAN2 como Dados Agrupados.

Vá à célula Z3 e digite com a fonte na cor azul e em negrito: Dados

Marque o intervalo Z4:Z504 e adote o verde como a cor de preenchimento das células.

Preencha os dados no intervalo da coluna Z (25 dados).

Na célula Y4 introduza o número 1

Na célula Y5 introduza a fórmula: **=SE(Z5<>"";1+Y4;"")**. Arraste para baixo, pela alça, esta fórmula até esgotarem os dados. Desta forma você enumera os dados.

Na célula A1 digite com fonte em azul: Medidas Estatísticas Com Intervalo de Classe

No intervalo B3:G3 digite os cabeçalhos de coluna: i, Classes, Freq., FR, FA, FAR, x_i

A partir da célula I31 monte a tabelinha:

		J31	K31	L31	M31	N31
I31	Menor Valor		Amplitude das Classes		Maior Valor	
I32	Número de Classes		Amplitude Total		Tamanho da Amostra	
		J32	K32	L32	M32	N32

Na célula J31: =MÍNIMO(Dados)

Na célula J32: =RAIZ(N32)

Na célula L31: =L32/J32

Na célula L32: =N31-J31

Na célula N31: =MÁXIMO(Dados)

Na célula N32: =CONT.VALORES(Dados)

Até aqui a planilha ficou assim:

Medidas Estatísticas Com Intervalo de Classe																										
i	Classes	Freq.	FR	FA	FAR	x _i																				
1	<280	0	0,00%	0	0,00%	280																				
2	[280, 310)	3	12,00%	3	12,00%	295																				
3	[310, 340)	4	16,00%	7	28,00%	325																				
4	[340, 370)	6	24,00%	13	52,00%	355																				
5	[370, 400)	7	28,00%	20	80,00%	385																				
6	[400, 430)	4	16,00%	24	96,00%	415																				
7	>=430	1	4,00%	25	100,00%	445																				
8																										
9																										
10																										
11																										
12																										
13																										
14																										
15																										
16																										
17																										
18																										
19																										
20																										
21																										
22																										
23																										
24																										
25																										
Total		25	100,00%																							

Menor Valor		Amplitude das Classes		Maior Valor	
Número de Classes		Amplitude Total		Tamanho da Amostra	

Na célula B29, digite: Total

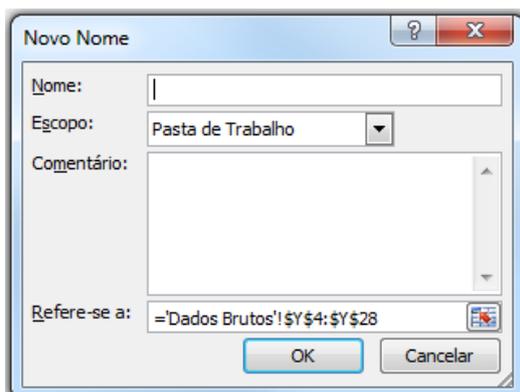
Na célula C29, digite a fórmula: **=SOMA(C4:C28)**

⁴ Procure ser um deles. Você poderá ganhar uma graninha extra e ajudar no seu salário de mestre. Que tal?

Vá à aba **Fórmulas** e no grupo de ferramentas **Nomes Definidos**, clique em **Definir Nome**.



Crie o nome **Dados** para as células do intervalo da coluna Z e na caixa **Refere-se a:** da janela **Novo Nome**, inserir a fórmula: **=DESLOC('Dados Brutos'!\$Z\$4;0;0;CONT.NÚM('Dados Brutos'!\$Z\$4:\$Z\$503);1)**



A função **DESLOC** retorna uma referência para um intervalo, que é um número especificado de linhas e colunas de uma célula ou intervalo de células. A referência retornada pode ser uma única célula ou intervalo de células. Você pode especificar o número de linhas e de colunas a serem retornadas.

A sua sintaxe é:

DESLOC(ref;lins;cols;altura;largura)

Ref: é a referência na qual você deseja basear o deslocamento. **Ref** deve ser uma referência a uma célula ou intervalo de células adjacentes; caso contrário, **DESLOC** retornará o valor de erro **#VALOR!**.

Lins: é o número de linhas, acima ou abaixo, a que se deseja que a célula superior esquerda se refira. Usar 5 como o argumento de linhas, especifica que a célula superior esquerda na referência está cinco linhas abaixo da referência. **Lins** podem ser positivas (que significa abaixo da referência inicial) ou negativas (acima da referência inicial).

Cols: é o número de colunas, à esquerda ou à direita, a que se deseja que a célula superior esquerda do resultado se refira. Usar 5 como o argumento de colunas, especifica que a célula superior esquerda na referência está cinco colunas à direita da referência. **Cols** pode ser positivo (que significa à direita da referência inicial) ou negativo (à esquerda da referência inicial).

Altura: é a altura, em número de linhas, que se deseja para a referência fornecida. **Altura** deve ser um número positivo.

Largura: é a largura, em número de colunas, que se deseja para a referência fornecida. **Largura** deve ser um número positivo.

Crie o nome **FrequenciaDeDados** para as células do intervalo da coluna C e na caixa **Refere-se a:** da janela **Novo Nome**, inserir a fórmula: **=DESLOC('Dados Brutos'!\$C\$4;0;1;CONT.NÚM('Dados Brutos'!\$C\$34:\$C\$58);1)**

Crie o nome **FreqRelDados** para as células do intervalo da coluna D e na caixa **Refere-se a:** da janela **Novo Nome**, inserir a fórmula: **=DESLOC('Dados Brutos'!\$D\$4;0;0;CONT.NÚM('Dados Brutos'!\$C\$34:\$C\$58);1)**

Crie o nome **ClassesDeDados** para as células do intervalo da coluna B e na caixa **Refere-se a:** da janela **Novo Nome**, inserir a fórmula: **=DESLOC('Dados Brutos'!\$B\$4;0;0;É.NÃO.DISP(('Dados Brutos'!\$C\$34:\$C\$58);1)**

Crie o nome **Bins** para as células do intervalo da coluna B e na caixa **Refere-se a:** da janela **Novo Nome**, inserir a fórmula: **=DESLOC('Dados Brutos'!\$B\$34;0;0;É.NÃO.DISP(('Dados Brutos'!\$C\$34:\$C\$58);1)**

Na célula **B34** introduza a fórmula: **=J31-0,01**

Na célula **B35** introduza a fórmula: **=SE(B34="";"";SE(ARREDONDAR.PARA.CIMA(B34;0)<\$N\$31;B34+\$L\$31;""))**. Arraste a fórmula para baixo, pela alça até a célula **B58**.

Na célula **B4** introduza a fórmula: **=" "&B34+0,01**.

Na célula B5 introduza a fórmula: `=SE(B34="";"";SE(B35<>"";"["&B34+0,01&" , "&B35+0,01&"]>"&B34+0,01))`. Arraste a fórmula para baixo, pela alça até a célula B28.

Marque o intervalo de células C34: C58 e introduza a fórmula: `=FREQUÊNCIA(Dados;Bins)`. Antes do ENTER, pressione F2 e depois, ao mesmo tempo, pressione CTRL+SHIFT+ENTER.

Na célula C4 introduza a fórmula: `=SE(É.NÃO.DISP(C34);"";C34)`. Arraste a fórmula para baixo, pela alça até a célula C28.

Na célula D4 introduza a fórmula: `=SE(C4<>"";C4/C29;"")`. Arraste a fórmula para baixo, pela alça até a célula D28.

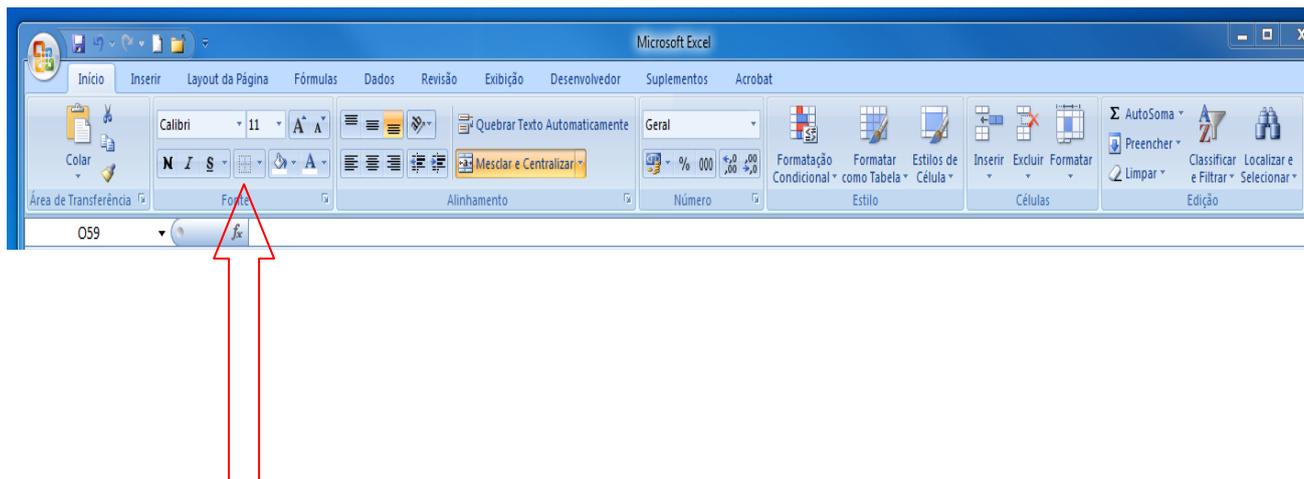
Na célula D29 introduza a fórmula: `=SOMA(D4:D28)`

Na célula E4 introduza a fórmula: `=SE(C4<>"";C4;"")`

Na célula E5 introduza a fórmula: `=SE(C5<>"";E4+C5;"")`

Vá aba Exibição, no grupo de ferramentas Mostrar/Ocultar, desmarque Linhas de Grade

Deixe a planilha com a aparência abaixo. Procure desenhar os contornos das células usando a ferramenta:



Medidas Estatísticas Com Intervalo de Classe

i	Classes	Freq.	FR	FA	FAR	x_i
1	<280	0	0,00%	0	0,00%	
2	[280, 310)	3	12,00%	3	12,00%	
3	[310, 340)	4	16,00%	7	28,00%	
4	[340, 370)	6	24,00%	13	52,00%	
5	[370, 400)	7	28,00%	20	80,00%	
6	[400, 430)	4	16,00%	24	96,00%	
7	>=430	1	4,00%	25	100,00%	
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
Total		25	100,00%			

i	Bin	Freq.
1	280	0
2	310	3
3	340	4
4	370	6
5	400	7
6	430	4
7		1
8		#N/D
9		#N/D
10		#N/D
11		#N/D
12		#N/D
13		#N/D
14		#N/D
15		#N/D
16		#N/D
17		#N/D
18		#N/D
19		#N/D
20		#N/D
21		#N/D
22		#N/D
23		#N/D
24		#N/D
25		#N/D

Menor Valor		Amplitude das Classes		Maior Valor	
Número de Classes		Amplitude Total		Tamanho da Amostra	

Agora vamos ao cálculo das estatísticas.

Média Aritmética Ponderada

Aqui, convencionamos que *todos os valores incluídos em um determinado intervalo de classe coincidem com o seu ponto médio*, e determinamos a média aritmética ponderada por meio da fórmula:

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

Onde x_i é o ponto médio da classe.

Em G4, introduza a fórmula: **=ARREDONDAR.PARA.CIMA(B34;0)**.

Em G5, introduza a fórmula: **=SE(C5<>"";ARREDONDAR.PARA.CIMA(B34;0)+\$L\$31/2;"")**. Arraste a fórmula até a G28.

Em K4, digite: Classe. Em L4, digite: Valores.

Em I5, digite: média. A seguir selecione I5 e J5 e clique no botão Mesclar e Centralizar que se encontra no grupo Alinhamento da guia Início.

Em I6, digite: mediana. A seguir selecione I6 e J6 e clique no botão Mesclar e Centralizar.

Em I7, digite: moda. A seguir selecione I7 e J7 e clique no botão Mesclar e Centralizar.

Em I8, digite: Percentil. A seguir selecione J8 e pinte-a de verde usando a ferramenta de preenchimento.

Em I9, digite: Quartil(1). A seguir selecione I9 e J9 e clique no botão Mesclar e Centralizar.

Em I10, digite: Quartil(3). A seguir selecione I10 e J10 e clique no botão Mesclar e Centralizar.

Usando a ferramenta de preenchimento, procure colorir os intervalos K4:K10 e L4:L10.

Procure deixar assim:

		Classe	Valores
média			
mediana			
moda			
Percentil	50		
Quartil(1)			
Quartil(3)			

Em L5, introduza a fórmula: **=SOMARPRODUTO(\$C\$4:\$C\$28;\$G\$4:\$G\$28)/\$C\$29**. O resultado será: **364,6000**

Em J8 digite 50.

Mediana

Precisamos procurar o ponto do intervalo em que está compreendida a mediana. Para tanto, inicialmente, precisamos determinar a classe mediana, que é aquela correspondente à frequência acumulada imediatamente superior a $\frac{\sum f_i}{2}$.

Na célula K6, introduza a fórmula:

=PROCV(CORRESP(PROCV(\$D\$29/2;\$F\$4:\$G\$28;2)+\$L\$31;\$G\$4:\$G\$28);\$A\$4:\$B\$28;2). A resposta é [340,370).

Na célula K7, introduza a fórmula:

=PROCV(MÁXIMO(\$C\$4:\$C\$28);\$C\$34:\$D\$58;2;FALSO). A resposta é [370,400).

O cálculo da Mediana neste caso exige o emprego da fórmula:

$$Md = l_i^* + \frac{\left[\frac{\sum f_i}{2} - FA_{anterior} \right] h}{f^*}$$

Nela se encontram valores correspondentes à classe mediana. Precisamos pesquisar estes valores na tabela e retorná-los no seu devido lugar na fórmula. Esta pesquisa é feita através da função PROCV e precisamos ampliar a nossa matriz devido à utilização desta função.

Vá ao intervalo de células D33:G33, coloque os cabeçalhos: Classes, l_i , Li , FA.

Em D34, introduza a fórmula: **= "<"&B34+0,01**.

Em D35, introduza a fórmula: **=SE(B34="";"";SE(B35<>"";["&B34+0,01&", "&B35+0,01&"]);">"&B34+0,01))**.

Arraste esta fórmula até a célula D58.

Na célula E34, digite: 0

Em E35, introduza a fórmula: `=SE(E34<N31;J31;"")`.

Em E36, introduza a fórmula: `=SE(E35<N31;E35+L31;"")`. Arraste esta fórmula até a E58.

Em F34, introduza a fórmula: `=J31`.

Em F35, introduza a fórmula: `=SE(E35<N31;E35+L31;"")`. Arraste esta fórmula até a E58.

Em G34, introduza a fórmula: `=E4`. Arraste esta fórmula até G58.

A tabela ficará assim:

i	Bin	Freq.	Classes	li	Li	FA
1	280	0	<280	0	280	0
2	310	3	[280, 310)	280	310	4
3	340	4	[310, 340)	310	340	13
4	370	6	[340, 370)	340	370	24
5	400	7	[370, 400)	370	400	32
6	430	4	[400, 430)	400	430	37
7		1	>430	430		40
8		#N/D				
9		#N/D				
10		#N/D				
11		#N/D				
12		#N/D				
13		#N/D				
14		#N/D				
15		#N/D				
16		#N/D				
17		#N/D				
18		#N/D				
19		#N/D				
20		#N/D				
21		#N/D				
22		#N/D				
23		#N/D				
24		#N/D				
25		#N/D				

Agora em K16, digite: Argumentos.

Em K17, digite: Mediana.

No intervalo L16:Q16 digite: li, Li, FA, FA-ant, fi, Cálculo

Veja como ficou⁵:

Argumentos	li	Li	FA	FA-ant	fi	Cálculo		
Mediana								

Em L17, introduza a fórmula: `=PROCV(K6;D34:F58;2;FALSO)`.

Em M17, introduza a fórmula: `=PROCV(K6;D34:F58;3;FALSO)`.

Em N17, introduza a fórmula: `=PROCV(K6;B4:G28;4;FALSO)`.

Em O17, introduza a fórmula: `=PROCV(C29/2;E4:F28;1)`.

Em P17, introduza a fórmula: `=PROCV(K6;B4:G28;2;FALSO)`.

⁵ Gostaria de lembrar a existência das teclas de atalho CTRL + C e CTRL + V. Isto evitará a digitação de fórmulas da apostila para o Excel. Por isso não coloquei segurança no arquivo pdf.

Em Q17, introduza a fórmula: $=\$L\$17+((\$C\$29/2-\$O\$17)/\$P\$17)*\$L\31 .

Aproveite para deixar a tabela como esta:

Argumentos	li	Li	FA	FA-ant	fi	Cálculo		
Mediana	340	370	13	7	6			
Percentil								
Quartil(1)								
Quartil(3)								

A resposta será: **367,5000**.

Em L6, introduza a fórmula: $=\$Q\17 .

Moda

Quando os dados estiverem agrupados em intervalos de classe, precisamos procurar a classe modal, isto é, aquela classe cuja frequência é máxima.

Na célula K7, introduza a fórmula:

$=PROCV(MÁXIMO(\$C\$4:\$C\$28);\$C\$34:\$D\$58;2;FALSO)$. A resposta é **[370,400]**.

Em L7, introduza a fórmula: $=PROCV(MÁXIMO(\$C\$4:\$C\$28);\$C\$4:\$G\$28;5;FALSO)$. A resposta é **385,0000**.

Percentil

Em K8, introduza a fórmula:

$=PROCV(CORRESP(PROCV(\$J\$8/100; \$F\$4:\$G\$28;2)+\$L\$31; \$G\$4:\$G\$28); \$A\$4:\$B\$28;2)$.

Em L8, introduza a fórmula: $=\$Q\18

Em L18, introduza a fórmula: $=PROCV(\$K\$8; \$D\$34:\$F\$58;2;FALSO)$.

Em M18, introduza a fórmula: $=PROCV(\$K\$8; \$D\$34:\$F\$58;3;FALSO)$.

Em N18, introduza a fórmula: $=PROCV(\$K\$8; \$B\$4:\$G\$28;4;FALSO)$.

Em O18, introduza a fórmula: $=PROCV(\$C\$29*\$J\$8/100; \$E\$4:\$F\$28;1)$.

Em P18, introduza a fórmula: $=PROCV(\$K\$8; \$B\$4:\$G\$28;2;FALSO)$.

Em Q18, introduza a fórmula: $=\$L\$18+((\$J\$8*\$C\$29/100)-\$O\$18)*\$L\$31/\$P\18 .

A resposta é **367,5000**, se você tivesse digitado 50 em J8.

Quartil(1)

Em K9, introduza a fórmula: $=PROCV(CORRESP(PROCV(25/100; \$F\$4:\$G\$28;2)+\$L\$31; \$G\$4:\$G\$28); \$A\$4:\$B\$28;2)$.

Em L9, introduza a fórmula: $=\$Q\19

Em L19, introduza a fórmula: $=PROCV(\$K\$9; \$D\$34:\$F\$58;2;FALSO)$.

Em M19, introduza a fórmula: $=PROCV(\$K\$9; \$D\$34:\$F\$58;3;FALSO)$.

Em N19, introduza a fórmula: $=PROCV(\$K\$9; \$D\$34:\$F\$58;3;FALSO)$.

Em O19, introduza a fórmula: $=PROCV(\$C\$29*25/100; \$E\$4:\$F\$28;1)$.

Em P19, introduza a fórmula: $=PROCV(\$K\$9; \$B\$4:\$G\$28;2;FALSO)$.

Em Q19, introduza a fórmula: $=\$L\$19+((25*\$C\$29/100)-\$O\$19)*\$L\$31/\$P\19 .

A resposta é **334,375**.

Quartil(3)

Em K10, introduza a fórmula:

$=PROCV(CORRESP(PROCV(75/100; \$F\$4:\$G\$28;2)+\$L\$31; \$G\$4:\$G\$28); \$A\$4:\$B\$28;2)$.

Em L10, introduza a fórmula: $=\$Q\20

Em L20, introduza a fórmula: $=PROCV(\$K\$10; \$D\$34:\$F\$58;2;FALSO)$.

Em M20, introduza a fórmula: $=PROCV(\$K\$10; \$D\$34:\$F\$58;3;FALSO)$.

Em N20, introduza a fórmula: $=PROCV(\$K\$10; \$B\$4:\$G\$28;4;FALSO)$.

Em O20, introduza a fórmula: **=PROCV(\$C\$29*75/100;\$E\$4:\$F\$28;1).**

Em P20, introduza a fórmula: **=PROCV(\$K\$10;\$B\$4:\$G\$28;2;FALSO).**

Em Q20, introduza a fórmula: **=\$L\$20+((75*\$C\$29/100)-\$O\$20)*\$L\$31/\$P\$20.**

A resposta é **394,6429**.

A planilha ficou assim:

Medidas Estatísticas Com Intervalo de Classe

i	Classes	Freq.	FR	FA	FAR	x_i
1	<280	0	0,00%	0	0,00%	280
2	[280, 310)	3	12,00%	3	12,00%	295
3	[310, 340)	4	16,00%	7	28,00%	325
4	[340, 370)	6	24,00%	13	52,00%	355
5	[370, 400)	7	28,00%	20	80,00%	385
6	[400, 430)	4	16,00%	24	96,00%	415
7	>=430	1	4,00%	25	100,00%	445
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
Total		25	100,00%			

	Classe	Valores
média		364,60000
mediana	[340, 370)	367,5000
moda	[370, 400)	385,00000
Percentil 50	[340, 370)	367,5
Quartil(1)	[310, 340)	334,375
Quartil(3)	[370, 400)	394,64286

Argumentos	li	Li	FA	FA-ant	fi	Cálculo	
Mediana	340	370	13	7	6	367,5000	
Percentil	340	370	13	7	6	367,5	
Quartil(1)	310	340	340	3	4	334,375	
Quartil(3)	370	400	20	13	7	394,6429	

Menor Valor	280	Amplitude das Classes	30	Maior Valor	430
Número de Classes	5	Amplitude Total	150	Tamanho da Amostra	25

É assim que se faz...É mole a vida???? Pare com isso, cara! Precisa gostar muito.....

Aproveite e brinque com esta planilha à vontade. Mostre isto para os seus amigos, parentes, etc..

Podemos montar tudo isto num histograma? Claro que sim, mas deixa isto para depois que o bicho pega ainda mais bravo...

2.6 – Medidas Estatísticas na Calculadora Casio fx-82MS - APÊNDICE

Em estatística o que não falta é conta a ser realizada. Para atender aqueles que não têm notebook colocamos este apêndice ensinando como realizar alguns cálculos importantes na calculadora científica mais popular do mercado. Como muitas outras concorrentes seguem os mesmos princípios de cálculo, este apêndice é de utilidade geral para as calculadoras.

2.6.1 - Configurando

Os cálculos estatísticos são realizados em dois modos:

- O chamado **Desvio Padrão - SD**
- E o modo de **Regressão – REG**

Para ativar estes modos utilizamos a tecla .

Após pressionada a tecla acima ficamos com o visor: 

Configurando o modo **Desvio Padrão - SD**

Após pressionada a tecla acima ficamos com o visor:



Digitando a tecla  o visor ficará assim:

No modo **SD** e no modo **REG**, a tecla  como a tecla .

2.6.2 – Entrando com Dados

➤ A **entrada de dados** sempre começa com uma sequência de teclas para **limpar** as memórias de estatística:

 aparece o menu: . A seguir digite:

Limpando memórias



➤ Entre com os dados usando a sequência de teclas : <dados x> .

2.6.3 – Calculando...

Estes dados serão usados por ela para calcular, automaticamente, os seguintes valores:

$$n, \sum x, \sum x^2, \bar{x}, \sigma_n \text{ e } \sigma_{n-1}.$$

Eles poderão ser recuperados usando-se as operações de teclas seguintes:

Para recuperar este tipo de valor	Realize esta operação de teclas:
$\sum x^2$	
$\sum x$	
n	
\bar{x}	
σ_n	
σ_{n-1}	

Exercícios de Aplicação

Calcular n, $\sum x$, $\sum x^2$, \bar{x} para os seguintes dados: 55, 54, 51, 55, 53, 53, 54, 52.

Solução

Primeiramente, no modo SD, devemos **limpar** as memórias estatísticas:



A seguir devemos entrar com os dados assim:

55 **DT** O visor irá mostrar:

SD
 n=
1

A cada vez que se introduz um dado, o número de dados que se introduziu até este ponto é indicado no visor (valor n).

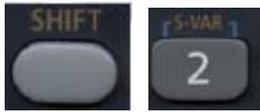
Continuemos introduzindo os demais dados:

54 **DT** 51 **DT** 55 **DT** 53 **DT** **DT** 54 **DT** 52 **DT**

Quando os valores forem repetidos fazemos desta forma

Recuperemos os valores calculados automaticamente:





... $\bar{x} = 53,375....$ esta é a média aritmética. E a ponderada? Calma, veja a seguir

2.6.4 – Observações quanto à entrada de Dados

1. introduz o dado duas vezes.

2. Podemos introduzir também múltiplas entradas do mesmo dado usando:



Por exemplo para introduzir o dado 110 vezes, pressione 110 10 .

3. As operações de teclas anteriores podem ser realizadas em qualquer ordem e, não necessariamente como mostrado acima.

4. Enquanto se introduz os dados ou depois de completar a introdução, pode-se usar as teclas para ir visualizando através dos dados que se introduziu. Se se introduzir múltiplas entradas do mesmo dado usando o que fora dito no item 2, ao se passar através dos dados é mostrado o item do dado e uma janela separada para a frequência de dados.

5. Os dados visualizados podem ser editados se assim o desejar. Introduza o novo valor e logo em seguida pressione a tecla para trocar o valor antigo pelo novo valor. Isto também significa que se desejar realizar alguma outra operação (cálculo, chamada de resultados de cálculos estatísticos, etc.), sempre deverá pressionar primeiro a tecla para sair da apresentação de dados.

6. Pressionando a tecla no lugar de depois de trocar um valor sobre a apresentação, registra o valor que se introduziu como um elemento de dado novo, e deixa o valor antigo tal como está.

7. Pode-se apagar o valor do dado visualizado usando e e logo pressionando . Apagando um valor de dado, ocasiona que todos os valores seguintes se desloquem para cima.

8. Os valores de dados que se registra normalmente se armazenam na memória da calculadora. Aparece a mensagem "Data Full" e não poderá introduzir nenhum dado se não houver memória disponível para o armazenamento de dados. Se isto acontecer, pressione a tecla para visualizar a janela mostrada a seguir.

Edit	OFF	ESC
1	2	

Pressione para sair da introdução de dados sem registrar o valor que ingressou recentemente.

Pressione se desejar registrar o valor introduzido recentemente, sem armazená-lo na memória. Se fizer isto, entretanto, não poderá visualizar ou editar nenhum dos dados que introduziu.

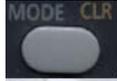
9. Para apagar os dados que introduziu recentemente, pressione .

10. Depois de se introduzirem os dados estatísticos no modo SD ou modo REG, não poderá visualizar ou editar mais os itens de dados individuais, depois de realizar qualquer das operações seguintes.

Trocando a um outro modo.

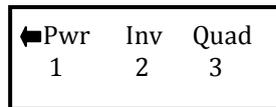
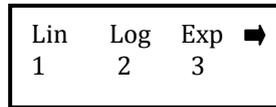
Trocando o tipo de regressão (Lin, Log, Exp, Pwr, Inv, Quad)

2.6.5 – Introdução de dados Bivariados.

Precisamos agora configurar a calculadora para o modo REG. Para isso utilizamos a tecla  para esta

seleção. Aqui (como no modo SD) a tecla  funciona como a tecla .

Ao selecionar o modo REG, serão exibidas as telas como mostradas abaixo



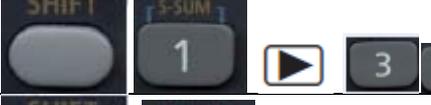
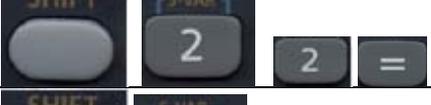
Pressione a tecla numérica (,  ou ) que corresponde ao tipo de regressão que você deseja utilizar. Por enquanto vamos usar a regressão linear para o que vamos explicar a seguir. Portanto, digite  na primeira tela.

Sempre inicie a introdução dos dados com    (Scl)  para apagar a memória de estatísticas.

Introduza os dados usando a sequência de teclas mostradas abaixo:

<Dados-x>  <Dados-y> .

Os valores produzidos por um cálculo de regressão dependem dos valores introduzidos, e os resultados podem ser chamados usando as operações de teclas mostradas abaixo:

Para recuperar este tipo de valor	Realize esta operação de teclas:
$\sum x^2$	
$\sum x$	
n	
$\sum y^2$	
$\sum y$	
$\sum xy$	
\bar{x}	
$x\sigma_n$	
$x\sigma_{n-1}$	
\bar{y}	
$y\sigma_n$	
$y\sigma_{n-1}$	

Use e abuse destes somatórios nas fórmulas que você desejar.

Exercícios de Aplicação

Calcular n , $\sum x$, $\sum x^2$, \bar{x} , $\sum y$, $\sum y^2$, \bar{y} , $x\sigma_n$, $x\sigma_{n-1}$, $y\sigma_n$, $y\sigma_{n-1}$, variância de x , variância de y e coeficiente de variação de x e de y para os seguintes dados do exemplo de tabela de frequência de dados agrupados sem intervalo de classe dada abaixo

x_i agrupado	frequência f_i
280	2
305	1
310	1
320	1
330	2
340	1
341	1
355	1
360	1
365	1
369	1
370	3
371	1
375	1
380	1
390	1
400	2
401	1
420	1
430	1

Solução

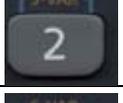
Introduza os dados ao lado:

Lembre-se de configurá-la para REG e limpara as memórias de estatísticas, depois.....

<Dados-x>  <Dados-y> . Assim,

280  2 . Continue você....

Agora mande a danada (hehehehe) fazer os cálculos para você. Ela não custou uma fortuna? Tem que mostrar serviço...

Para recuperar este tipo de valor	Realize esta operação de teclas:	Resultado:
$\sum x^2$	  1  =	2.630.205,00
$\sum x$	  1  =	7.213,00
n	  1  =	20,00
$\sum y^2$	  1  1 =	37,00
$\sum y$	  1  2 =	25,00
$\sum xy$	  1  3 =	8.963,00
\bar{x}	  2  =	360,65
$X\sigma_n$	  2  =	37,9714
$X\sigma_{n-1}$	  2  =	38,9578

\bar{y}	SHIFT	S-VAR	2	▶	1	=	1,25
$y\sigma_n$	SHIFT	S-VAR	2	▶	2	=	0,5362
$y\sigma_{n-1}$	SHIFT	S-VAR	2	▶	3	=	0,5501

Variância amostral de x : $38,9578^2 = 1.517,71$

Variância amostral de y : $0,5501^2 = 0,3026$

Variância populacional de x : $37,9714^2 = 1.441,8272$

Variância populacional de y : $0,5362^2 = 0,2875$

Coefficiente de variação de x: $CV = \sigma/\bar{x} \Rightarrow CV_x = 37,9714/360,65 = 0,1053$

Coefficiente de variação de y: $CV = \sigma/\bar{y} \Rightarrow CV_y = 0,5362/1,25 = 0,4290$.

Media ponderada de x: $\frac{\sum x f_i}{\sum f_i} = 8.963,00/25,00 = 358,52$

Procure outras medidas e realize os seus cálculos (a calculadora está “carregada” mesmo)...Seja feliz, e saia pelo mundo afora mostrando os seus dotes de grande “calculista” do planalto paulista. Mostre isto no “zero grau” de SJRP, o pessoal vai delirar e pedir mais um chopp.



UM ESTATÍSTICO É AQUELE QUE,
TENDO A CABEÇA A ARDER E OS
PÉS ENTERRADOS NO GELO,
AINDA DIZ QUE NA MÉDIA
ESTÁ TUDO BEM!...

2.7 – Medidas Estatísticas na Calculadora HP-12C - APÊNDICE⁶

Na HP12C, os dados estatísticos são armazenados como um conjunto de somatórios resultantes dos dados coletados originalmente. O conjunto dos dados coletados originalmente deve ser digitado antes de se usar quaisquer características estatísticas disponíveis na HP12C porque todos os valores produzidos por estas ferramentas estatísticas dependem deles. A organização da memória da HP12C permite o estudo dos dados estatísticos organizados como amostras de uma ou duas variáveis. Como um *procedimento geral*, os dados são sempre coletados como um par de números, ou valores (x,y), e a HP12C calcula para eles as seguintes somas:

$$\sum x_n \quad \sum y_n \quad \sum (x_n)^2 \quad \sum (y_n)^2 \quad \sum (x_n \times y_n) \quad \text{Figura 1}$$

2.7.1 – Introdução dos Dados

No caso de um par de dados, digita-se o dado **y** **ENTER**, o dado **x** e depois pressione a tecla **Σ+**. A HP-12C calcula automaticamente as estatísticas e armazena nos registradores R1 a R6, como mostra a Tabela abaixo:

Registrador	Estatística
R1 e visor	N
R2	Σx
R3	Σx²
R4	Σy
R5	Σy²
R6	Σxy

← = nº de pares de dados acumulados

Recuperando estes somatórios, você faz o que bem entender com eles.

2.7.2 – Corrigindo as estatísticas acumuladas

Se você descobrir que introduziu dados incorretos, as estatísticas acumuladas podem ser facilmente corrigidas:

- Se o dado (ou par de dados) tiver acabado de ser introduzido, e **Σ+** tiver sido pressionada, pressione **g** **LSTx** **g** **Σ-**.
- Se o dado (ou par de dados) não for o mais recentemente introduzido, introduza o dado (ou par de dados) incorreto novamente, como se fosse um novo dado, e então pressione **g** **Σ-**, ao invés de **Σ+**. Tal procedimento cancela o efeito do dado (ou par de dados) incorreto. Basta agora introduzir o dado corretamente, usando **Σ+** como se fosse um novo dado.

Com estes valores atualizados e armazenados na memória, a HP12C calcula a média (\bar{x} , \bar{y}) para cada variável com a seguinte expressão:

$$\bar{x} = \frac{\sum x_n}{n} \quad \text{e} \quad \bar{y} = \frac{\sum y_n}{n}$$

Figura 2

As expressões seguintes são usadas pela HP12C para calcular o desvio padrão⁷ de uma amostra:

$$S_x = \sqrt{\frac{n \sum (x_n)^2 - (\sum x_n)^2}{n(n-1)}} \quad \text{e} \quad S_y = \sqrt{\frac{n \sum (y_n)^2 - (\sum y_n)^2}{n(n-1)}}$$

Figura 3

Prática em encontrar preços de venda médios e desvios padrão na HP-12C

Exemplo 1: Os preços de vendas das últimas 10 casas vendidas na comunidade *Guarulhos Green* foram: R\$198.000; R\$185.000; R\$205.200; R\$225.300; R\$206.700; R\$201.850; R\$200.000; R\$189.000; R\$192.100; R\$200.400. Qual foi a **média** destes preços de venda e qual é o **desvio padrão** da amostra? Um preço de venda de \$240.000 será considerado não usual na mesma comunidade?

Solução: certifique-se de apagar as memórias somatório/estatística antes de iniciar o problema.

f **Σ**

⁶ Este material foi montado para um curso *in company* de Finanças na HP 12C que ministrei na empresa de filtros de óleo automotivo **Tecfil** (www.sofape.com) em Guarulhos-SP através do **INVENT** (www.inventrade.com.br).

⁷ O **DESVIO PADRÃO** é a medida da **DISPERSÃO** dos dados em torno da média. O s_x e o s_y na HP 12C dão a **melhor estimativa** do $s_{população}$, a partir dos dados da amostra.



Figura 4

Cada valor de dado entrado faz o visor mudar e mostrar o número da entrada atual (n). Agora entre cada valor de dado com $\Sigma+$:



Figura 5

O visor representado na Figura 1 mostra o valor atual 1 das n entradas.

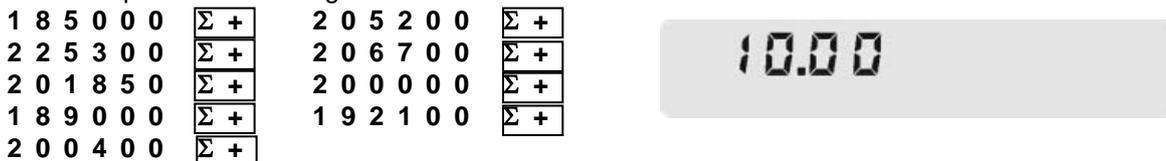


Figura 6

A Figura 6 representa o visor após a última entrada. Com todos os dados entrados anteriormente entrados, todos os somatórios estão prontos e é possível calcular a média e o desvio padrão. Para calcular a **média** aperte:



Figura 7

\bar{x} é a função azul na frente da face inclinada da tecla 0 , assim g (a tecla de prefixo azul) deve ser pressionada primeiro.

Para calcular o desvio padrão, aperte: g s



Figura 8

s é a função azul na frente da face inclinada da tecla . .

Baseado nestes números, aproximadamente 68% dos preços estão no intervalo de \$200.355,00 \pm \$11.189,04. Aproximadamente 95% dos preços estão no intervalo \$200.355,00 \pm 2x(\$11.189,04).

A seqüência de teclas seguinte dá o limite inferior:



Figura 9

O visor mostra o limite inferior. Para calcular o limite superior, se nenhuma operação foi realizada após as teclas acima, pressione:



Figura 10

O visor mostra o limite superior.

Resposta: \$240.000,00 é um preço não usual para uma casa na comunidade *Parkdale* baseado nos últimos 10 preços de venda.

Prática com média e desvio padrão com duas variáveis

Exemplo 2: Um recenseador de terrenos quer calcular a relação entre a área construída e a área do terreno de oito casas localizadas na sua vizinhança. Inicialmente ele precisa saber a média e o desvio padrão para ambos parâmetros. Suas medidas permitiram-lhe construir o seguinte quadro:

Área do Terreno (m ²)	Área Construída (m ²)	Área do Terreno (m ²)	Área Construída(m ²)
12000	3120	9000	2080
10000	2560	10000	2700
11000	2920	13000	3280
14000	3300	12000	3080

Solução: Certifique-se em apagar as memórias estatísticas/somatório antes de iniciar o problema.

f Σ

0.00

Figura 11

Cada par deve ser entrado para adicioná-lo aos somatórios estatísticos.

3 1 2 0 ENTER 1 2 0 0 0 $\Sigma +$

10.00

Figura 12

O primeiro valor entrado (área construída) é calculado como a variável y, e o segundo, o valor (área do terreno), é calculado como a variável x. O visor mostra o número de entradas. Certifique-se de que todos os dados foram entrados:

2 5 6 0 ENTER 1 0 0 0 0 $\Sigma +$
 2 9 2 0 ENTER 1 1 0 0 0 $\Sigma +$
 3 3 0 0 ENTER 1 4 0 0 0 $\Sigma +$
 2 0 8 0 ENTER 9 0 0 0 $\Sigma +$
 2 7 0 0 ENTER 1 0 0 0 0 $\Sigma +$
 3 2 8 0 ENTER 1 3 0 0 0 $\Sigma +$
 3 0 8 0 ENTER 1 2 0 0 0 $\Sigma +$

8.00

Figura 13

Para calcular a média da de terreno:

g \bar{x}

11,375.00

Figura 14

Área de terreno média: 11.375 m². Agora pressionando $x \geq y$

2,880.00

Figura 15

Temos área média de construção: 2.880 m².

Para calcular o desvio padrão:

g s

1,685.02

Figura 16

Desvio padrão para a área de terreno: 1.685,02 m². Pressionando $\boxed{x \geq y}$



Figura 17

Temos o desvio padrão para a área construída: 415,83 m².

Resposta: A média da área de terreno para esta amostra é 11.375 m² e o desvio padrão é 1.685,02 m². Para a área de construção, a média é 2.880 m² e o desvio padrão é 415,83 m².

Exercícios Propostos

1. Uma pesquisa feita com sete vendedores de sua empresa revelou os dados da tabela dada a seguir. Quantas horas um vendedor trabalha, em média, por semana? Quanto ele vende, em média, por mês?

Vendedor	Horas por Semana	Vendas por Mês
1	32	R\$ 1.700.000,00
2	40	R\$ 2.500.000,00
3	45	R\$ 2.600.000,00
4	40	R\$ 2.000.000,00
5	38	R\$ 2.100.000,00
6	50	R\$ 2.800.000,00
7	35	R\$ 1.500.000,00

2. No exercício anterior, encontrar o desvio padrão das vendas e das horas trabalhadas por semana.

Resposta 1: Média das vendas é R\$ 2.171.428,57.

Média das horas de trabalho por semana é 40 h.

Resposta 2: Desvio padrão das vendas: R\$ 482.059,08

Desvio padrão das horas trabalhadas: 6,03 h.

2.7.3 – Média Ponderada na HP 12C

Na HP12C, dados estatísticos são armazenados como um conjunto de somatórios resultante dos dados coletados originalmente. O conjunto de dados coletados deve ser digitado antes de se usar qualquer função estatística disponível na HP12C porque todas as funções estatísticas usam valores produzidos durante os somatórios. A média ponderada é calculada com o uso da tecla \bar{x}_w e os conteúdos de dois somatórios são usados.

$$\bar{x}_w = \frac{\sum w x}{\sum w} \quad \text{Figura 2}$$

onde:

x é o valor que se repete
 w é o número de ocorrências de x (peso)
 \bar{x}_w é a média ponderada

Exercícios de Aplicação

1- Um grande *shopping center* quer saber a média ponderada dos preços de venda de 2.000 unidades de um produto que tem o seu preço final ajustado de acordo com os primeiros dez dias de vendas. A tabela abaixo resume a relação entre o preço final e o número de unidades vendidas.

Preço por unidade	# de unidades vendidas	Preço por unidade	# de unidades vendidas
\$24.20	354	\$24.14	288
\$24.10	258	\$24.06	240
\$24.00	209	\$23.95	186
\$23.90	133	\$23.84	121
\$23.82	110	\$23.75	101

Figura 3

Calcule o **preço médio** e a **média ponderada dos preços** de vendas deste produto.

Solução: Certifique-se em apagar as memórias estatísticas/somatório antes de iniciar o problema.

f Σ

0.00

Figura 4

Médias regulares e médias ponderadas podem ser calculadas dos mesmos dados acumulados na HP12C, desde que a ordem dos valores seja entrada corretamente: *valor* **ENTER** *peso*.

2 4 . 2 0 **ENTER** 3 5 4 **Σ +**

1.00

Figura 5

Os valores restantes e os seus pesos são entrados da mesma forma.

2 4 . 1 4 **ENTER** 2 8 8 **Σ +** 2 4 . 1 0 **ENTER** 2 5 8 **Σ +**
 2 4 . 0 6 **ENTER** 2 4 0 **Σ +** 2 4 . 0 0 **ENTER** 2 0 9 **Σ +**
 2 3 . 9 5 **ENTER** 1 8 6 **Σ +** 2 3 . 9 0 **ENTER** 1 3 3 **Σ +**
 2 3 . 8 4 **ENTER** 1 2 1 **Σ +** 2 3 . 8 2 **ENTER** 1 1 0 **Σ +**
 2 3 . 7 5 **ENTER** 1 0 1 **Σ +**

10.00

Figura 6

Para calcular a média ponderada dos preços de venda:

g \bar{x}_w

24.03

Figura 7

Para calcular o preço médio:

g \bar{x} **R**↓

23.98

Figura 8

Resposta: Embora o preço médio para este produto seja \$23,98, a média ponderada para as vendas ocorridas nos primeiros dez dias foi \$24,03. Note que a tecla **R**↓ é pressionada porque o valor que aparece no visor após **g** \bar{x} ser pressionados é a média dos pesos e não será de nenhuma utilidade neste exemplo.

2 - Estimar os custos do combustível numa viagem de férias permite planejar melhor as próximas viagens. A média ponderada é uma referência melhor para se calcular a média atual quando se compra gasolina em postos com diferentes preços por litro. A tabela abaixo se refere a uma viagem de férias regular e relaciona a gasolina comprada (peso) em litros pelo preço do litro (valor).

Quantidade de gasolina (litros)	Preço por litro	Quantidade de gasolina (litros)	Preço por litro
12	\$1.26	9	\$1.32
13	\$1.20	29	\$1.12
31	\$1.18	13	\$1.25

Figura 9

Baseado nestes valores calcule a média ponderada e o custo médio por litro de gasolina comprada.

Solução: Certifique-se de apagar as memórias estatísticas/somatório antes de iniciar o problema.

f Σ

0.00

Figura 10

Médias regulares e médias ponderadas podem ambas ser calculadas dos mesmos dados acumulados na HP12C, desde que os valores sejam corretamente entrados em ordem: *valor ENTER peso*.

1 . 2 6 ENTER 3 2 $\Sigma +$

1.00

Figura 11

Os valores restantes e seus pesos são entrados da mesma maneira.

1 . 3 2 ENTER 9 $\Sigma +$

1 . 2 0 ENTER 1 3 $\Sigma +$

1 . 1 2 ENTER 2 9 $\Sigma +$

1 . 1 8 ENTER 3 1 $\Sigma +$

1 . 2 5 ENTER 1 3 $\Sigma +$

6.00

Figura 12

Para calcular a média ponderada da gasolina comprada:

g \bar{x}_w

1.20

Figura 13

Para calcular o preço médio da gasolina:

g \bar{x} $R\downarrow$

1.22

Figura 14

Resposta: Embora o preço médio da gasolina seja \$1,22 por litro, a média ponderada para esta viagem foi de \$1.20 por litro. Note que a tecla $R\downarrow$ é pressionada porque o valor que aparece no visor após g \bar{x} ser pressionados é a média dos pesos e não será de nenhuma utilidade neste exemplo.

2.7.4 – Média, desvio padrão e erro padrão de Dados Agrupados na HP 12C

Dados agrupados são apresentados em distribuições de frequência, em vez de inserir cada observação individualmente. Dado um conjunto de pontos de dados:

X_1, X_2, \dots, X_n
 f_1, f_2, \dots, f_n

com as respectivas frequências:
 este procedimento computa a média, o desvio padrão e o erro padrão da média.

Instruções

Siga estes passos para computar a média, o desvio padrão e o erro padrão da média:

1. Pressione **f** e, em seguida, **CLEAR [REG]**.
2. Digite o primeiro valor e pressione **ENTER** duas vezes.
3. Digite a respectiva frequência e pressione **STO**, **[+]**, **0**, **[X]** e, em seguida, **$\Sigma+$** . O visor mostra o número de pontos de dados inseridos.
4. Repita os passos 2 e 3 para cada ponto de dados.
5. Para calcular a média, pressione **RCL**, **0**, **STO**, **1**, **RCL**, **6**, **STO**, **3**, **g** e, em seguida, \bar{x} (a tecla zero).
6. Pressione **g** e, em seguida, **s** para encontrar o desvio padrão.
7. Pressione **RCL**, **0**, **g**, a tecla $\sqrt{\quad}$ e \div para encontrar o erro padrão da média.

Exemplo - Uma pesquisa de aluguéis de 266 apartamentos de um quarto revela que 54 são alugados por R\$ 190,00 por mês, 32 são alugados por R\$ 195,00 por mês, 88 são alugados por R\$ 200,00 por mês e 92 são alugados por R\$ 206,00 por mês. Qual é o aluguel mensal médio, o desvio padrão e o erro padrão da média?

Toques no teclado (modo RPN)	Exibição	Comentários
f , CLEAR [REG] , 190 , ENTER , ENTER , 54 , STO , [+] , 0 , [X] e, em seguida, $\Sigma+$	1,00	Calculadora limpa e primeiro ponto de dados inserido
195 , ENTER , ENTER , 32 , STO , [+] , 0 , [X] e, em seguida, $\Sigma+$	2,00	Segundo ponto de dados inserido
200 , ENTER , ENTER , 88 , STO , [+] , 0 , [X] e, em seguida, $\Sigma+$	3,00	Terceiro ponto de dados inserido
206 , ENTER , ENTER , 92 , STO , [+] , 0 , [X] e, em seguida, $\Sigma+$	4,00	Quarto ponto de dados inserido
RCL , 0 , STO , 1 , RCL , 6 , STO , 3 , g e, em seguida \bar{x}	199,44	Aluguel mensal médio
g e, em seguida, s	5,97	Desvio padrão
RCL , 0 , g , $\sqrt{\quad}$ e, em seguida, \div	0,37	Erro padrão da média

E aí, gostou?

Bacana, né !!! (hehehehehe)

Nas questões matemáticas não se compreende a incerteza nem a dúvida, assim como tampouco se podem estabelecer distinções entre verdades médias e verdades de grau superior.
(Hilbert)

2.8 – Algumas funções da Categoria Estatística do Excel - APÊNDICE

Cara, não tem coisa mais útil nos aplicativos que o Ajuda (Help). Antigamente os autores de livro de informática ficavam traduzindo o tal Ajuda para o Português e editavam livros. Eram conhecidos como exímios conhecedores do assunto, mas na verdade apenas sabiam o Inglês ou dispunham de um bom dicionário. Agora a mamata acabou. As coisas já estão traduzidas e, então, vamos usá-la. Você já pesquisou no Ajuda alguma vez? Senão, faça isto sempre. É bom, ajuda.....

Vejam algumas funções de medidas de tendência central e de dispersão que tirei de lá, portanto, não precisa falar bem de mim. Tô citando a fonte.

DESV.MÉDIO(núm1;núm2;...)

Retorna a média dos desvios absolutos dos pontos de dados a partir de sua média. Os argumentos podem ser números ou nomes, matrizes ou referências que contenham números.

núm1: núm1; núm2; ... de 1 a 255 argumentos cuja média dos desvios absolutos se deseja calcular.

DESVPAD(núm1;núm2;...)

Calcula o desvio padrão a partir de uma amostra (ignora os valores lógicos e texto da amostra)

núm1: núm1; núm2; ... de 1 a 255 argumentos cuja média dos desvios absolutos se deseja calcular.

MÉDIA(núm1;núm2;...)

Retorna a média (aritmética) dos argumentos que podem ser números ou nomes, matrizes ou referências que contêm números

núm1: núm1; núm2; ... de 1 a 255 argumentos cuja média se deseja calcular.

Para ser chamada no **VBA**, temos a Sub:

MÉDIA.GEOMÉTRICA(núm1;núm2;...)

Retorna a média geométrica de uma matriz ou um intervalo de dados numéricos positivos

núm1: núm1; núm2; ... de 1 a 255 números ou nomes, matrizes ou referências que contenham números cuja média você deseja calcular.

MED(núm1;núm2;...)

Retorna a mediana, ou o número central de um determinado conjunto de números.

núm1: núm1; núm2; ... de 1 a 255 números ou nomes, matrizes ou referências que contêm números cuja mediana se deseja obter

MODOS(núm1;núm2;...)

Retorna o valor mais repetido ou que ocorre com maior frequência, em uma matriz ou um intervalo de dados.

núm1: núm1; núm2; ... de 1 a 255 números ou nomes, matrizes ou referências que contêm números cuja moda se deseja obter.

VAR(núm1;núm2;...)

Estima a variância com base em uma amostra (ignora valores lógicos e texto na amostra).

núm1: núm1; núm2; ... de 1 a 255 argumentos numéricos que correspondem a uma amostra de uma população

VARP(núm1;núm2;...)

Estima a variância com base na população total (ignora valores lógicos e texto da população).

núm1: núm1; núm2; ... de 1 a 255 argumentos numéricos que correspondem a uma população

PERCENTIL(matriz;k)

Retorna o k-ésimo percentil de valores em um intervalo

Matriz é a matriz ou intervalo de dados que define a posição relativa

K é o valor do percentil no intervalo 0 a 1, inclusive.

QUARTIL(matriz;quarto)

Retorna o quartil do conjunto de dados

Matriz é a matriz ou intervalo de células de valores numéricos cujo valor quartil você deseja obter

Quarto é um número; valor mínimo = 0, primeiro quartil = 1, valor mediano = 2, terceiro quartil = 3, valor máximo = 4.

CURT(núm1;núm2;...)

Retorna a curtose de um conjunto de dados.

núm1: núm1; núm2; ... de 1 a 255 números ou nomes, matrizes ou referências que contenham números cuja curtose se deseja obter.

Na planilha intitulada *Funções Estatísticas* apresentamos os cálculos de algumas medidas de tendência central e de dispersão.

	A	B	C	D	E
1	Calculando as Estatísticas com as Funções Embutidas do Excel				
2					
3	Dados				
4	25			No Excel	
5	23	Média Aritmética	23,40740741	<---=MÉDIA(A4:A30)	
6	24	Média Geométrica	23,33324073	<---=MÉDIA.GEOMÉTRICA(A4:A30)	
7	25	Mediana	24	<---=MED(A4:A30)	
8	25	Moda	25	<---=MODO(A4:A30)	
9	26	Percentil - 30	22	<---=PERCENTIL(A4:A30;0,3)	
10	22	Quartil (3)	25	<---=QUARTIL(A4:A30;3)	
11	23				
12	24	Desvio Médio	1,577503429	<---=DESV.MÉDIO(A4:A30)	
13	24	Desvio Padrão	1,886373388	<---=DESVPAD(A4:A30)	
14	25	Variância de Amostra	3,558404558	<---=VAR(A4:A30)	
15	26	Variância da População	3,426611797	<---=VARP(A4:A30)	
16	27	Curtose	-0,768545507	<---=CURT(A1:A30)	
17	23				
18	21				
19	22				
20	20				
21	22				
22	24				
23	25				
24	22				
25	23				
26	21				
27	24				
28	25				
29	21				
30	20				
31					

Ou

	A	B	C	D	E	F	G	H
1	Calculando as Estatísticas com as Funções Embutidas do Excel							
2								
3	Dados							
4	25			No Excel				
5	23		Média Aritmética	23,40740741	<---=MÉDIA(A4:A30)			
6	24		Média Geométrica	23,33324073	<---=MÉDIA.GEOMÉTRICA(A4:A30)			
7	25		Mediana	24	<---=MED(A4:A30)			
8	25		Moda	25	<---=MODO(A4:A30)			
9	26		Percentil- 30	22	<---=PERCENTIL(A4:A30;0,3)			
10	22		Quartil (3)	25	<---=QUARTIL(A4:A30;3)			
11	23							
12	24		Desvio Médio	1,577503429	<---=DESV.MÉDIO(A4:A30)			
13	24		Desvio Padrão	1,886373388	<---=DESVPAD(A4:A30)			
14	25		Variância de Amostra	3,558404558	<---=VAR(A4:A30)			
15	26		Variância de População	3,426611797	<---=VARP(A4:A30)			
16	27		Curtose	-0,768545507	<---=CURT(A4:A30)			
17	23		Distorsão	-0,160300135	<---=DISTRORÇÃO(A4:A30)			
18	21							
19	22							
20	20							
21	22							
22	24							
23	25							
24	22							
25	23							
26	21							
27	24							
28	25							
29	21							
30	20							

1.2.2 - Agrupamentos de Dados e Distribuição de Freqüências

1.2.3 - Classes

1.2.4 – Tabelas – Gráficos – Porcentagens.

1.2.4.1 -Tabela

Exemplos de Tabelas

1.2.4.2 - Gráficos estatísticos

Exercícios Propostos

04 Foi aplicada, em duas séries, A e B, de um colégio X, uma prova com 5 questões. A representação dos resultados da série A foi feita por meio de uma tabela, e os resultados da série B foram apresentados em um gráfico.

Série A – Acertos das questões da prova

07 É dada a tabela abaixo, que corresponde à distribuição de salários de uma empresa A. Complete a tabela com valores percentuais e, em seguida, faça um gráfico de barras **número de funcionários x salários**.

Distribuição de salários