

INSTITUTO FEDERAL
MINAS GERAIS

“Erros são, no final das contas, fundamentos da verdade.
Se um homem não sabe o que uma coisa é,
já é um avanço do conhecimento saber o que ela não é.”

Carl Jung (1875-1961)

ESTATÍSTICA

Conceitos básicos da Estatística

A **estatística** é uma parte da Matemática Aplicada que pode ser entendida como um conjunto de métodos empregados no planejamento de experimentos, na obtenção, organização e resumo de dados coletados, bem como na análise e interpretação de tais dados, a fim de que conclusões possam ser tiradas.

Em geral, as fases de coleta, organização e descrição dos dados ficam a cargo da **estatística descritiva**, enquanto que a análise e interpretação de tais dados são de competência da chamada **estatística indutiva** ou **inferencial**, objeto de estudo de grande parte dos cursos superiores e programas de pós-graduação.

A fim de que o leitor possa se familiarizar com este importante ramo da Matemática, algumas definições prévias se fazem necessárias, a saber:

População e amostra

Se desejarmos saber, por exemplo, qual o esporte favorito entre os alunos de uma turma, podemos consultar todos os alunos da turma e obter tal informação.

Tal procedimento, entretanto, não é possível (ou não é viável), caso queiramos pesquisar a respeito da intenção de voto dos eleitores do Brasil, com relação à eleição para presidente da República.

Entre os fatores que inviabilizam a pesquisa acerca da intenção de voto, acima descrita, estão: o *tempo* e o *dinheiro* gastos com a sua realização.

Felizmente, resultados confiáveis (o mais próximos o possível da realidade) podem ser obtidos desde que a pesquisa seja realizada com uma parte representativa da população à qual pertence, cuidadosamente selecionada.

Uma **população estatística** (ou universo estatístico) pode ser entendida como um conjunto de entes portadores de, no mínimo, uma característica comum.

Já uma **amostra** é um subconjunto finito, cujos elementos são extraídos de uma população que se deseja estudar.

Variável

Uma fábrica de automóveis deseja lançar um novo modelo que seja do agrado do público.

Para tanto, contrata uma pesquisa de mercado a fim de conhecer as preferências dos consumidores no que diz respeito à motorização, itens de segurança, cor da pintura e material utilizado no revestimento dos bancos.

Cada uma das características consultadas na pesquisa constitui uma **variável**, que pode ser entendida como o conjunto de resultados possíveis de um fenômeno.

Por exemplo, no caso da variável “cor da pintura”, preta, prata, vermelha e azul poderiam ser os resultados possíveis.

Vale ressaltar que os valores que uma variável pode assumir não são necessariamente numéricos, o que nos remete à seguinte classificação:

- **Variáveis qualitativas** – são aquelas cujos valores podem ser separados em diferentes categorias que se distinguem por alguma característica não numérica. Por exemplo: sexo (masculino e feminino), cor dos cabelos (preto, loiro, ruivo, castanho, etc). Podem ser subdivididas em:
 - **Variáveis qualitativas ordinais** – quando existe uma ordem nos seus valores. Por exemplo, a variável “Grau de instrução” pode ter seus valores ordenados (fundamental, médio, superior, etc). O mesmo não ocorre com a variável “cor da pele”.
 - **Variáveis qualitativas nominais** – quando uma ordem não pode ser estabelecida entre seus valores.
- **Variáveis quantitativas** – São aquelas cujos valores são expressos em números. As variáveis quantitativas se subdividem, ainda, em:
 - **Variáveis quantitativas discretas** – quando resultam de um conjunto finito (ou enumerável) de valores possíveis. Por exemplo, o número de livros de Matemática da biblioteca da escola e a quantidade de carros presentes, agora, num estacionamento, são variáveis discretas.
 - **Variáveis quantitativas contínuas** – quando resultam de um número infinito de valores possíveis que podem ser associados a pontos em uma escala contínua, de modo que não haja lacunas ou interrupções. Por exemplo, o peso de um pacote de arroz e o comprimento de um parafuso constituem variáveis contínuas.

Em geral, adotamos o seguinte critério de decisão:

Quando os dados representam *contagens*, são discretos; caso sejam resultantes de *medições*, são contínuos.

Questões propostas

Q1. Classifique as variáveis abaixo em qualitativas ou quantitativas (contínuas ou discretas):

- a) Número de partidas num torneio esportivo: _____
- b) Diâmetro interno de uma tubulação: _____
- c) cor dos cabelos: _____
- d) Número de filhos de um casal: _____
- e) Número de ações negociadas num dia na bolsa de valores: _____
- f) Teor alcoólico de uma bebida: _____
- g) Sexo de um filhote nascido no zoológico: _____

Distribuição de frequências

Denomina-se **frequência absoluta** o número de vezes que um valor da variável ocorre numa pesquisa.

Há, ainda, a **frequência relativa**, que indica, em geral, na forma de porcentagem, a razão entre a frequência absoluta e o total de elementos integrantes da amostra considerada na pesquisa.

A título de exemplo, suponhamos que, numa locadora de DVD's, em certa sexta-feira, tenha sido feita uma pesquisa com os clientes presentes, acerca do gênero de filme predileto, e que o resultado tenha sido o seguinte:

Das pessoas consultadas, 4 preferem filmes de terror, 6 preferem assistir a filmes de romance, 7 têm preferência por filmes de suspense, 5 têm predileção por filmes de ação e 4 só gostam de comédia.

No exemplo apresentado, a variável qualitativa é “gênero de filme predileto” e seus valores são “terror”, “romance”, “suspense”, “ação” e “comédia”.

Consoante as respostas dadas, pode-se dizer que a frequência absoluta do valor suspense é 7, enquanto que o valor comédia tem frequência absoluta 4.

A soma das frequências absolutas dos valores declarados para a variável “gênero de filme predileto” fornece o total de elementos da amostra, que, nesse caso, é igual a 26.

Portanto, a frequência relativa do valor “terror” é $\frac{4}{26} \approx 15,38\%$.

Já a frequência relativa do valor “ação” é $\frac{5}{26} \approx 19,23\%$

Tabela de frequências

Uma tabela de frequências relaciona categorias (ou classes) de valores com as contagens (frequências) do número de valores que se enquadram em cada categoria.

As tabelas de frequências constituem importante instrumento de organização de dados brutos.

Os dados coletados sobre o gênero de filme predileto presentes no exemplo anterior deram origem à seguinte tabela:

Gênero de filme predileto	f	f _r (%)
Terror	4	15,385
Romance	6	23,077
Suspense	7	26,923
Ação	5	19,231
Comédia	4	15,385
Total	26	100

Quando a variável assume muitos valores diferentes, fica inviável a construção de uma tabela de distribuição de frequências com uma linha para cada valor.

Nesse caso, pode-se agrupar valores próximos em intervalos e construir uma tabela de distribuição de frequências com intervalos de classe.

É comum utilizar-se intervalos reais fechados à esquerda e abertos à direita, que podem ser representados por $[a; b[$ ou por $a |— b$, tal que $a \leq x < b$, sendo que a amplitude do intervalo $a |— b$ é dada pela diferença $b - a$.

Construindo uma tabela de distribuição de frequências com intervalos de classe

- As notas obtidas por 50 alunos de uma turma foram:

84 68 33 52 47 73 68 61 73 77
 74 71 81 91 65 55 57 35 85 88
 59 80 41 50 53 65 76 85 73 60
 67 41 78 56 94 35 45 55 64 74
 65 94 66 48 39 69 89 98 42 54

Forme, com tais notas, uma tabela de distribuição de frequências com intervalos de classe, contendo 7 classes.

Inicialmente determinaremos a amplitude do conjunto de dados, calculando a diferença entre o maior e o menor valor observado da variável. Nesse caso, $98 - 33 = 65$.

Como fora solicitado que a tabela contenha 7 classes (intervalos), determinaremos a amplitude de cada intervalo de classe dividindo 65 por 7. Geralmente arredondamos o valor obtido “para mais”, a fim de garantir que todos os valores sejam incluídos na tabela. Dessa forma, a amplitude de classe será igual a 10.

Em seguida, escolheremos um limite inferior para a primeira classe, que deve ser o menor valor observado ou um valor ligeiramente inferior a ele. Optaremos, nesse exercício, pelo valor 30.

Dessa forma, o primeiro intervalo de classe será $30 \text{ |---} 40$.

Somando a amplitude de classe (10) com o número escolhido como ponto de partida (30), obtemos, assim, o limite inferior da segunda classe. No caso desse exemplo, o limite inferior da segunda classe será $30 + 10 = 40$ e, por conseguinte, o segundo intervalo de classe será $40 \text{ |---} 50$.

Adicionando a amplitude de classe ao segundo limite inferior, obtemos o terceiro limite e repetimos o processo até obtermos o último intervalo de classe.

Em cada linha da tabela, na coluna correspondente à frequência absoluta, indicamos a quantidade de notas pertencentes ao correspondente intervalo.

Assim, obtemos a seguinte tabela de distribuição de frequências com intervalos de classe:

Classe	Nota	f_i	$f_r(\%)$
1	$30 \text{ ---} 40$	4	8
2	$40 \text{ ---} 50$	6	12
3	$50 \text{ ---} 60$	9	18
4	$60 \text{ ---} 70$	11	22
5	$70 \text{ ---} 80$	9	18
6	$80 \text{ ---} 90$	7	14
7	$90 \text{ ---} 100$	4	8
		$\Sigma = 50$	100%

Observação: Embora não haja uma regra única a ser utilizada na determinação do número de classes de uma tabela, sugere-se que o número de classes deva ficar entre 5 e 20, sempre primando pela construção de uma tabela com valores convenientes e compreensíveis. Recomenda-se, no entanto, a construção de classes com mesma amplitude.

Questões propostas

Q2. (UFPB) A tabela abaixo apresenta o percentual de candidatos por faixa de pontuação, na prova discursiva de Matemática de PSS-2005/UFPB.

Pontos	%
0	10,1
1 a 4	36,3
5 a 8	31,3
9 a 12	13,2
13 a 16	5,6
17 a 20	2,6
21 a 24	0,9

Fonte: Coperve/UFPB

Com base nesses dados, é correto afirmar:

- a) Mais de 10% obtiveram, no mínimo, 13 pontos.
- b) No máximo, 40% obtiveram até 4 pontos.
- c) Mais de 70% obtiveram, no máximo, 8 pontos .
- d) Mais de 3% obtiveram de 17 a, no máximo, 20 pontos.
- e) Mais de 4% obtiveram de 17 a 24 pontos.

Q3. (UEL-PR) De acordo com os dados apresentados pela tabela, é correto afirmar:

Rendimento/hora em reais, segundo nível de instrução e sexo
Regiões Metropolitanas – 1999

	Analfabeto	Ensino Fundamental incompleto	Ensino Fundamental completo	Ensino médio incompleto	Ensino médio completo	Ensino superior	Total
Distrito Federal							
Homens	1,98	2,47	3,57	3,49	6,68	16,64	6,56
Mulheres	-(1)	1,51	2,04	2,19	4,47	21,52	4,87
Porto Alegre							
Homens	1,69	2,42	3,01	2,79	4,41	9,09	3,87
Mulheres	1,27	1,60	2,02	2,04	3,15	6,90	2,96
Salvador							
Homens	1,09	1,53	2,18	2,17	3,94	10,12	3,410
Mulheres	0,71	0,87	1,26	1,80	2,61	7,10	2,31
São Paulo							
Homens	2,02	2,94	3,74	3,38	5,71	14,33	5,28
Mulheres	1,69	1,96	2,58	2,48	3,90	10,03	4,03

Fonte: Convênio DIEESE/SEADE, MTB/FAT e convênio regionais, PED – Pesquisa de Emprego e Desemprego

Elaboração: DIEESE/janeiro de 2000

- a) O ingresso de mulheres no ensino superior proporcionou a equiparação dos rendimentos salariais entre os sexos nas regiões metropolitanas.
- b) Nas regiões apresentadas, os homens são mais bem remunerados do que as mulheres, porque possuem nível de instrução mais elevado.
- c) A relação entre as variáveis sexo e escolaridade permite inferir que a diferença de gênero determina rendimentos menores às mulheres.
- d) A diferença entre a remuneração de homens e mulheres é menor na coluna “Ensino superior”, se comparada à das demais colunas.
- e) A diferença absoluta dos rendimentos entre homens e mulheres, na coluna “Ensino fundamental incompleto”, é maior na cidade da região Nordeste.

Gráficos estatísticos

Vimos, anteriormente, que os dados podem ser organizados na forma de uma tabela de distribuição de frequências (com ou sem intervalos de classe).

Todavia, a representação gráfica de uma distribuição de frequências, em geral, transmite uma idéia melhor de um levantamento estatístico do que uma tabela “recheada” de números.

Por ora apresentaremos alguns dos principais tipos de gráficos estatísticos.

Gráfico de linha

É um tipo de gráfico traçado num plano cartesiano, geralmente utilizado para mostrar tendências de crescimento, decrescimento ou estabilidade dos valores numéricos assumidos por uma variável num certo período.

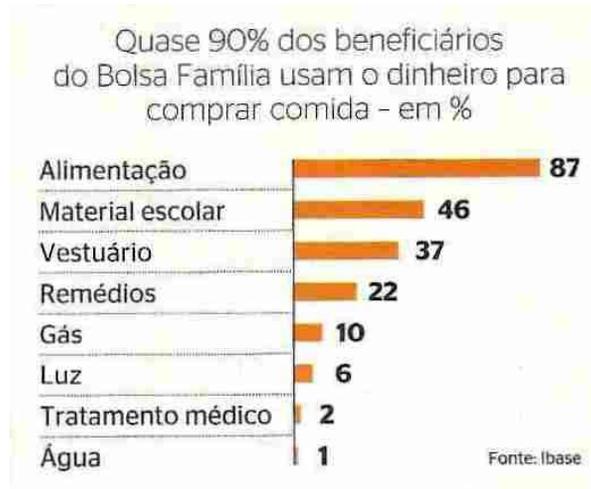
Também pode ser chamado de gráfico poligonal ou de gráfico de segmentos.



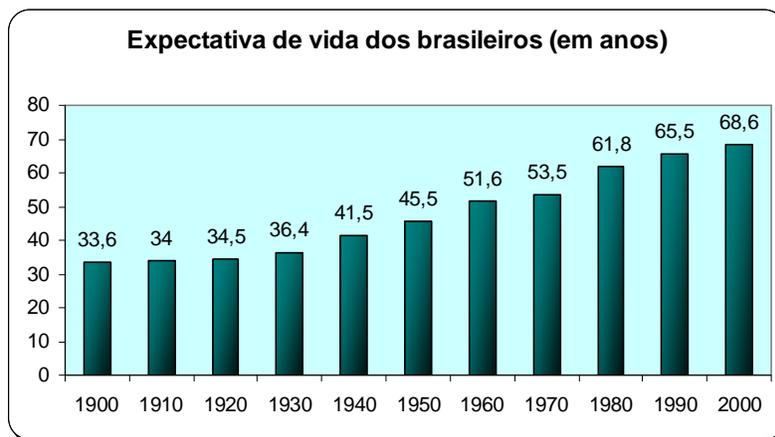
Observando-se o gráfico acima conclui-se que, de 2003 ao 2º trimestre de 2008, o preço, em dólares, da tonelada do minério de ferro só aumentou. Além disso, percebe-se uma maior intensidade do crescimento, de 2007 ao 2º trimestre de 2008, vista que o segmento correspondente a tal período é o que apresenta maior inclinação.

Gráfico de barras ou de colunas

Trata-se de um tipo de gráfico formado por retângulos paralelos, horizontais ou verticais, utilizado, em geral, para comparar as frequências dos valores assumidos por uma variável.



Fonte: Revista ÉPOCA nº 547. 10 nov. 2008. p.75



Fonte: Folha de S.Paulo, 30/9/2003.

Gráfico de setores (gráfico de pizza)

Este gráfico é construído a partir de um círculo e é empregado sempre que se deseja ressaltar a participação de uma categoria em relação ao total.

A construção de tal gráfico exige a divisão do círculo em fatias cuja área é *proporcional* às frequências observadas na tabela a partir da qual ele foi elaborado, sendo que o ângulo central correspondente a cada setor pode facilmente ser obtido a partir de uma regra de três simples, lembrando que 360° correspondem a 100% dos dados.



Quando da observação do gráfico acima, fica claro para o leitor o quão considerável é a participação da televisão no volume total de anúncios do país.

Histograma

Na representação gráfica de uma distribuição de freqüências é utilizado um outro tipo de gráfico, denominado histograma.

O histograma é um gráfico de colunas justapostas, cujas bases estão localizadas sobre o eixo horizontal, de modo que seus pontos médios coincidem com os pontos médios dos intervalos de classe.

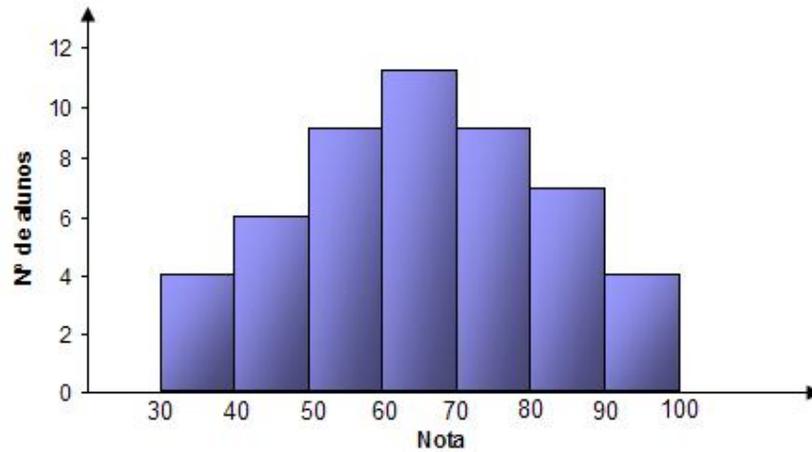
No eixo vertical são marcadas as freqüências absolutas, que correspondem às alturas dos retângulos. Pode-se também marcar, no eixo vertical, as freqüências relativas.

Construindo um histograma

- Seja a seguinte tabela de freqüências na qual se organizou, em 7 classes, a nota obtida por 50 alunos de uma turma.

Classe	Nota	f_i	$f_r(\%)$
1	30 — 40	4	8
2	40 — 50	6	12
3	50 — 60	9	18
4	60 — 70	11	22
5	70 — 80	9	18
6	80 — 90	7	14
7	90 — 100	4	8
		$\Sigma = 50$	100%

Construa um histograma que relacione as classes com as frequências absolutas:



Observemos que cada coluna corresponde a uma classe e que a altura de cada coluna corresponde à frequência absoluta da respectiva classe.

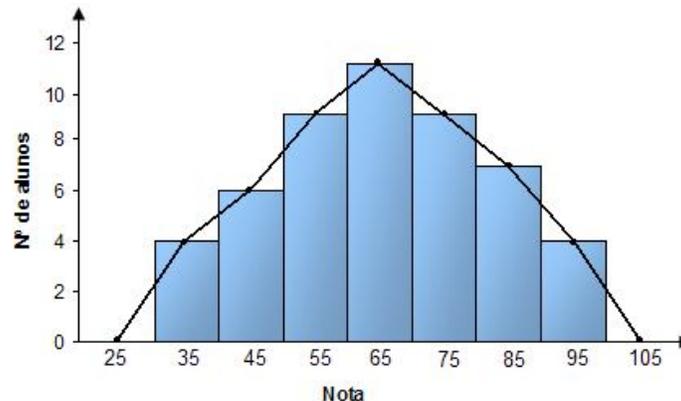
Polígono de frequência

Podemos representar um conjunto de dados agrupados por meio de um polígono de frequências.

Para tanto, no eixo horizontal, cada classe será representada pelo ponto médio do intervalo correspondente. Por exemplo, a classe 30 — 40 é representada pelo valor 35, obtido pela média aritmética dos valores 30 e 40.

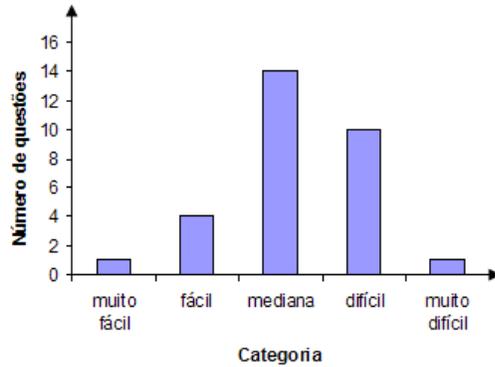
Os segmentos que ligam, em seqüência, os pontos médios das bases superiores das colunas de um histograma formam o polígono de frequências, sendo que, nas extremidades do polígono, utilizamos pontos do eixo horizontal cujas abscissas estejam a uma mesma distância das demais abscissas.

O polígono de frequência que se segue foi formado a partir do histograma anterior.



Questões propostas

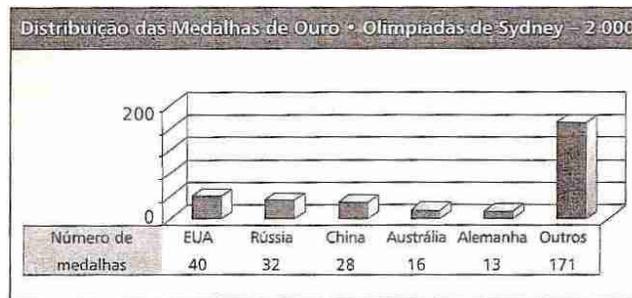
Q4. (UFRGS) As questões de Matemática do Concurso Vestibular da UFRGS de 2004 foram classificadas em categorias quanto ao índice de facilidade, como mostra o gráfico de barras a seguir:



Se essa classificação fosse apresentada em um gráfico de setores, a cada categoria corresponderia um setor circular. O ângulo do maior desses setores mediria:

- a) 80° b) 120° c) 157° d) 168° e) 172°

Q5. (ENEM) As Olimpíadas são uma oportunidade para o conagraçamento de um grande número de países, sem discriminação política ou racial, ainda que seus resultados possam refletir características culturais, socioeconômicas e étnicas. Em 2000, nos Jogos Olímpicos de Sydney, o total de 300 medalhas de ouro conquistadas apresentou a seguinte distribuição entre os 196 países participantes, como mostra o gráfico.

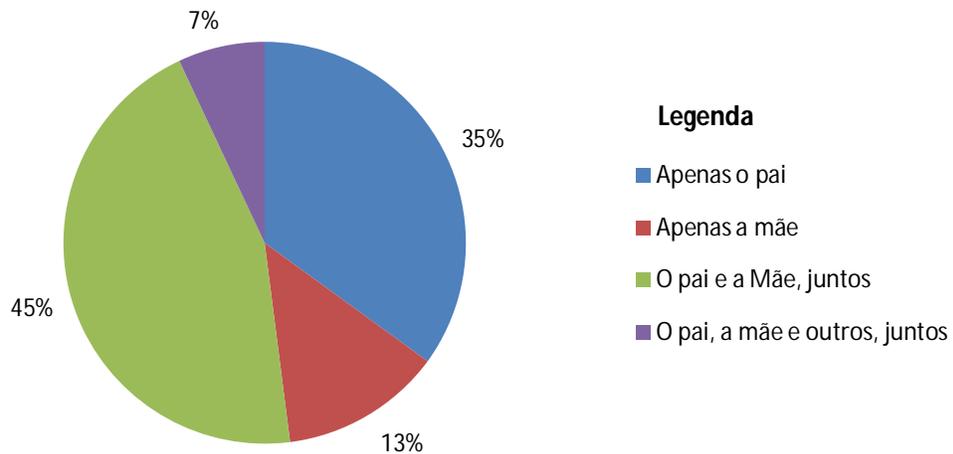


Esses resultados mostram que, na distribuição das medalhas de ouro em 2000,

- a) cada país participante conquistou pelo menos uma.
b) Cerca de um terço foi conquistado por apenas três países
c) Os cinco países mais populosos obtiveram os melhores resultados.
d) Os cinco países mais desenvolvidos obtiveram os melhores resultados.
e) Cerca de um quarto foi conquistado pelos Estados Unidos.

Q6. (UFMG) Este gráfico representa o resultado de uma pesquisa realizada com 1000 famílias com filhos em idade escolar:

Responsáveis pela renda familiar



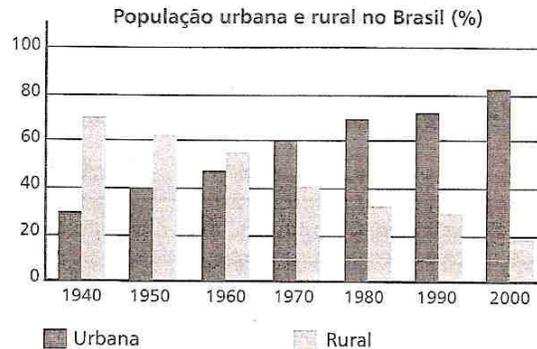
Considere estas afirmativas referentes às famílias pesquisadas:

- I) O pai participa da renda familiar em menos de 850 dessas famílias.
- II) O pai e a mãe participam, juntos, da renda familiar em mais de 500 dessas famílias.

Então, é correto afirmar que:

- a) Nenhuma das afirmativas é verdadeira.
- b) Apenas a afirmativa I é verdadeira.
- c) Apenas a afirmativa II é verdadeira.
- d) Ambas as afirmativas são verdadeira.

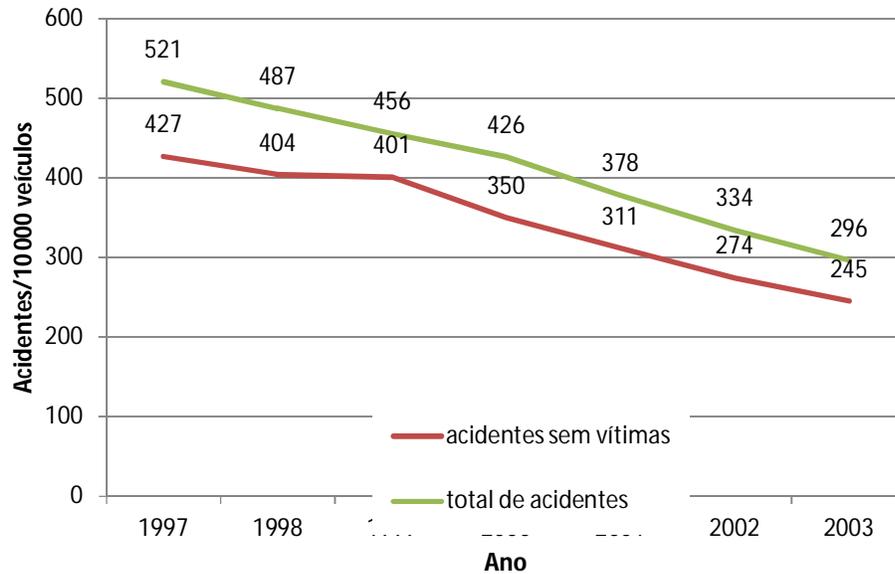
Q7. (ENEM) Ao longo do século XX, as características da população brasileira mudaram muito. Os gráficos mostram as alterações da distribuição da população da cidade e do campo e na taxa de fecundidade (número de filhos por mulher) no período entre 1940 e 2000.



Comparando-se os dados dos gráficos, pode-se concluir que

- o aumento relativo da população rural é acompanhado pela redução da taxa de fecundidade.
- Quando predominava a população rural, as mulheres tinham em média três vezes menos filhos do que hoje.
- A diminuição relativa da população rural coincide com o aumento do número de filhos por mulher.
- Quanto mais aumenta o número de pessoas morando em cidades, maior passa a ser a taxa de fecundidade.
- Com a intensificação do processo de urbanização, o número de filhos por mulher tende a ser menor.

Q8. (Unicamp-SP) O gráfico abaixo mostra o total de acidentes de trânsito na cidade de Campinas e o total de acidentes sem vítimas, por 10 000 veículos, no período entre 1997 e 2003. Sabe-se que a frota da cidade de Campinas era composta por 500 000 veículos em 2003 e era 4% menor em 2002.



Adaptado de: Sumário Estatístico da Circulação em Campinas 2001-2003. Campinas, EMDEC, 2004, p.12

- Calcule o número total de acidentes de trânsito ocorridos em Campinas em 2003.
- Calcule o número de acidentes com vítimas ocorridos em Campinas em 2003.

Medidas de tendência central

A partir do estudo de uma amostra podem ser obtidas informações essenciais à pesquisa estatística, a saber:

- O valor numérico mais provável de se obter;
- A capacidade de variação dos dados ao se examinarem várias observações simultaneamente;
- A frequência com que cada um dos valores observados ocorre dentro do universo de respostas possíveis.

Esses três itens correspondem, respectivamente, aos valores da média aritmética, do desvio padrão e ao tipo de distribuição da resposta que está sendo caracterizada pelo conjunto de dados e somente através deles a análise estatística poderá ser devidamente caracterizada.

Por ora apresentaremos algumas medidas de tendência central importantes, que são assim denominadas por serem constituídas por valores que, geralmente, estão localizados no meio de um conjunto de dados ordenados.

Tais valores são representativos do conjunto de dados ao qual pertencem.

Média aritmética

Denomina-se média aritmética de um conjunto de valores de uma variável o quociente entre a soma dos valores da variável pelo número deles, isto é,

$$\bar{x} = \frac{\sum x_i}{n}$$

em que:

- \bar{x} é a média aritmética
- x_i representa os valores que a variável pode assumir
- n é o número de valores

Calculando a média aritmética de um conjunto de dados não agrupados

- Um funcionário registrou, em uma tabela, o tempo gasto, em minutos, para se deslocar de sua residência ao seu local de trabalho, ao longo de uma semana.

Dia da semana	Tempo gasto (min)
Segunda-feira	32
Terça-feira	28
Quarta-feira	29
Quinta-feira	31
Sexta-feira	25
Sábado	20

Qual foi o tempo médio gasto, com o deslocamento?

$$\bar{x} = \frac{\sum x}{n} \Rightarrow \bar{x} = \frac{32 + 28 + \dots + 20}{6} = 27,5 \text{ minutos.}$$

É possível calcular-se a média aritmética a partir de uma tabela de freqüências, desde que se utilize o valor correspondente a cada classe (ou o ponto médio da classe, no caso de a tabela conter intervalos de classe) e as respectivas freqüências absolutas.

Nesse caso, a média é obtida a partir da fórmula

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

em que:

- x_i é o valor da classe i ou o ponto médio do intervalo da i -ésima classe;
- f_i representa a frequência absoluta da i -ésima classe;
- $\sum f_i$ é a soma das frequências absolutas de todas as classes.

Calculando a média aritmética de uma tabela de frequências

- Uma imobiliária gerencia o aluguel de residências particulares, segundo a tabela a seguir. Calcule o valor médio de um aluguel, para a distribuição de frequências nela contida.

Aluguel (R\$)	Número de casas
0 - 200,00	30
200,00 - 400,00	52
400,00 - 600,00	28
600,00 - 800,00	7
800,00 - 1000,00	3

Como os dados estão agrupados em intervalos de classe, será necessário determinar os pontos médios das classes e criar, também, uma nova coluna ($x_i f_i$) da qual nos interessará o somatório.

Aluguel (R\$)	Número de casas – f_i	x_i	$x_i f_i$
0 - 200	30	100	3000
200 - 400	52	300	15600
400 - 600	28	500	14000
600 - 800	7	700	4900
800 - 1000	3	900	2700
	$\sum f_i = 120$		$\sum x_i f_i = 40200$

Dessa forma, calculamos a média aritmética a partir da fórmula $\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$, a

saber:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} \Rightarrow \bar{x} = \frac{40200}{120} = 335$$

Logo, o valor médio do aluguel é R\$ 335,00.

Mediana

A mediana (M_e) de um conjunto de dados é definida como o número que se encontra no centro, quando esses dados estão dispostos em ordem crescente ou decrescente.

Caso o conjunto de dados tenha um número par de valores, a mediana será a média aritmética dos dois valores centrais.

Moda

Denomina-se moda (M_o) o valor que ocorre com maior frequência em um conjunto de dados.

Ao contrário da média e da mediana, que sempre existem, alguns conjuntos de dados não apresentam moda (são amodais).

Por outro lado, há conjuntos de dados que podem apresentar dois ou mais valores com a frequência máxima. O conjunto de dados que apresenta dois valores modais é dito bimodal. Se houver três ou mais valores modais, o conjunto chamado de multimodal.

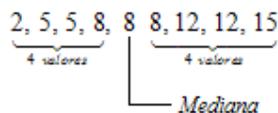
Calculando medianas e modas

- Calcule a mediana e a moda de cada um dos seguintes conjuntos de valores:

- 2, 5, 8, 8, 12, 15, 8, 5, 12
- 3,4; 5,2; 4,7; 6; 8,4; 9,3; 2,1; 4,8
- 1,2, 2, 2, 3, 4, 5, 5, 5

Em ordem crescente, temos:

a)



Como a quantidade de valores do conjunto é um número ímpar, o termo médio é o 5º e, portanto, $M_e = 8$.

O valor mais freqüente no conjunto de dados é 8; logo, $M_o = 8$.

- b) 2,1; 3,4; 4,7; $\overbrace{4,8; 5,2}^{\text{As duas posições centrais}}$; 6; 8,4; 9,3

O segundo conjunto de valores apresenta 8 elementos (quantidade par). Portanto, a mediana será obtida a partir da média aritmética dos dois valores centrais. Dessa forma, a mediana será dada por:

$$M_e = \frac{4,8 + 5,2}{2} = 5$$

Não há, nesse conjunto, um valor mais freqüente que os demais. Portanto, trata-se de um conjunto de valores amodal.

c) 1, 2, 2, 2, 3, 4, 5, 5, 5
 4 valores 4 valores
 └──────────┬──────────
 Mediana

$M_e = 3$, posto que 3 é o valor central do conjunto de dados ordenado.

Como há dois valores que ocorrem com frequência máxima, o conjunto de valores é bimodal, sendo que $M_o = 2$ e $M_o = 5$.

Questões propostas

Q9. (UFC-CE) A média aritmética das notas dos alunos de uma turma formada por 25 meninas e 5 meninos é igual a 7. Se a média aritmética das notas dos meninos é igual a 6, a média aritmética das notas das meninas é igual a:

- a) 6,05 b) 7,2 c) 7,4 d) 7,8 e) 8,0

Q10. Antes do lançamento de um novo tipo de sabonete, uma indústria realizou uma pesquisa de mercado na qual apurou os valores cobrados por dez marcas concorrentes e obteve os seguintes valores, em reais: 1,12; 1,00; 1,07; 1,18; 1,60; 1,90; 0,92; 2,02; 1,70; 1,12. Calcule a média aritmética, a moda e a mediana desses valores.

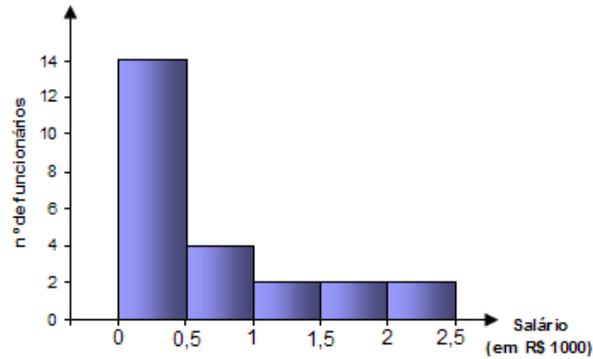
Q11. (Uece) Considere três números inteiros e positivos **a**, **b** e **c**, onde um deles é igual à média aritmética dos outros dois. A soma $2a+2b+2c$ é igual ao:

- a) Dobro de um dos números dados.
 b) Triplo de um dos números dados.
 c) quádruplo de um dos números dados.
 d) Sêxtuplo de um dos números dados.

Q12. (Uece) A média aritmética dos elementos de um conjunto de números inteiros é 14,625. Se **n** é o número de elementos desse conjunto, então o menor valor de **n** é:

- a) 4. b) 6. c) 8. d) 16.

Q13. (PUC-SP) O histograma a seguir apresenta a distribuição de frequências das faixas salariais numa pequena empresa.



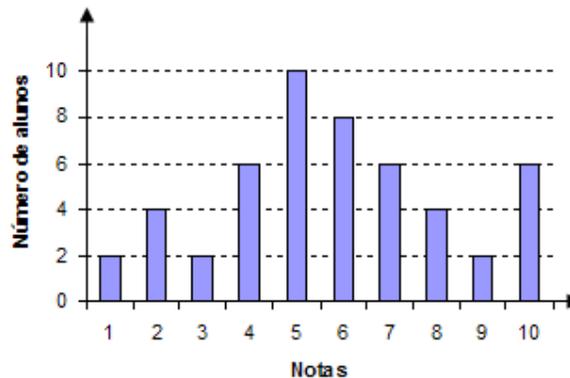
Com os dados disponíveis, pode-se concluir que a média desses salários é, aproximadamente,

- a) R\$ 420,00
- b) R\$ 536,00
- c) R\$ 562,00
- d) R\$ 640,00
- e) R\$ 708,00

Q14. (FUVEST-SP) Numa classe com 20 alunos as notas do exame final podiam variar de 0 a 100 e a nota mínima para aprovação era 70. Realizado o exame, verificou-se que 8 alunos foram reprovados. A média aritmética das notas desses 8 alunos foi 65, enquanto a média dos aprovados foi 77. Após a divulgação dos resultados, o professor verificou que uma questão havia sido mal formulada e decidiu atribuir 5 pontos a mais a todos os alunos. Com essa decisão, a média dos aprovados passou a ser 80 e a dos reprovados 68,8.

- a) Calcule a média aritmética das notas da classe toda antes da atribuição dos 5 pontos extras.
- b) Com a atribuição dos 5 pontos extras, quantos alunos, inicialmente reprovados, atingiram nota para aprovação?

Q15. (Ibmec-SP) Chama-se mediana de um conjunto de 50 dados ordenados em ordem crescente o número x dado pela média aritmética entre o 25º e o 26º dado. Observe no gráfico a seguir uma representação para as notas de 50 alunos do primeiro semestre de Ciências Econômicas numa determinada prova.



A mediana das notas do 50 alunos de Ciências Econômicas nesta prova é igual a:

- a) 3. b) 4. c) 5. d) 6. e) 7.

Medidas de variação (ou medidas de dispersão)

Sejam os seguintes conjuntos de valores das variáveis **A**, **B** e **C**:

A: 60, 60, 60, 60, 60

B: 58, 59, 60, 61, 62

C: -5, 5, 40, 110, 150

A média aritmética de cada um desses conjuntos é igual a 60 (sugerimos que você, leitor, verifique isso como exercício).

Todavia, não é difícil notar que o conjunto **A** é o mais homogêneo, pois todos os seus valores são iguais entre si, e que o conjunto **B** é mais homogêneo que o conjunto **C**, uma vez que há uma menor diversificação entre cada um dos valores de **B** e a média do conjunto.

Ora apresentaremos algumas medidas de variação, que nos permitirão medir a menor ou maior diversificação dos valores de uma variável em torno de um valor de tendência central, tomado como ponto de comparação.

Vale ressaltar que números relativamente próximos uns dos outros têm baixas medidas de variação, enquanto que números mais dispersos têm maior medida de variação.

Amplitude total

Define-se como **amplitude total** (AT) a diferença entre o maior e o menor valor observado num conjunto de dados.

A amplitude total é considerada uma medida instável por não levar em consideração todos os valores e por se deixar influenciar pelos valores extremos.

Desvio em relação à média

Denomina-se **desvio em relação à média** a diferença entre cada valor x_i de um conjunto de dados e a média aritmética, isto é, $d_i = x_i - \bar{x}$.

Desvio médio

O **desvio médio** (DM) é definido como a média aritmética dos módulos dos desvios em relação à média, isto é,

$$DM = \frac{|d_1| + |d_2| + \dots + |d_n|}{n},$$

em que d_1, d_2, \dots, d_n são os desvios em relação à média e n é a quantidade de valores do conjunto de dados.

Determinando amplitudes totais, desvios em relação à média e o desvio médio

- A quantidade de pessoas que visitou uma exposição, de 3ª feira a domingo, foi de 21, 18, 19, 20, 20 e 28. Determine a amplitude total, os desvios em relação à média e o desvio médio para esse conjunto de valores.

Observando os valores, verifica-se que o maior deles é 28 e o menor é 18. Logo, $AT = 28 - 18 = 10$.

A média aritmética é $\bar{x} = \frac{21+18+19+20+20+28}{6} = 21$.

Logo, os desvios em relação à média são:

$$\begin{aligned}d_1 = x_1 - \bar{x} &\Rightarrow d_1 = 21 - 21 = 0 & d_4 = x_4 - \bar{x} &\Rightarrow d_4 = 20 - 21 = -1 \\d_2 = x_2 - \bar{x} &\Rightarrow d_2 = 18 - 21 = -3 & d_5 = x_5 - \bar{x} &\Rightarrow d_5 = 20 - 21 = -1 \\d_3 = x_3 - \bar{x} &\Rightarrow d_3 = 19 - 21 = -2 & d_6 = x_6 - \bar{x} &\Rightarrow d_6 = 28 - 21 = 7\end{aligned}$$

$$DM = \frac{|d_1| + |d_2| + \dots + |d_n|}{n} \Rightarrow$$

$$DM = \frac{|0| + |-3| + |-2| + |-1| + |-1| + |7|}{6} = \frac{7}{3}$$

Obs: A soma algébrica dos desvios em relação à média é sempre nula.

Variância

A **variância**, indicada por s^2 , é definida como o valor correspondente à média aritmética dos quadrados dos desvios em relação à média, isto é,

$$s^2 = \frac{|d_1|^2 + |d_2|^2 + \dots + |d_n|^2}{n} \text{ ou seja,}$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

Uma vez que a variância é calculada a partir dos quadrados dos desvios em relação à média, ela é um número em unidade quadrada em relação à variável observada. A fim de evitar os inconvenientes relacionados às unidades de medida,

criou-se uma nova medida de variação, com utilidade e interpretação práticas, chamada desvio padrão, que definiremos na seqüência.

Desvio padrão

O desvio padrão é uma medida da variação representada por “s”, definida como a raiz quadrada da variância.

Assim:

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

Obs.: Tanto a variância quanto o desvio padrão apresentam como principal vantagem sobre a amplitude total o fato de levarem em consideração todos os valores.

Utilizando o desvio padrão na análise da dispersão das notas de duas turmas

- Após a aplicação e correção de uma mesma avaliação em duas turmas de inglês, cada uma com 6 alunos, o professor responsável pôde confeccionar a seguinte tabela:

	Notas dos alunos					
Turma α	12	15	12	16	14	15
Turma β	12	11	18	9	19	15

Tomando por base somente essa avaliação, qual turma se mostrou mais homogênea?

A turma mais homogênea é aquela que apresentou menor dispersão das notas em relação à nota média.

Calcularemos, inicialmente, a nota média para cada turma.

$$\bar{x}_\alpha = \frac{12+15+12+16+14+15}{6} = 14$$

$$\bar{x}_\beta = \frac{12+11+18+9+19+15}{6} = 14$$

Em seguida, calcularemos o desvio padrão para os dois conjuntos de notas.

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} \Rightarrow$$

$$s_\alpha = \sqrt{\frac{(12-14)^2 + (15-14)^2 + (12-14)^2 + (16-14)^2 + (14-14)^2 + (15-14)^2}{6}}$$

$$s_\alpha = \sqrt{2,33} \Rightarrow s_\alpha \approx 1,53$$

$$s_\beta = \sqrt{\frac{(12-14)^2 + (11-14)^2 + (18-14)^2 + (9-14)^2 + (19-14)^2 + (15-14)^2}{6}}$$

$$s_\beta = \sqrt{13,33} \Rightarrow s_\beta \approx 3,65$$

Como o desvio padrão correspondente às notas da turma α é menor que o desvio padrão correspondente às notas da turma β , podemos afirmar que, entre as duas, a turma α é a mais homogênea.

Questões propostas

Q16. Dado o conjunto de valores 2, 3, 5, 4 e 6, obtenha o desvio médio correspondente.

Q17. O quadro a seguir apresenta a idade (em anos) de cada um dos onze jogadores titulares de um time de futebol.

22	25	20	20	22	
					26
23	22	26	31	23	

Utilizando os conceitos de média aritmética (\bar{x}) e desvio padrão (d), é possível determinar a dispersão das idades dos atletas em torno da média de suas idades. Assim, é correto afirmar que o número de jogadores com idade pertencente ao intervalo $[\bar{x} - d; \bar{x} + d]$ é

- a) 5 b) 6 c) 7 d) 8 e) 9

Q18. (Unicamp-SP) Para um conjunto $X = \{x_1, x_2, x_3, x_4\}$, a média aritmética de X é definida por $\bar{x} = \frac{x_1 + x_2 + x_3 + x_4}{4}$ e a variância de X é definida por

$$v = \frac{1}{4}[(x_1 - \bar{x})^2 + \dots + (x_4 - \bar{x})^2].$$

Dado o conjunto $X = \{2, 5, 8, 9\}$, pede-se:

- a) Calcular a média aritmética de X .
b) Calcular a variância de X .

c) Quais elementos de X pertencem ao intervalo $[\bar{x} - \sqrt{v}; \bar{x} + \sqrt{v}]$?

Questões complementares

Q19. (UERJ) “O engenheiro Ronaldo Belassiano descobriu que o carioca é o povo mais ágil para embarcar nos coletivos. Ele leva, em média, apenas 1,85segundo contra 2,4 segundos gastos, em média, pelos londrinos.” (Revista Superinteressante, set.96, com adaptações.)

Com base no texto, considere que um ônibus no Rio de Janeiro fique parado num ponto, durante 74 segundos, e embarque passageiros de acordo com a média apresentada. Em Londres, para embarcar essa mesma quantidade de passageiros, o ônibus deverá ficar parado durante:

- a) 96s b) 104s c) 108s d) 220s

Q20. (ESPM-SP) Considere todos os pares ordenados (x,y) do produto cartesiano $A \times B$ em que $A = \{1, 2, 3, 4\}$ e $B = \{1, 3, 5\}$. Tomando-se todos os 12 produtos $x.y$, podemos afirmar que a média, a moda e a mediana desse conjunto são, respectivamente:

- a) 9,5; 7,5 e 5,5
b) 7,5; 5,5 e 3,0
c) 7,5; 3,0 e 5,5
d) 5,5; 5,5 e 5,5
e) 7,5; 3,0 e 6,0

Q21. (FGV-SP, adaptada) Em um conjunto de 100 observações numéricas distintas, podemos afirmar que:

- a) a média aritmética é maior que a mediana.
b) A mediana é maior que a moda
c) 50% dos valores estão acima da média aritmética
d) 50% dos valores estão abaixo da mediana
e) 25% dos valores estão entre a moda e a mediana.

Q22. (UFPR) Dado um conjunto $\mathbf{X} = \{x_1, x_2, x_3, \dots, x_n\}$ com n elementos, definimos a média \bar{x} e o desvio padrão d de \mathbf{X} por:

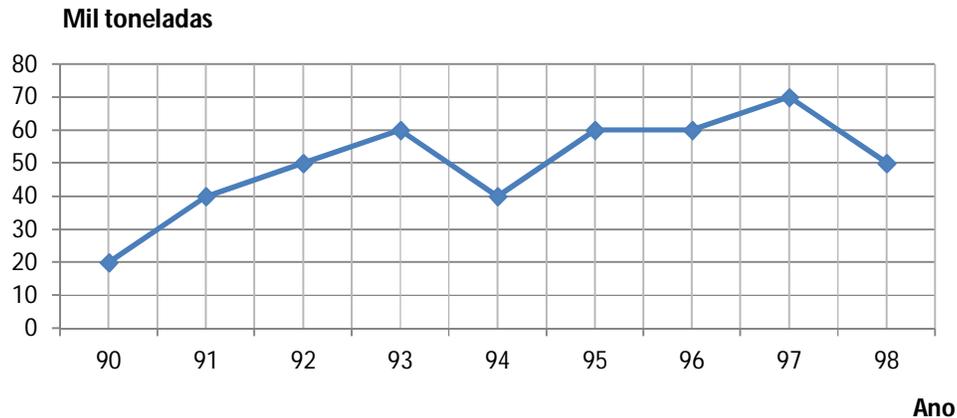
$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \text{e} \quad d = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

Uma informação útil para quem analisa um conjunto de dados como \mathbf{X} é que a maioria desses dados pertence ao intervalo $\mathbf{C} = [\bar{x} - 2d, \bar{x} + 2d]$. Sendo $\mathbf{X} = \left\{\frac{5}{2}, 4, \frac{7}{2}, 3\right\}$ um conjunto de dados:

- 1) Calcule a média \bar{x} e o desvio padrão d .
- 2) Verifique quais elementos do conjunto \mathbf{X} acima pertencem ao intervalo \mathbf{C} .

Q23. (UFABC/2009) A média aritmética das idades de um grupo de x pessoas é 25 anos. Com a entrada de mais uma pessoa no grupo, a nova média passou a ser 26 anos. Determine a idade do novo integrante do grupo em função de x .

Q24. (Vunesp) O gráfico representa, em milhares de toneladas, a produção no estado de São Paulo de um determinado produto agrícola entre os anos de 1990 e 1998.



Analisando o gráfico, observa-se que a produção:

- Foi crescente entre 1992 e 1995.
- Teve média de 40 mil toneladas ao ano.
- Em 1993 teve acréscimo de 30% em relação ao ano anterior.
- A partir de 1995 foi decrescente.
- Teve média de 50 mil toneladas ao ano.

Gabarito – ESTATÍSTICA

- Q1. a) quantitativa discreta
a) quantitativa contínua
b) qualitativa
c) quantitativa discreta
d) quantitativa discreta
e) quantitativa contínua
f) qualitativa

Q2. c

Q3. c

Q4. d

Q5. b

Q6. c

Q7. e

Q8. a) 14800 b) 2880

Q9. b

Q10. $\bar{x} \approx 1,36$, $M_e = 1,15$ e $M_o = 1,12$

Q11. d

Q12. c

Q13. e

Q14. a) 72,2 b) 3

Q15. d

Q16. DM = 1,2

Q17. d

Q18. a) $\bar{x} = 6$ b) $v = \frac{15}{2}$ c) 5 e 8

Q19. a

Q20. c

Q21. d

Q22. 1) $\bar{x} = 3,25$ e $d = \frac{\sqrt{5}}{4}$

2) Todos

Q23. $x + 26$

Q24. e