



ESTATÍSTICA APLICADA no Excel

Tópicos de Matemática Aplicada

07/08/2008

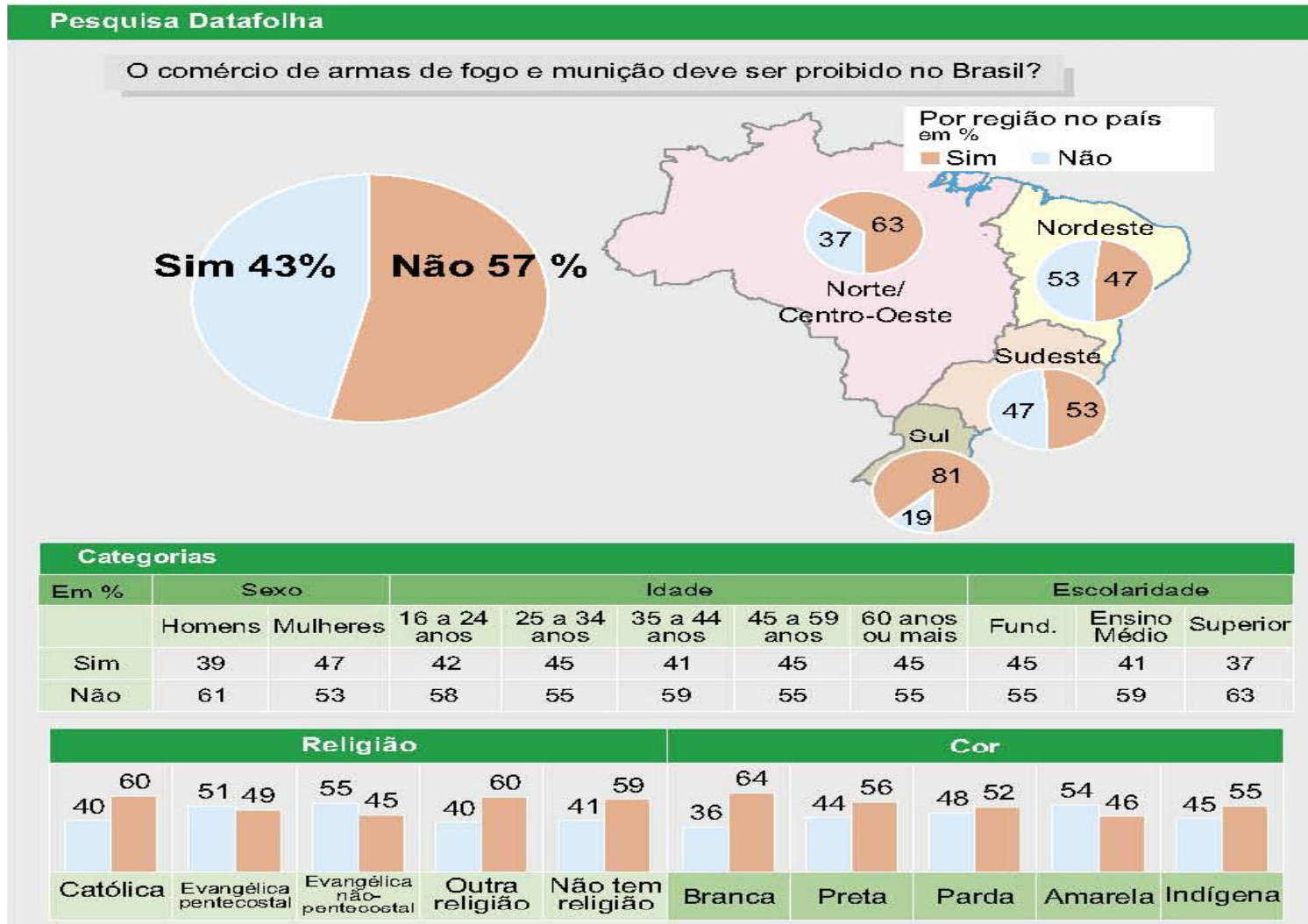
Bertolo

1

O que é Estatística?

- A estatística pode ser entendida como um conjunto de métodos usados para
 1. *coleccionar (planejamento e obtenção),*
 2. *classificar (organizar, resumir, interpretar) e*
 3. *analisar dados (extrair conclusões).*
- A origem da palavra Estatística está associada à palavra *status* (Estado em latim

Exemplo

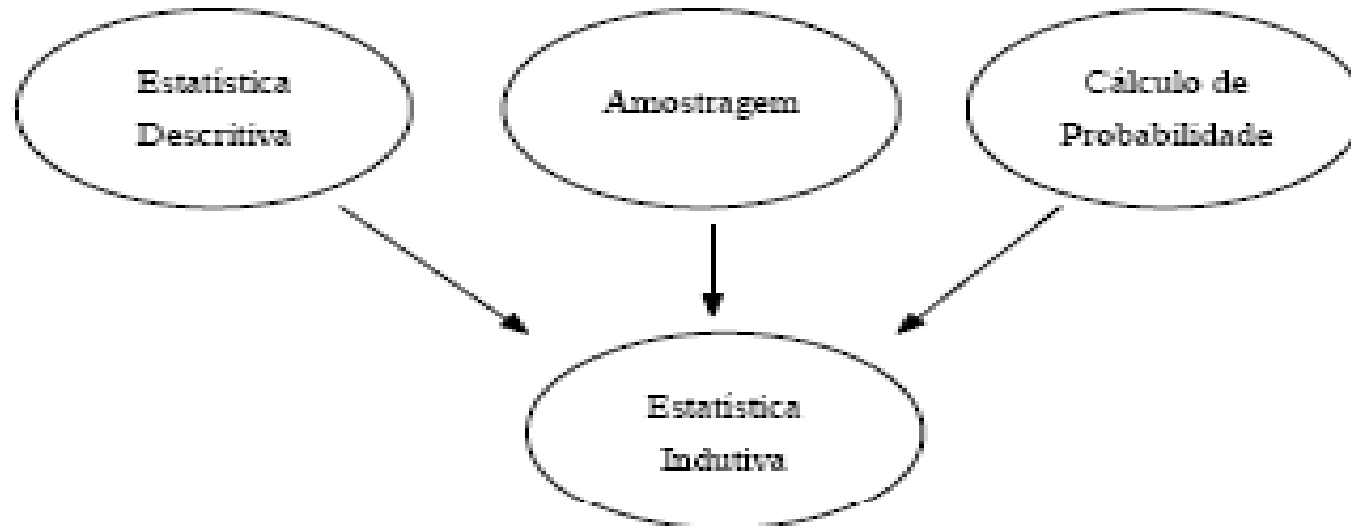


Folha de S. Paulo, 22/10/2005.

Divisões da Estatística

- **ESTATÍSTICA DESCRITIVA**: Envolve a coleta, a organização, a descrição, dos dados e o cálculo e interpretação de coeficientes. Esta parte está associada a cálculos de médias, variâncias, estudo de gráficos, tabelas, etc.. É a parte mais conhecida.
- **ESTATÍSTICA INDUTIVA** ou **INFERENCIAL**: Envolve a análise e a interpretação dos dados (associados a uma margem de incerteza) cujos métodos que se fundamentam na teoria da probabilidade. Nela encontramos a *Estimação de Parâmetros*, *Testes de Hipóteses*, *Modelagens*, etc.

Diagrama



- AMOSTRAGEM: é o ponto de partida (na prática) para todo um Estudo Estatístico.
- Um dos objetivos da análise e interpretação de dados é buscar um modelo para as observações. Estes modelos podem ser essencialmente *determinísticos* ou *não-determinísticos* (*probabilísticos* ou *estocásticos*).

População e Amostra

- *População* é a coleção de todas as observações *potenciais* sobre determinado fenômeno.
- *Amostra* da população é o conjunto de dados efetivamente observados, ou extraídos. Sobre os dados da amostra é que se desenvolvem os estudos, com o objetivo de se fazerem inferências sobre a população.

Os Conflitos e as Contradições

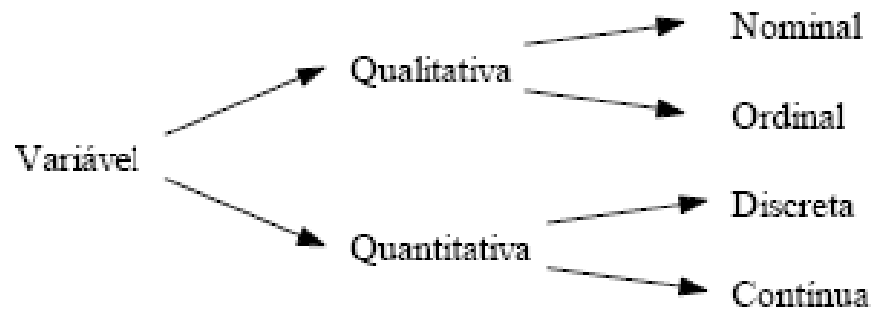
- A inferência estatística permite chegar a conclusões amplas (gerais, sobre o universo estudado) a partir de uma série mais restrita (amostra) de informações (dados estatísticos). Daí a importância do ensino da estatística na formação de várias categorias profissionais de nível superior.
- Por meio da estatística fazemos, por exemplo, melhor análise dos conflitos e contradições que estamos pesquisando no momento e, assim, nós tomamos decisões que irão influenciar a resolução dos conflitos até então discutidos.

Análise Exploratória dos Dados

- Veja como funciona a estatística para que não haja conflitos nem contradições:
- 1º) coleta de dados;
- 2º) apreciação ou crítica dos dados;
- 3º) apuração dos dados;
- 4º) exposição dos dados.
- A pesquisa é feita através de questionários, que podem ser genéricos ou especiais.
- Os erros e inconsistências ocorridos na coleta de dados devem ser corrigidos. As amostras de dados devem ser agrupadas de forma que seu manuseio, visualização e compreensão sejam simplificados.

Tipos de Variáveis

- Os dados coletados em uma primeira fase podem ser definidos como variáveis qualitativas ou quantitativas, de acordo com a seguinte figura:



Exemplo

Para uma população de peças produzidas em um determinado processo, poderíamos ter:

Variável	Tipo
<i>Estado: Perfeita ou defeituosa</i>	Qualitativa Nominal
<i>Qualidade: 1ª, 2ª ou 3ª categoria</i>	Qualitativa Ordinal
<i>Nº de peças defeituosas</i>	Quantitativa Discreta
<i>Diâmetro das peças</i>	Quantitativa Contínua

Agrupamentos de Dados

- Quando se vai fazer um levantamento de uma população, um dos passos é retirar uma amostra desta população e obter dados relativos à variável desejada nesta amostra.
- Cabe à Estatística sintetizar estes dados na forma de tabelas e gráficos que contenham, além dos valores das variáveis, o número de elementos correspondentes a cada variável. A este procedimento está associado o conceito de:
 - **Dados brutos**: É o conjunto de dados numéricos obtidos e que ainda não foram organizados.
 - **Rol**: É o arranjo dos dados brutos em ordem crescente (ou decrescente).
 - **Amplitude** (H): É a diferença entre o maior e o menor dos valores observados.

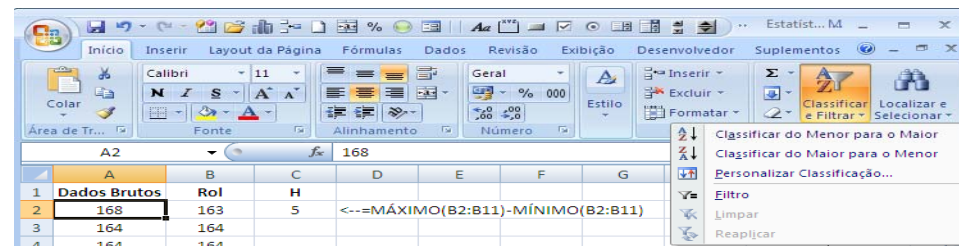
Exemplo


- População = Diâmetro de determinada peça (em mm).
- Dados brutos: { 168, 164, 164, 163, 165, 168, 165, 164, 168, 168 }
- Rol : { 163, 164, 164, 164, 165, 165, 168, 168, 168, 168 }
- $H = 168 - 163 = 5$
- Como Fazer isso no Excel?
- Abra uma pasta e na Plan1 insira o intervalo dos dados:

	A
1	Dados Brutos
2	168
3	164
4	164
5	163
6	165
7	168
8	165
9	164
10	168
11	168

Agora copie o intervalo para B1, re-nomeando o título para ROL.

A seguir procure classificar os dados em ordem crescente, usando a guia **Início** e o grupo **Edição**. Assim:



Clicando no botão  **Classificar de A a Z** teremos um intervalo de dados ordenados do menor para o maior.

Exemplo

- População = Diâmetro de determinada peça (em mm).
- Dados brutos: { 168, 164, 164, 163, 165, 168, 165, 164, 168, 168 }
- Rol : { 163, 164, 164, 164, 165, 165, 168, 168, 168, 168 }
- $H = 168 - 163 = 5$
- Como encontrar a **Amplitude** no Excel?

	A	B
1	Dados Brutos	Rol
2	168	163
3	164	164
4	164	164
5	163	164
6	165	165
7	168	165
8	165	168
9	164	168
10	168	168
11	168	168

Dê o título **H** para a célula C1.

Na célula C2 introduza a função composta
=MÁXIMO(B2:B11)-MÍNIMO(B2:B11)

O resultado final de sua planilha será:

	A	B	C	D	E	F	G
1	Dados Brutos	Rol	H				
2	168	163	5	<---=MÁXIMO(B2:B11)-MÍNIMO(B2:B11)			
3	164	164					
4	164	164					
5	163	164					
6	165	165					
7	168	165					
8	165	168					
9	164	168					
10	168	168					
11	168	168					

Freqüências

- **Freqüência absoluta** (n_i): É o número de vezes que um elemento aparece na amostra.

Temos,
$$\sum_i^k n_i = n$$

, onde n é o número total de dados da amostra e k é o número de valores diferentes na amostra.

- **Freqüência Relativa** (f_i): $f_i = \frac{n_i}{n}$ e $\sum_i^k f_i = 1$

- **Freqüência Absoluta Acumulada** (N_i): É a soma da freqüência absoluta do valor da variável i com todas as freqüências absolutas anteriores.

- **Freqüência Relativa Acumulada** (F_i): $F_i = \frac{N_i}{N}$

Exemplo

- População = Diâmetro de determinada peça (em mm).
- Dados brutos: { 168, 164, 164, 163, 165, 168, 165, 164, 168, 168 }
- Rol : { 163, 164, 164, 164, 165, 165, 168, 168, 168, 168 }
- $H = 168 - 163 = 5$
- Como calcular essas freqüências no Excel?
- Dê os títulos x_i e freqüência (n_i) para as células D1 e E1 respectivamente.
- Com o cursor na célula D2 insira =B2.
- Pule para a célula D3 e introduza a fórmula:
=SE(B3:\$B\$11<>B2;B3;"")
- Arraste pela alça, copiando esta fórmula até a D11 e a sua planilha ficará assim:

Continuação

	A	B	C	D	E
1	Dados Brutos	Rol	H	xi	freqüência(ni)
2	168	163	5	163	
3	164	164		164	
4	164	164			
5	163	164			
6	165	165		165	
7	168	165			
8	165	168		168	
9	164	168			
10	168	168			
11	168	168			

Como juntar os valores em células adjacentes? Ajude o Bertolo a fazer isso e serás recompensado!!! Pesquise kra.... Você não quer ser cientista?

Calculando as freqüências no Excel

- Agora selecione a célula E2 e introduza a função:
- **=FREQUÊNCIA(B2:B11;D2:D12)** e não dê **ENTER** não!. Calma!!!!
- Agora pressione **F2** e após, ao mesmo tempo, **CTRL +SHIFT + ENTER**.
- O que aconteceu?
- Apareceu 1 em E2 e olhando para a barra de fórmulas vemos:
- **{=FREQUÊNCIA(B2:B11;D2:D12)}** , a fórmula ficou entre chaves. O que significa isto?
- O Excel reconheceu a fórmula como uma matriz (array).
- Você viu que o intervalo foi de D2 até D12. É isso mesmo. A explicação encontra-se no Apêndice da apostila.
- A coisa agora ficou assim:

Calculando as freqüências no Excel

The screenshot shows the Excel interface with the following data in the spreadsheet:

	A	B	C	D	E	F	G	H	I	J
1	Agora	Rol	H	xi	freqüência(ni)					
2	168	163	5	163	1					
3	164	164		164						
4	164	164								
5	163	164								
6	165	165		165						
7	168	165								
8	165	168		168						
9	164	168								
10	168	168								
11	168	168								
12										

Agora é só arrastar pela alça até D11 e obter:

Calculando as freqüências no Excel

	A	B	C	D	E
1	Agora	Rol	H	xi	freqüência(ni)
2	168	163	5	163	1
3	164	164		164	3
4	164	164			4
5	163	164			3
6	165	165		165	2
7	168	165			5
8	165	168		168	4
9	164	168			0
10	168	168			0
11	168	168			0
12					0

Agora **delete** os valores das células E's que não correspondem a valores das células D's

Para encontrar o total introduzimos na célula E12 a função SOMA:
=SOMA(E2:E11)

	A	B	C	D	E
1	Dados Brutos	Rol	H	xi	freqüência(ni)
2	168	163	5	163	1
3	164	164		164	3
4	164	164			
5	163	164			
6	165	165		165	2
7	168	165			
8	165	168		168	4
9	164	168			
10	168	168			
11	168	168			
12				Total	10

Resultado Final

	A	B	C	D	E	F	G	H
1	Dados Brutos	Rol	H	X_i	freqüência(n_i)	fi	Ni	Fi
2	168	163	5	163	1	0,1	1	0,1
3	164	164		164	3	0,3	4	0,4
4	164	164		165	2	0,2	6	0,6
5	163	164		168	4	0,4	10	1,0
6	165	165			0			
7	168	165						
8	165	168						
9	164	168						
10	168	168						
11	168	168						
12				Total	10	1		

A fórmula no exemplo precisa ser inserida como uma fórmula de matriz. Selecione o intervalo E19:E23. Pressione F2, insira a função FREQUÊNCIA e, em seguida, pressione CTRL+SHIFT+ENTER. Se não for inserida como uma fórmula de matriz, haverá somente um resultado na célula E19 (1).

E aí? Quebre a cabeça e preencha (calculando no Excel) as colunas F, G e H

Classes

As classes são um artifício para condensar o número de elementos diferentes de uma amostra. Imagine construir uma tabela para 200 valores diferentes, nos moldes do problema anterior?

Os principais pré-requisitos para uma boa definição de classes em um conjunto de dados são:

- a) As classes devem abranger todas as observações;
- b) O extremo superior de uma classe é o extremo inferior da classe subsequente (simbologia: $\left[\text{---} \right)$, intervalo fechado à esquerda e aberto à direita);
- c) Cada valor observado deve enquadrar-se em apenas uma classe;
- d) $k \leq 25$, de um modo geral, sendo k o número de classes;
- e) As unidades das classes devem ser as mesmas dos dados.

Cálculo de k (opções não rígidas):

- Fórmula de Sturges: $k = 1 + \log_2 N = 1 + 3,32 \log N = 1 + 1,44 \ln N$

- $k \cong \sqrt{N}$

Obs.: N é o número de elementos diferentes da amostra e em muitas vezes pode ser considerado $N = n$

Geralmente, temos ainda:

- Intervalo da classe (h): $h \cong H / k$
- Ponto médio da classe (x_i): Ponto médio entre o limite inferior e o limite superior de cada classe .

Exemplo

- A tabela seguinte registra uma amostra aleatória de tamanho vinte e cinco das Vendas diárias em milhares de uma empresa. Construir a **distribuição de freqüências absolutas**.

280	305	320	330	310	340	330	341	369	355	370	360	370
365	280	375	380	400	371	390	400	370	401	420	430	

Solução

- Abrindo uma pasta para este exercício, escolhemos Plan1 e introduza os dados brutos, classifique em ordem crescente e determine o valor máximo, mínimo, o tamanho da amostra, o número de classes e a amplitude total. No final a sua planilha deverá ficar parecida com esta:

	A	B	C	D	E	F	G	H
1	Funções Estatísticas							
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						
10	330	330						
11	380	340						
12	310	341						
13	400	355						
14	340	360						
15	371	365						
16	330	369						
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						

Solução - Continuação

- Agora vamos agrupar os dados em classes e calcular as freqüências de valores em cada classe. Para isso, coloque os títulos Classe, Limite Inferior e Limite Superior, respectivamente, nas células **C11**, **D11** e **E11**. A seguir, introduza os rótulos dos intervalos de classe de **C12** até **C16**, como mostrado na planilha abaixo. Use a sua criatividade e espírito de pesquisa e encontre uma maneira para o Excel fazer este preenchimento automaticamente. Depois coloque o cursor na célula **D12** e introduza a fórmula **=D5**. Na célula **D13**, introduza a seguinte fórmula: **=D12+\$D\$9** e arraste a alça até **D16**. Na célula **E12**, coloque a fórmula: **=D12+\$D\$9-0,01**. Por que esse tal 0,01? Para o intervalo ficar aberto à direita. Arraste o resultado até **E16**. Agora arredonde o resultado 429,99 para 430, para tomar esse valor e os outros superiores a ele.
- Agora selecione o intervalo **F12** a **F16** e introduza a fórmula: **=FREQUÊNCIA(B4:B28;E12:E\$17)**. Não aperte o ENTER. Calma, tá com pressa? Vá pescar... Pressione depois a **tecla F2** e, ao mesmo tempo, as teclas **CTRL + SHIFT + ENTER**, para transformar numa fórmula de matriz. Você agora obterá os resultados da planilha abaixo.

Solução - Continuação

	A	B	C	D	E	F	G
1	Funções Estatísticas						
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	Classe	Limite Inferior	Limite Superior		
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					

Tabelas

- É uma pequena tábua, ou quadro, composta de linhas e colunas, onde se registram dados de pessoas ou de coisas.
- A construção de uma tabela depende dos dados coletados que serão resumidos e dispostos em forma tabular, o que significa dizer que **são colocados em série e apresentados em quadros ou tabelas.**
- Tabela é a disposição gráfica das séries de acordo com uma determinada ordem de classificação.
- As partes principais de uma tabela são:
 - – Cabeçalho (com o título);
 - – Corpo (com as informações organizadas em linhas e colunas);
 - – Rodapé (com a fonte e eventuais legendas).

Exemplos de Tabelas

- **Tabela 1** – Série geográfica ou territorial, cujos dados são dispostos em correspondência com a região geográfica, isto é, variam com o local.

População brasileira de alguns estados – 1970	
Estados	Número de Habitantes
Bahia	7.583.140
Rio de Janeiro	4.794.578
São Paulo	17.958.693
Paraná	6.997.682

Exemplos de Tabelas

- Tabela 2 – Série específica ou categórica, cujos dados são dispostos de acordo com itens específicos ou categorias.

Índice de Preços ao Consumidor Semanal (IPC-S), itens com maiores influências positivas, variações percentuais ao mês		
Item	Variação percentual ao mês até	
	08-05-2004	17-05-2004
Mamão da Amazônia (papaia)	23,64	33,34
Tarifa elétrica residencial	1,58	1,61
Batata - inglesa	20,12	15,42
Manga	15,63	23,14
Leite tipo longa vida	3,25	3,25
Tomate	-3,97	9,76
Plano e seguro saúde	0,72	0,72
Cebola	1,69	8,38
Vasodilatador para pressão arterial	2,47	2,63
Açúcar cristal	8,68	8,26

Exemplos de Tabelas

- Tabela 3 – Série categórica.

Índice de Preços ao Consumidor Semanal (IPC-S), itens com maiores influências negativas, variações percentuais ao mês		
Item	Variação percentual ao mês até	
	08-05-2004	17-05-2004
Passagem aérea	-6,78	-6,81
Melão	-1,19	-23,01
Laranja-pêra	-1,90	-3,64
Frango em pedaços	2,03	-1,83
Pimentão	3,76	-4,70
Maçã argentina	-6,95	-8,22
Contra-filé	-3,69	-2,93
Corvina	-2,83	-5,12
Óleo de soja	-0,16	-1,45
Maçã nacional	-7,98	-5,41

POVIBRE – Divisão de Gestão de Dados

Exemplos de Tabelas

- Tabela 4 – Série geográfica

Índice de Preços ao Consumidor Semanal (IPC-S), segundo municípios de 12 capitais, variações percentuais ao mês		
Município	Variação percentual ao mês até	
	08-05-2004	17-05-2004
IPC-S	0,52	0,59
Belém	0,59	0,50
Belo Horizonte	1,05	0,99
Curitiba	0,96	1,37
Distrito Federal	0,35	0,54
Florianópolis	0,86	0,85
Fortaleza	0,89	0,98
Goiânia	0,56	1,06
Porto Alegre	0,37	0,58
Recife	0,14	0,36
Rio de Janeiro	0,48	0,59
Salvador	0,59	0,57
São Paulo	0,30	0,27

POVIERE - Divisão de Gestão de Dados

Gráficos Estatísticos

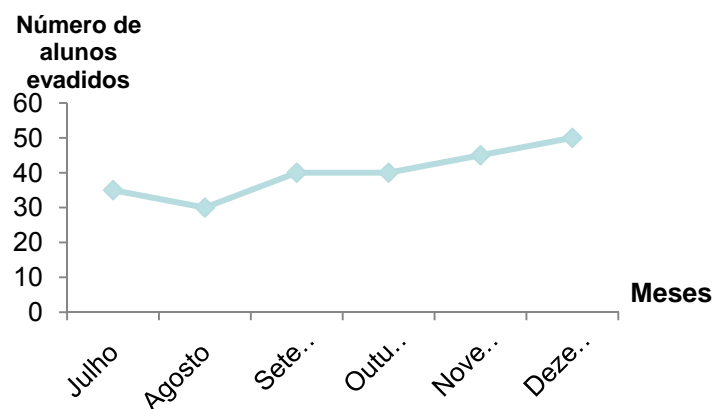
- Os gráficos também são uma forma de apresentação dos dados coletados na pesquisa. O gráfico precisa ter clareza, simplicidade e veracidade.
- Existem vários tipos de gráficos: gráfico de segmento, de setor, de barras e de colunas.
- A palavra, gráfico, refere-se à grafia, ou às artes gráficas, ou ao que delas se ocupa. Um gráfico pode ser representado por desenho ou figuras geométricas.
- Os principais gráficos serão sucintamente descritos a seguir. É importante lembrar que os modernos programas computacionais de Edição de Texto, Planilha Eletrônica e Banco de Dados facilitam em muito a manipulação com gráficos. Dentre esses programas o destaque é o Excel.

Gráfico de Linha

- A tabela seguinte mostra o número de alunos evadidos de uma determinada escola de Ensino Médio no segundo semestre de 2002 (uma série temporal, cujos dados são dispostos de acordo com o tempo).

Meses do 2º semestre	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Nº de alunos evadidos	35	30	40	40	45	50

- De acordo com a tabela, construiremos um gráfico de segmentos cujos pares ordenados serão (julho, 35), (agosto, 30), (setembro, 40), (outubro, 40), (novembro, 45) e (dezembro, 50).



Os gráficos de linhas são muito utilizados para mostrar a evolução durante um certo período (séries temporais). O gráfico permite visualizar muito bem o crescimento, o decréscimo ou a estabilidade do objeto a ser analisado.

Gráfico de Barras

- No geral, é usado para comparar as freqüências de mesma variável em um determinado momento.
- Freqüência é o número de vezes que cada resposta ou cada intervalo de resposta aparece na tabela do pesquisador, a qual pode ter sido feita por amostragem.

A opinião dos paulistanos sobre o salário mínimo (em porcentagem)

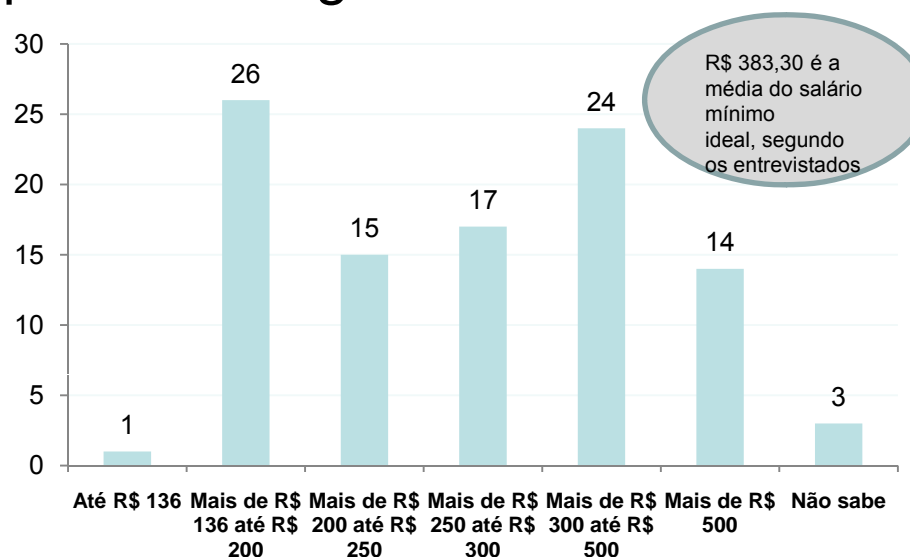


Gráfico de Setores

- O gráfico de setores também é muito usado e faz enxergar melhor as partes do todo (representadas em porcentagens). Para ser traçado, é necessário saber que o círculo todo tem 360° , o que corresponde a 100%. De acordo com a pesquisa, o círculo será repartido, efetuando-se uma regra de três simples.
- Para a construção do gráfico de setores, são necessários a pesquisa, a tabela, a fonte, o compasso, o transferidor e as cores.
- Exemplo - O gráfico de setores a seguir foi construído com os resultados de uma pesquisa que perguntou a 1.500 pessoas qual jornal diário elas mais gostam de ler:

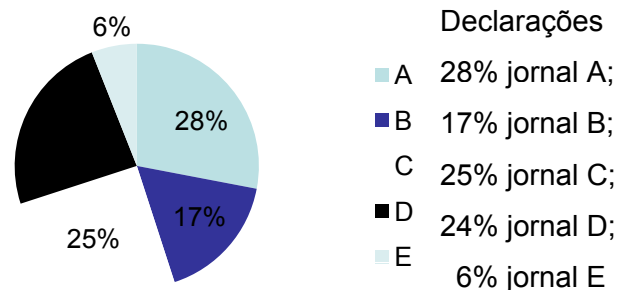
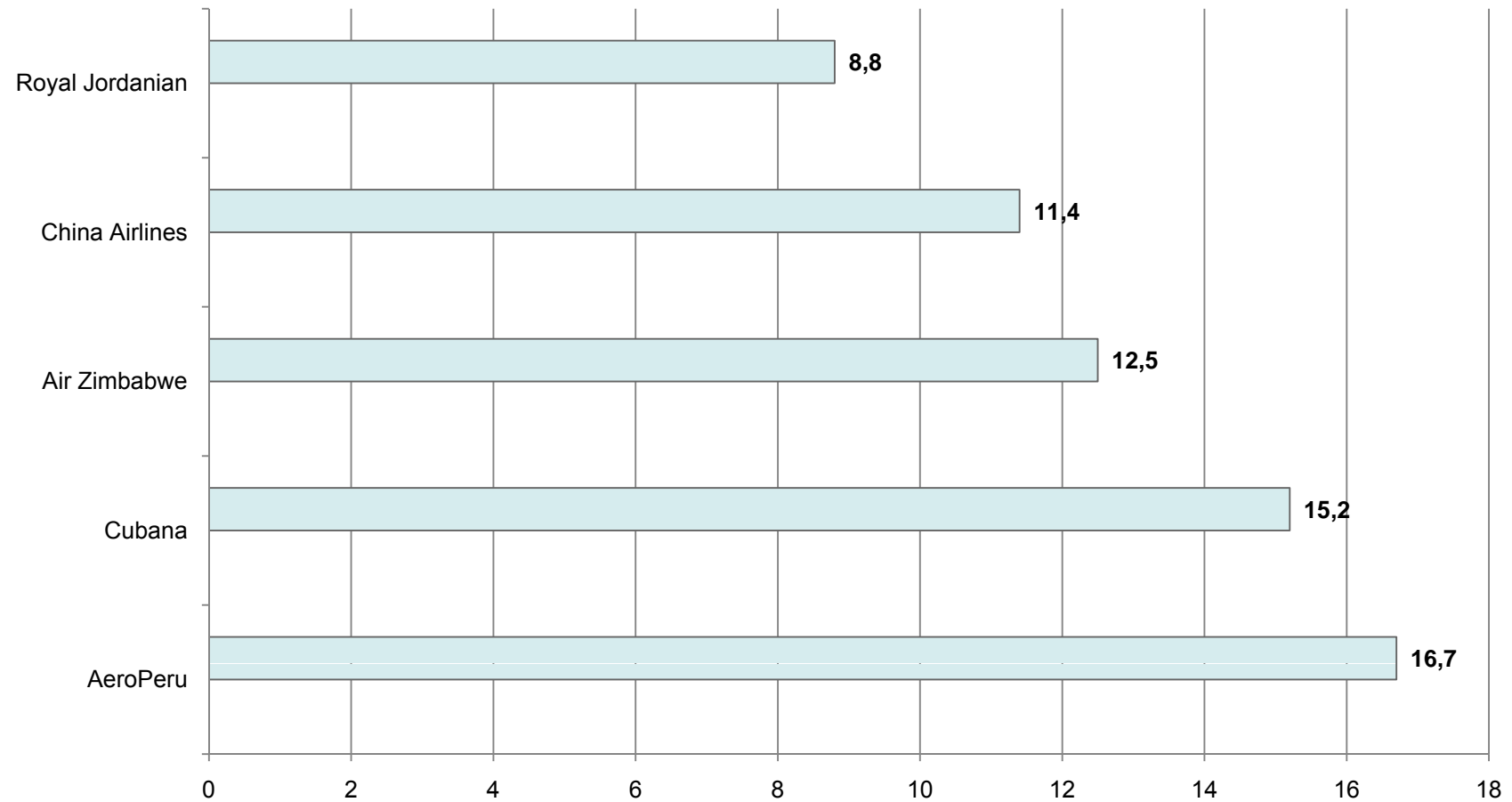


Gráfico de Barras Horizontais

Número de acidentes por milhão de decolagens



Tarefa para casa

- **01** Numa pequena empresa com 20 funcionários a distribuição dos salários, em reais, é a seguinte:
- 800,00, 800,00, 12.000,00, 12.000,00, 800,00, 600,00, 1.300,00, 600,00, 800,00, 1.000,00, 500,00, 600,00, 700,00, 800,00, 700,00, 12.000,00, 500,00, 600,00, 800,00 e 500,00.

Coloque os valores em ordem crescente e, em seguida, monte uma tabela.

Salário dos Funcionários	Freqüência
500,00	??????????
12.000,00	????????????????

Cap 02 - Medidas Estatísticas

- A redução dos dados através de tabelas de frequências e gráficos fornece muito mais informações sobre o comportamento de uma variável do que a própria série original de dados. Contudo, muitas vezes queremos resumir ainda mais esses dados, apresentando um ou alguns valores que sejam representativos da série toda. Quando usamos um só valor, obtemos uma redução drástica dos dados.
- As principais medidas estatísticas (ou simplesmente estatísticas) referem-se às **medidas de posição** (locação ou tendência central) ou às **medidas de dispersão** (ou variabilidade):

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 frequência.
- 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 da **média aritmética, da moda ou da mediana**, e 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.

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.

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

2.1.2 – Média Aritmética Ponderada (\bar{x}_p)

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

$$\bar{x} = \frac{p_1x_1 + p_2x_2 + p_3x_3 + \dots + p_nx_n}{p_1 + p_2 + \dots + p_n} = \frac{\sum_{i=1}^n p_i x_i}{\sum_{i=1}^n p_i}$$

p_i = peso da amostra x_i .

Para dados agrupados em classes, temos:

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

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 ().

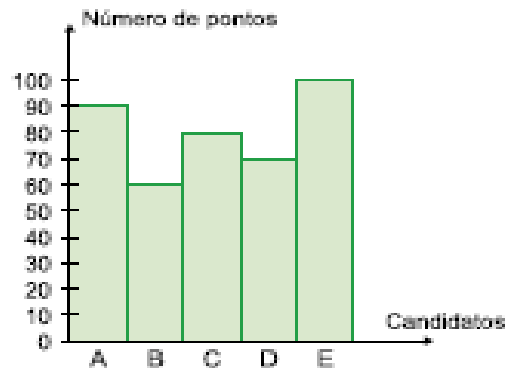
Média Geométrica (X_G)

- É definida como a raiz de ordem n do produto desses números.

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$$

Exercícios Modelos

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



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

b. 80 pontos

c. Mais pontos E e menos pontos B

Exercício 2

- 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.
- Para resolver esse exercício, desconsidere a média ponderada.
média final do aluno foi 6,5.
- $\bar{x}_p = (6 \cdot 2 + 5 \cdot 4 + 3 \cdot 2 + 10 \cdot 4) / (2 + 4 + 2 + 4)$

Exercício 3

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:

- a. R\$ 2.637,00
- b. R\$ 2.500,00
- c. R\$ 2.420,00
- d. R\$ 2.400,00

Os valores centrais das classes 1, 2, 3 e 4 são, respectivamente, 1.500, 2.500, 3.500 e 4.500 reais.

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

Médiana ()

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 x_{\sim}

Para dados não agrupados em classes:

Se n é **ímpar** : é o termo

$$x_{\sim} = \left(\frac{n+1}{2} \right)^{\text{º}}$$

Se n é **par** → é o termo

$$x_{\sim} = \frac{\frac{n}{2} \text{ termo} + \left(\frac{n}{2} + 1 \right) \text{ termo}}{2}$$

EX1: Em um colégio, estão matriculados numa determinada classe 21 alunos. Durante o 1º bimestre foi feito um levantamento da freqüência 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 das faltas? Resposta: 3

EX2: 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

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.”
- **Ex:** Para valores de 51 a 100, ordenados crescentemente:
- $P_{25} = 25$ 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:

$$P_{80} = \frac{(90 + 91)}{2} = 90,5$$

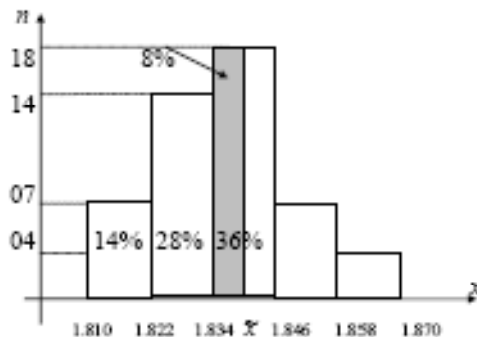
Percentis de dados agrupados

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

Ex.: Dada a distribuição de freqüência 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



$$\bar{x} = (62 + 310 + 621 + 201) / 4 = 1.753 / 4 = 438,2!$$

Um 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.

Moda e Classe Modal

- 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.
- **Exemplos**
- **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: Portanto bimodal

Mais Exemplos

3. 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.

4. Tempo, em anos, que um grupo de pessoas frequentou 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 e 5, portanto o conjunto de dados é bimodal.

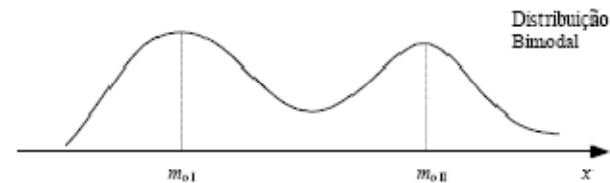
Moda de Dados Agrupados

Nota importante

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

Quando os dados estão agrupados em classes,

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



Exemplo

Nesta série, 1, 7, 9, 12, 17, não há moda, pois não há repetição de número.

Mais Exemplos de Moda

Considere os números 621, 310, 621, 201 e calcule:

a) a média aritmética ();

b) a média aritmética ponderada (), com pesos 2, 3, 1 e 2, respectivamente;

c) a moda (Mo).

Resposta

Primeiramente, monta-se a tabela:

Números	621	310	201
Freqüência	2	1	1

$$\bar{x} = (621 + 310 + 621 + 201) / 4 = 1.753 / 4 = 438,25$$

$$x = (621 \cdot 2 + 310 \cdot 3 + 621 \cdot 1 + 201 \cdot 2) / (2 + 3 + 1 + 2) = 3.195 / 8 = 399,375$$

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 $Mo = 621$.

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.

2.2 – Medidas de Dispersão ou Variabilidade – Cont...

- 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.

2.2 – Medidas de Dispersão ou Variabilidade – Cont...

- 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

Pessoas	Peso (kg)
Agulha	30
Aderbal	15
Corá	55
Renato	52
Guilherme	60
Bruno	53
Bertolo	75
Alexandre	20
Fábio Thomáz	40

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

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 negativa. E o desvio médio, então, passa a ser encontrado a partir da média aritmética de todos os desvios.

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

Desvio Médio - Exemplo

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

Com os dados do exercício anterior, temos:

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

Desvio Médio

$$\begin{aligned} & \frac{|30 - 44,4| + |15 - 44,4| + |55 - 44,4| + |52 - 44,4| + |60 - 44,4| + |53 - 44,4| + |75 - 44,4| + |20 - 44,4| + |40 - 44,4|}{9} \\ & = 16,17 \end{aligned}$$

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

2.2.2 – 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, toma-mos 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.

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

Variância - Exemplo

Variância

$$= \frac{(30 - 44,4)^2 + (15 - 44,4)^2 + (55 - 44,4)^2 + (52 - 44,4)^2 + (60 - 44,4)^2 + (53 - 44,4)^2 + (75 - 44,4)^2 + (20 - 44,4)^2 + (40 - 44,4)^2}{9}$$

$$= 345,57$$

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 m² (metro quadrado), que representa áreas. E é aí que entra o desvio-padrão, ou seja, extraíndo a raiz quadrada da variância.

$$\text{Desvio - padrão } \sigma = \sqrt{\text{Variância}}$$

Desvio Padrão - Exemplo

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

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 – Momentos de uma distribuição de freqüências

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 ,
dado por:

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

Mais Momentos

- Conforme já vimos nos casos da média e da variância, as expressões precedentes podem ser reescritas levando-se em consideração as freqüências 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 $M_1 = \bar{x}$; $m_1 = 0$; $m_2 = \sigma^2$.

2.2.5 – 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.

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

Exemplo: 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.