Estatística Básica*

Fabricio Ferrari
fabricio@ferrari.pro.br

2004

Parte I

1 Método científico

Método: conjunto de meios e rotinas dispostos convenientemente para se chegar a um fim que se deseja.

Método experimental: método que consiste em manter constantes todas as causas (fatores), menos uma, e variar esta causa de modo que se possa descobrir seus efeitos.

Método estatístico: método que admite todas as causas presentes variando-as, dada a impossibilidade de manter as causas constantes, registrando estas variações e procurando determinar que influências cabem a cada uma delas.

Estatística: a estatística é a parte da Matemática Aplicada que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados e para a utilização dos mesmos na tomada de decisões.

2 Fases do método estatístico

As várias fases do método estatístico estão delineadas a seguir. As etapas descritas nas seções de planejamento, coleta, crítica, apuração e exposição constituem a **Estatística Descritiva**, enquanto a seção de análise constitui a **Estatística Indutiva ou Inferencial**

*Baseado em *Estatística Fácil* de Antônio Arnot Crespo, Editora Saraiva, 1999.

2.1 Planejamento

A primeira etapa consiste em planejar o modo como serão realizadas as fazes seguintes, determinando o objetivo da pesquisa e os métodos que serão utilizados. Nesta etapa são definidos os objetivos, as características da amostra, o método de aquisição e de processamento de dados.

2.2 Coleta de dados

Coleta indireta

A **coleta direta** de dados é quando os dados são obtidos pelo próprio pesquisador através de levantamento de registros (nascimentos, óbitos, notas fiscal, impostos, etc.) ou coletados diretamente através de inquéritos, questionários, etc.

A **coleta direta** pode ser classificado quanto ao fator **tempo** como:

- **contínua:** quando feita de forma continuada, como registro de nascimentos e óbitos, frequência de alunos às aulas, etc.

- **periódica:** quanto feita em intervalos constantes de tempo, como censos (10 em 10 anos), avaliações mensais dos alunos, etc.

- **ocasional:** quanto feitas em determinada situação para atender a um objetivo, como pesquisa de mortalidade de um rebanho, pesquisa de um produto no mercado, etc.

Coleta indireta

A **coleta indireta** é inferida de elementos conhecidos, através de uma coleta direta, ou do conhecimento de fenômenos relacionados ao fenômeno estudado. Por exemplo, pesquisa sobre mortalidade infantil que é feita sobre a coleta direta de dados de nascimentos e óbitos.

2.3 Crítica dos dados

Os dados obtidos devem ser criticados à procura de falhas sistemáticas no planejamento, aquisição e armazenamento dos dados.

2.4 Apuração dos dados

É a etapa de soma e processamento dos dados obtidos mediante critérios de classificação. Pode ser **manual**, **eletromecânica** ou **eletrônica**.

2.5 Exposição ou apresentação dos dados

Os dados sempre devem ser apresentados de forma adequada, seja através de tabelas ou gráficos, seguindo os critérios determinados no planejamento e utilizados no processamento dos dados. A exposição dos dados tem o objetivo de facilitar a análise daquilo que é objeto do estudo estatístico.

2.6 Análise

A última etapa do processo estatístico consiste em tirar conclusões sobre os dados levantados e processados, inferindo conclusões sobre o todo (**população**) a partir de dados coletados de uma parte representativa da população (**amostra**)

3 População e amostra

3.1 Variáveis

Variável é o conjunto de resultados possíveis de um fenômeno.

As variáveis podem ser:

- **qualitativa:** quando seus valores são expressos por atributos, de forma não numérica. Por exemplo sexo (M ou F), cor da pele (branca, preta, amarela, ...), etc.

- **quantitativa:** quando seus valores são expressos por números. Por exemplo idade, salário, volume, etc. As variáveis quantitativas ainda são classificadas como:

  – *discreta:* quando os seus valores podem enumerados. Ex. de contagem: do número de pessoas numa sala (1, 2, 3, . . .)

  – *contínua:* quando os seus valores podem ser qualquer um num intervalo. Ex. de medições: volume de uma caixa d'agua (1 m^3, 1.1 m^3, 1.01 m^3, . . .)

3.2 População e amostra

População e amostra referem-se ao conjunto de entes cujas propriedades desejamos averiguar.

População estatística ou **universo estatístico** é o conjunto de entes portadores de pelo menos uma característica em comum. Por exemplo, os estudantes constituem uma população com uma característica em comum: são os que estudam.

Muitas vezes, por motivos práticos ou econômicos, limitam-se os estudos estatísticos somente a uma parte da população, a amostra. A **amostra** é um subconjunto finito de uma população.

Como todo a análise estatística será inferida a partir das características obtidas da amostra, é importante que a amostra seja **representativa** da população, isto é, que as suas características de uma parte (amostra) sejam em geral as mesmas que do todo (população).

3.3 Amostragem

Amostragem é a técnica especial de escolher amostras que garanta o acaso na escolha. Assim cada elemento da população tem a mesma chance de ser escolhido, o que garante à amostra um caráter de representatividade da população.

3.3.1 Amostragem casual ou aleatória simples

Este tipo de amostragem é baseado no sorteio da amostra. Numera-se a população de 1 a n e depois, utilizando um dispositivo aleatório qualquer, escolhem-se k números desta sequência, que corresponderão aos elementos da amostra.
EXEMPLO:

pesquisa da estatura de uma escola com 90 alunos (população: 90 alunos) usando uma amostra de 10% da população:

1. numeram-se os alunos de 1 a 90;
2. sorteiam-se 9 números (10% de 90) usando algum mecanismo aleatório ou através de uma Tabela de Números Aleatórios (veja o Apêndice D). Por exemplo escolhendo-se a 5^a linha da tabela do Apêndice D), tem-se: `14 35 30 19 66 27 77 45 38`
3. os alunos numerados de acordo com a lista acima são escolhidos e tomados os valores das suas estaturas, obtendo assim uma amostra da população dos 90 alunos.

3.3.2 Amostragem proporcional estratificada

Quando a população se divide em sub-populações – **estratos** – é necessário utilizar um amostragem proporcional estratificada, que considera os estratos (subgrupos) e obtém a amostragem proporcional a estes.
EXEMPLO: Suponha que no exemplo anterior, dos noventa alunos, 54 sejam meninos e 36 sejam meninas. Neste caso precisamos obter a amostra estratificada. Serão dois estratos (sexo masculino e sexo feminino) e queremos uma amostra de 10% da população. Assim,

1. Definimos a amostra em estratos:

SEXO	POPULAÇÃO	10%	AMOSTRA
M	54	5,4	5
F	36	3,6	4
Total	90	9,0	9

2. Numeram-se os alunos de 1 a 90 sendo que 1 a 54 correspondem a meninos e de 55 a 90, a meninas. Tomando a 2 coluna, de cima para baixo, tem-se: `56 05 46 74 90 17 75 63 31`.

3. Neste caso serão obtidas as características dos seguintes alunos:
   `56 05 46 74 90` – meninos
   `17 75 63 31` – meninas.

3.3.3 Amostragem sistemática

Quando os elementos da população já estão ordenados, não é necessário construir um sistema de referência ou de amostragem. Neste caso a amostragem é **sistemática**.
EXEMPLO:

Suponha uma rua que tenha 500 prédios e desejamos obter uma amostra de 40 prédios (8%). Como os prédios já estão ordenados na rua, podemos usar o seguinte procedimento:

1. como $500/40 = 12.5$, então temos de selecionar um prédio para a amostra a cada 12.

2. sorteamos um número entre 1 e 12 inclusive, digamos que seja 5.

3. vamos amostrando os prédios iniciando pelo 5^o e pulando de 12 em 12. Assim, iniciamos pelo prédio 5, depois usamos o prédio 12+5, depois 12+12+5, e assim por diante.

4. No final teremos amostrado os 40 prédios.

4 Séries Estatísticas

4.1 Tabelas

A tabela é um quadro que resume um conjunto de observações. Compõe-se de:

- **corpo:** linhas e colunas que contém os valores das variáveis em estudo.
- **cabeçalho:** parte superior que especifica o conteúdo das colunas.
- **coluna indicadora:** coluna que indica o conteúdo das linhas.
- **casa ou célula:** espaço destinado a uma só informação.
- **título:** conjunto de informações sobre a tabela (*O quê? Quando? Onde?*) localizada no topo da tabela.

EXEMPLO:

PRODUÇÃO DE CAFÉ
BRASIL

Anos	Produção (1000 ton)
1991	1221
1992	2234
1993	1254
1994	1445
1995	1112

FONTE: IBGE.

Normas para células

- usar um traço horizontal (—) quando o valor é nulo quanto à natureza das coisas ou resultado do inquérito.
- três pontos (. . .) quando não temos dados.
- um ponto de interrogação (?) quando temos dúvida quanto à exatidão do valor.
- zero (0; 0,0; 0,00) quando o valor é muito pequeno para ser expresso pela grandeza utilizada.

4.2 Séries históricas, cronológicas, temporais ou marchas

Descrevem os valores da variável, em determinado local, discriminados segundo intervalos de tempo variáveis.
EXEMPLO:

PREÇO DO ACÉM
SÃO PAULO

Anos	Preço médio (US$)
1989	2,24
1990	2,73
1991	2,12
1992	1,89
1993	2,04

FONTE: APA.

4.3 Séries geográficas, espaciais, territoriais ou de localização

Descrevem os valores da variável, em determinado instante, discriminados segundo regiões.
EXEMPLO:

DURAÇÃO MÉDIA DOS
ESTUDOS SUPERIORES
1994

Países	**Núm. de anos**
Itália	7,5
Alemanha	7,0
França	7,0
Holanda	5,9
Inglaterra	< 4

FONTE: Revista *Veja*.

4.4 Séries específicas ou categóricas

Descrevem os valores da variável, em determinado tempo e local, discriminando segundo especificações ou categorias.
EXEMPLO:

REBANHOS BRASILEIROS
1992

Espécies	**Quantidade** (1.000 cabeças)
Bovinos	154.440,8
Ovinos	19.955,9
Caprinos	12.159,6
Suínos	34.532,2

FONTE: IBGE.

4.5 Séries conjugadas – tabela de dupla entrada

Constituem-se da conjugação de uma ou mais séries.
EXEMPLO:

Podemos ter a conjugação de uma série **geográfica** com uma série **histórica**.

TERMINAIS TELEFÔNICOS EM SERVIÇO

Regiões	**1991**	**1992**
Norte	342.938	375.658
Sudeste	6.234.501	6.729.467
Sul	1.497.315	1.608.989

FONTE: Ministério das Comunicações.

4.6 Distribuição de frequência

São dados agrupados de acordo com intervalos de valores das variáveis.
EXEMPLO:

ESTATURA DE 100 ALUNOS
DA ESCOLHA X – 1994

Estaturas (cm)	**Núm. de alunos**
$140 \vdash 145$	2
$145 \vdash 150$	5
$150 \vdash 155$	11
$155 \vdash 160$	39
$160 \vdash 165$	32
$165 \vdash 170$	10
$170 \vdash 175$	1
Total	100

FONTE: dados fictícios.

4.7 Dados absolutos e dados relativos

Dados absolutos são aqueles resultantes da coleta direta da fonte, sem outra manipulação senão contagem ou medida. Os **dados relativos** são resultados de especificações por quociente (razões) para facilitar a compreensão entre as quantidades.

4.7.1 Porcentagem

Os dados relativos são especificados como uma razão relativa ao total, que equivale a uma centena (100) ou uma unidade (1).
EXEMPLO:
Total do rebanho: 1456 (100%)
Bovinos: $860/1456 = 0.59 \Rightarrow 59\%$
Ovinos: $354/1456 = 0.243 \Rightarrow 24\%$
Caprinos: $30/1456 = 0.02 \Rightarrow 2\%$
Suínos: $212/1456 = 0.1456 \Rightarrow 15\%$

RELATIVO A 100:

REBANHOS DE UMA FAZENDA – 1992

Espécies	**Quantidade** (cabeças)	**Porcentagem** %
Bovinos	860	59
Ovinos	354	24
Caprinos	30	2
Suínos	212	15
Total	1456	100

FONTE: DADOS FICTÍCIOS.

RELATIVO A 1:

REBANHOS DE UMA FAZENDA – 1992

Espécies	**Quantidade** (cabeças)	**Proporção**
Bovinos	860	0,590
Ovinos	354	0,243
Caprinos	30	0,020
Suínos	212	0,146
Total	1456	1

FONTE: DADOS FICTÍCIOS.

4.7.2 Índices

Os **índices** são razões entre duas grandezas tais que uma não inclui a outra.
EXEMPLO:

$$\text{densidade demográfica} = \frac{\text{população}}{\text{superfície}}$$

$$\text{renda per capita} = \frac{\text{renda}}{\text{população}}$$

4.7.3 Coeficientes

Os **coeficientes** são razões entre o número de ocorrências e o número total (ocorrências e não ocorrências).
EXEMPLO:

$$\text{coeficiente de natalidade} = \frac{\text{número de nascimentos}}{\text{população total}}$$

$$\text{coeficiente de evasão escolar} = \frac{\text{número de evadidos}}{\text{total inicial de matrículas}}$$

4.7.4 Taxas

As taxas são os coeficientes multiplicados por uma potência de 10 (10, 100, 1000, ...) para tornar o resultado mais legível.
EXEMPLO:
taxa de mortalidade = coeficiente de mortalidade $\times$ 1000

5 Gráficos estatísticos

O **gráfico estatístico** é uma forma de apresentação dos dados estatísticos cujo objetivo é o de produzir uma impressão mais rápida e viva do fenômeno em estudo.

A seguir são apresentados vários tipos de gráficos baseados na mesma série estatística apresentada na tabela abaixo.

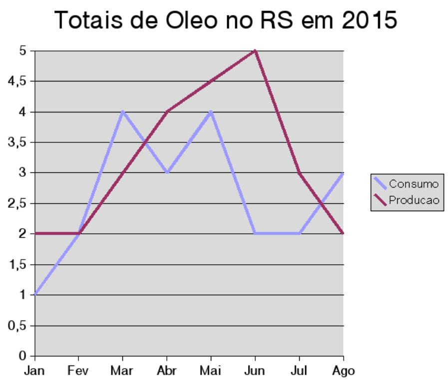
Totais de Óleo no RS em 2015

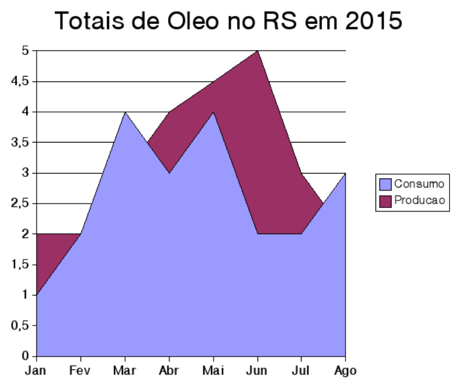
Meses	Consumo	Produção
Jan	1	2
Fev	2	2
Mar	4	3
Abr	3	4
Mai	4	4,5
Jun	2	5
Jul	2	3
Ago	3	2

FONTE: DADOS FICTÍCIOS.

5.1 Gráficos em linha ou curva

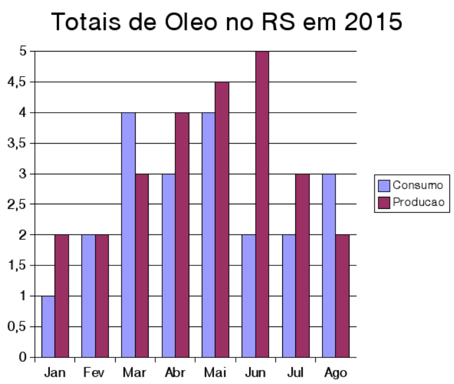
Este tipo de gráfico usa uma linha poligonal para representar a série estatística. Para ficar mais claro pode ser hachurado (preenchido).

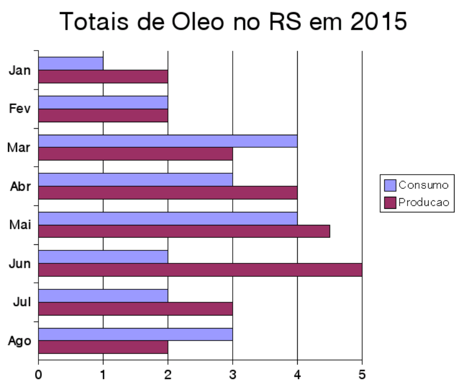




5.2 Gráficos em colunas ou em barras

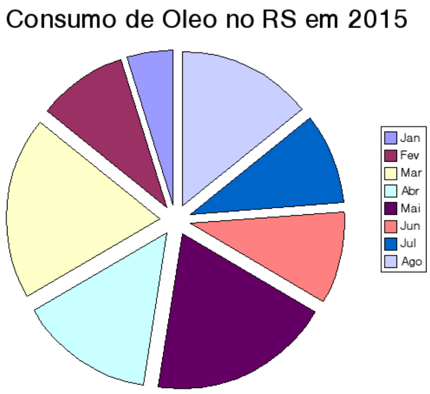
Este tipo de gráfico usa colunas para representar a série estatística. Podem ser verticais ou horizontais e conter barras múltiplas.

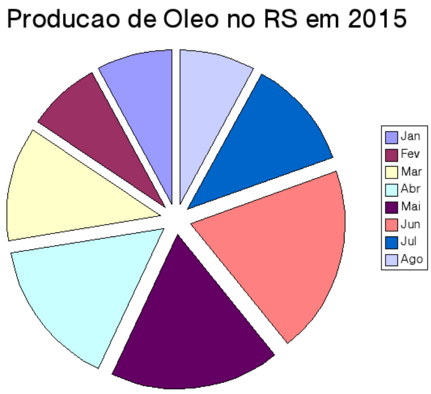




5.3 Gráfico em setores

É o tipo de gráfico construído com base num círculo. É útil para representar frações em relação ao total.





Parte II

6 Distribuição de Frequência

6.1 Tabela Primitiva e Rol

A tabela em que os elementos não foram organizados numericamente chama-se **tabela primitiva**. Por exemplo, considere o levantamento de dados da estatura de 40 alunos da escola A (variável x), cujos resultados, em centímetros, mostrados na tabela a seguir, estão colocados na sequência como foram obtidos.

Estatura de 40 Alunos da Escola A (cm)

166	160	161	150	162	160	165	167	164	160
162	161	168	163	156	173	160	155	164	168
155	152	163	160	155	155	169	151	170	164
154	161	156	172	153	157	156	158	158	161

O primeiro passo para a organização dos dados é ordená-los de forma crescente ou decrescente. A tabela assim organizada recebe o nome de **rol**.

Estatura de 40 Alunos da Escola A (cm)

150	154	155	157	160	161	162	164	166	169
151	155	156	158	160	161	162	164	167	170
152	155	156	158	160	161	163	164	168	172
153	155	156	160	160	161	163	165	168	173

A simples organização dos dados em um rol de ordem crescente já permite determinar diretamente o menor valor ($x = 150$ cm), o maior valor ($x = 173$ cm), o valor que mais ocorre ($x = 160$ cm), e a amplitude da variação (a distância entre o maior e o menor, $\Delta x = 173 - 150 = 23$ cm).

6.2 Distribuição de Frequência

Uma maneira mais concisa de mostrar os dados do rol é apresentar cada um seguido pelo número de vezes que ocorre, ao invés de repetí-los. O número de ocorrências de um determinado valor recebe o nome de **frequência**. Por exemplo, a estatura de 155 cm ocorre 4 vezes que se escreve $f(155) = 4$; a estatura de 150 ocorre 1 vez ou $f(150) = 1$.

A tabela que contém todos os valores com a sua frequência recebe o nome de **distribuição de frequência**. Veja abaixo uma distribuição de frequência construída a partir do rol anterior (separada em 3 partes):

Estat.	**Freq.**
150	1
151	1
152	1
153	1
154	1
155	4
156	3
157	1

Estat.	**Freq.**
158	2
160	5
161	4
162	2
163	2
164	3
165	1
166	1

Estat.	**Freq.**
167	1
168	2
169	1
170	1
172	1
173	1
Total	40

Ainda assim, o processo exige muito espaço em especial quando o número de valores da variável (n) aumenta. O mais razoável nestes casos, em especial quando a variável é contínua, é agrupar os valores por intervalos. Deste modo, ao invés de listar cada um dos valores que ocorrem, listam-se os intervalos de valores e a frequência correspondente, isto é, ao invés de colocar 1 aluno com 150 cm, 1 aluno com 151 cm, etc., coloca-se 4 alunos entre 150 e 154 cm. Este intervalo é escrito como $150 \vdash 154$ que corresponde a $150 \leqslant x < 154$ (a variável pode estar desde 150 inclusive até 154 exclusive), portanto valores 150, 150.1, 151, 152, 153, 153.5, 153.99 estariam neste intervalo, mas 154 não. Definindo o rol de acordo com intervalos, tem-se a seguinte tabela:

Estatura de 40 Alunos do Colégio A

Estaturas (cm)	**Frequência**
$150 \vdash 154$	4
$154 \vdash 158$	9
$158 \vdash 162$	11
$162 \vdash 166$	8
$166 \vdash 170$	5
$170 \vdash 174$	3
Total	40

Procedendo desta forma perde-se a informação detalhada das estaturas, mas ganha-se em simplicidade, pois a análise dos dados fica simplificada. Examinando a tabela acima, podemos facilmente verificar que a maioria dos alunos tem estaturas entre 154 e 166 cm e que uma minoria é menor que 154 cm ou maior que 170. Esta análise não é imediata da tabela em que todos os valores são listados. Por outro lado, se desejarmos saber quantos alunos tem 150 cm de altura, esta informação não estará disponível pois somamos os alunos de 150, 151, 152 e 153 cm numa única classe da distribuição de frequência.

Frequentemente procedemos desta forma numa análise estatística, pois o objetivo da estatística é justamente fazer o apanhado geral das características de um conjunto de dados, desinteressando-se por casos particulares.

6.3 Elementos de uma Distribuição de Frequência

6.3.1 Classe

As **classes** são intervalos de variação de uma variável. As classes são representadas simbolicamente por i, sendo $i = 1, 2, \ldots, k$, onde k é o número total de classes. O número total de valores é simbolizado por n.

Assim, no exemplo, o intervalo $154 \vdash 158$ define a segunda classe ($i = 2$), o intervalo $166 \vdash 170$ define a quinta classe ($i = 5$) e assim por diante. Como a distribuição tem seis classes, logo $k = 6$. a variável x assume 40 valores, logo $n = 40$.

6.3.2 Limites de Classe

Os **limites de classe** são os extremos de cada classe. Para uma determinada classe i, o **limite inferior** é simbolizado por l_i e o **limite superior** por L_i.

O limite inferior da segunda classe é escrito como $l_2 = 154$, enquanto o limite superior da segunda classe é escrito como $L_2 = 158$.

De acordo com o IBGE[1] as classes devem ser escritas como *desta quantidade até menor que aquela*, usando para isso o símbolo $\vdash$. Assim, $l_i \vdash L_i$ significa inclusão de l_i e exclusão de L_i. O indivíduo com estatura 158 cm estaria na terceira classe ($i = 3$) e não na segunda.

6.3.3 Intervalo de Classe

A **amplitude de um intervalo de classe** ou simplesmente **intervalo de classe** é o tamanho do intervalo que define a classe. O intervalo da classe i é simbolizado por h_i e é obtido pela diferença entre os seus limites:

$$h_i = L_i - l_i.$$

No exemplo que usamos, o tamanho do intervalo da segunda classe (h_2) vale

$$h_2 = L_2 - l_2 = 158 - 154 = 4 \text{ cm}.$$

Todos as outras classes do exemplo também tem intervalo de 4 cm, pois este é o intervalo entre cada um dos limites inferiores e os limites superiores correspondentes.

6.3.4 Amplitude Total da Distribuição

A **amplitude total** da distribuição (AT) é o intervalo total compreendido por todas as classes da distribuição, isto é, desde o limite inferior da primeira classe (l_1) até o limite superior da última classe (L_k). Matematicamente, escrevemos isso como

$$AT = L_k - l_1.$$

[1] Instituto Brasileiro de Geografi a e Estatística

Ainda no nosso exemplo, temos seis classes ($k = 6$). O limite superior da última classe ($i = 6$) vale $L_6 = 174$, enquanto o limite inferior da primeira classe ($i = 1$) vale $l_1 = 150$. Portanto,

$$AT = L_6 - l_1 = 174 - 150 = 24 \text{ cm}.$$

Numa distribuição em que as classes que possuem o mesmo intervalo, a amplitude total pode ser escrita como o intervalo de classe multiplicado pelo número de classes

$$AT = h_i\ k.$$

6.3.5 Amplitude Amostral

A **amplitude amostral** (AA) é o intervalo entre o maior valor ($m\acute{a}x(x)$) e o menor valor ($m\acute{i}n(x)$) dos dados da amostra:

$$AA = m\acute{a}x(x) - m\acute{i}n(x)$$

No exemplo a maior estatura é 173 e a menor 150, logo $AA = 173 - 150 = 23$ cm.

6.3.6 Ponto Médio de uma Classe

O **ponto médio de uma classe** é o ponto que divide a classe ao meio. O ponto médio da classe i é simbolizado por x_i e calculado efetuando-se a média entre os limites da classe:

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

No nosso exemplo, o ponto médio da segunda classe é

$$x_2 = \frac{l_2 + L_2}{2} = \frac{154 + 158}{2} = 156 \text{ cm}.$$

O ponto médio de uma classe é o valor representativo da classe.

6.3.7 Frequência Simples ou Absoluta

A **frequência simples** ou **frequência absoluta** ou simplesmente **frequência** de uma classe ou de um valor individual é o número de vezes que o valor ocorre numa amostra. A frequência da classe i é representada por f_i. Assim, no exemplo temos

$$f_1 = 4, \quad f_2 = 9, \quad f_3 = 11, \quad f_4 = 8, \quad f_5 = 5, \quad f_6 = 3.$$

A soma de todas as frequências é representada pelo símbolo de somatório ($\sum$). $\sum_{i=1}^{k} f_i$ significa a soma dos f_i sendo que i vai desde 1 até k. Pode-se entender que a soma de todas as frequências é igual ao número total de valores na amostra:

$$\sum_{i=1}^{k} f_i = n.$$

Quando não há dúvidas, podemos escrever simplesmente:

$$\sum f_i = n$$

No nosso exemplo, escrever $\sum_{i=1}^{6} f_i$ é como escrever $f_1 + f_2 + f_3 + f_4 + f_5 + f_6$, ou seja:

$$\sum_{i=1}^{6} f_i \;=\; f_1 + f_2 + f_3 + f_4 + f_5 + f_6 \;=\; 4 + 9 + 11 + 8 + 5 + 3 \;=\; 40.$$

Neste ponto podemos reescrever a distribuição de frequência com a seguinte representação técnica da tabela:

Estatura de 40 Alunos do Colégio A

i	**Estaturas** (cm)	f_i
1	$150 \vdash 154$	4
2	$154 \vdash 158$	9
3	$158 \vdash 162$	11
4	$162 \vdash 166$	8
5	$166 \vdash 170$	5
6	$170 \vdash 174$	3
		$\sum f_i = 40$

6.4 Determinação do Número de Classes e Intervalos de Classe

Quando dispomos de uma tabela primitiva ou de um rol, precisamos estabelecer a quantidade e o intervalo das classes que vamos criar, de outro modo a distribuição de frequência pode não ser útil para a nossa análise.

Uma das maneiras de determinar o número de classes é usando a **Regra de Sturges** que determina k em função de n:

$$k \simeq 1 + 3.3\log(n)$$

onde k é o número de classes e n o número de dados. Da mesma forma podemos usar outra regra que associa k e n de outra forma:

$$k \simeq \sqrt{n}.$$

No nosso exemplo, usando a Regra de Sturges temos $n = 40$, logo $k = 1 + 3.3\log(40) = 6.28 \simeq 6$, portanto utilizamos 6 classes. Com a outra regra, temos $k = \sqrt{40} = 6.32 \simeq 6$, cujo resultado para o número de classes é o mesmo.

Sabendo o número de classes (k) que vamos usar, podemos determinar o intervalo de classes através da amplitude total da distribuição (AT)

$$h \simeq \frac{AT}{k}.$$

Nas equações acima foi usado o símbolo de aproximadamente ($\simeq$) ao invés de igualdade ($=$) porque estas fórmulas representam valores típicos a serem usados mas que podem ser alterados ligeiramente de acordo com o objetivo da distribuição ou para evitar classes com frequências nulas enquanto outras tem valores muito altos. Com relação ao intervalo de classe, lembre-se que a amplitude total (AT) deve ser ligeiramente maior que a amplitude amostral (AA) para que a distribuição tenha intervalos para incluir todos os valores da amostra.

6.5 Tipos de Frequências

6.5.1 Frequências Simples ou Absoluta

Frequências simples ou absoluta (f_i) são os valores que diretamente representam o número de dados de cada classe. A soma de todas as ocorrências em cada classe é igual ao número total de dados:

$$\sum_{i=1}^{k} f_i = n.$$

6.5.2 Frequências Relativas

Frequências relativas (fr_i) são as razões entre as frequências simples (f_i) e a frequência total (n):

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

A frequência relativa de uma classe mostra a parcela que aquela classe representa da amostra. Assim, a frequência relativa da terceira classe do nosso exemplo é:

$$fr_3 = \frac{f_3}{\sum f_i} = \frac{11}{40} = 0.275,$$

então a terceira classe corresponde a uma fração de 0.275 do total ou 27.5 %.

6.5.3 Frequência Acumulada

Frequência acumulada (F_i) é a soma das frequências simples de todas as classes com intervalos inferiores a um determinada classe:

$$F_j = \sum_{i=1}^{j} f_i = f_1 + f_2 + \ldots + f_j.$$

Assim, ainda no exemplo dos alunos, a frequência acumulada correspondente à terceira classe é

$$F_3 = \sum_{i=1}^{3} f_i = f_1 + f_2 + f_3 = 4 + 9 + 11 = 24,$$

que significa que existem 24 alunos com estatura inferior a 162 cm (limite superior da terceira classe.)

6.5.4 Frequência Acumulada Relativa

Frequência acumulada relativa (Fr_i) é a frequência acumulada da classe dividida pela frequência total da distribuição:

$$Fr_i = \frac{F_i}{\sum f_i} = \frac{F_i}{n}.$$

Logo, para a terceira classe, temos:

$$Fr_3 = \frac{F_3}{n} = \frac{24}{40} = 0.6$$

que significa que a fração de 0.6 alunos (ou 60%) tem estaturas inferiores à 162 cm (limite superior da terceira classe.)

A tabela completa do nosso exemplo fica assim:

Estatura de 40 Alunos do Colégio A

i	**Estaturas** (cm)	x_i	f_i	fr_i	F_i	Fr_i
1	$150 \vdash 154$	152	4	0.100	4	0.100
2	$154 \vdash 158$	156	9	0.225	13	0.325
3	$158 \vdash 162$	160	11	0.275	24	0.600
4	$162 \vdash 166$	164	8	0.200	32	0.800
5	$166 \vdash 170$	168	5	0.125	37	0.925
6	$170 \vdash 174$	172	3	0.075	40	1.000
			$\sum f_i = 40$	$\sum fr_i = 1$		

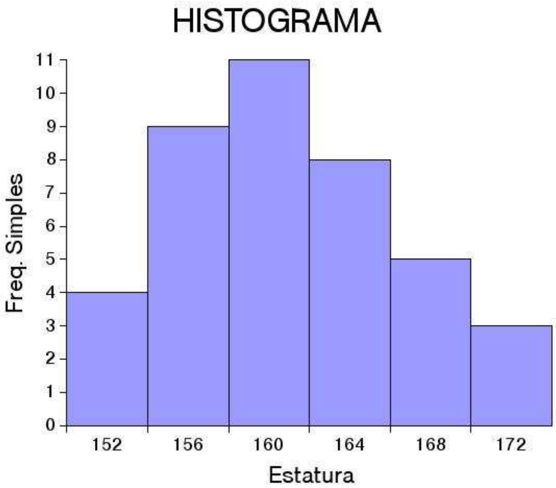
Examinando a tabela, vemos por exemplo que a terceira classe corresponde a maior fração de alunos ($fr_3 = 0.275$), isto é, a maioria dos alunos tem estatura entre 158 cm (inclusive) e 162 cm (exclusive). Também e'possível ver que 80% dos alunos têm estatura inferior a 166 cm pois a frequência acumulada até a quarta classe ($L_4 = 166$) é 0.800 que corresponde a 80%.

6.6 Representações Gráfi cas de um Distribuição

6.6.1 Histograma

O **histograma** é formado por um conjunto de retângulos justapostos cujas bases se localizam sobre o eixo horizontal, de tal modo que os seus pontos médios coincidam com os pontos médios dos intervalos de classe e seus limites coincidam com os limites da classe.

Um histograma para a frequência simples é mostrado abaixo:


HISTOGRAMA
Freq. Simples
0
1
2
3
4
5
6
7
8
9
10
11
152
156
160
164
168
172
Estatura

Parte III

7 Medidas de Posição

7.1 Média Aritmética

A **média aritmética**, simbolizada por $\bar{x}$, é o quociente entre a soma dos valores de uma variável pelo número de valores

Média simples – dados não agrupados:

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

Média ponderada – dados agrupados:

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

Desvio da média – a diferença entre o valore e a média:

$$d_i = x_i - \bar{x}$$

7.1.1 Propriedades

Desvios A soma dos desvios é nula:

$$\sum_{i=1}^{n} d_i = 0$$

Constante aditiva Somando uma constante C a todos os valores de uma variável, a média do conjunto fica aumentada desta constante[2]:

$$\text{valores: } x_i \rightarrow \text{média: } \bar{x}$$

$$\text{valores: } y_i = (x_i + C) \rightarrow \text{média: } \bar{y} = (\bar{x} + C)$$

[2]Subtrair equivale a somar uma constante negativa ($-C$)

Constante multiplicativa Multiplicando uma constante C a todos os valores de uma variável, a média do conjunto fica multiplicada desta constante[3]:

$$\text{valores: } x_i \rightarrow \text{média: } \bar{x}$$

$$\text{valores: } y_i = (C\, x_i) \rightarrow \text{média: } \bar{y} = (C\, \bar{x})$$

EXEMPLOS:

1. A média do conjunto $x_i = \{1, 2, 3, 4, 5\}$ é $\bar{x} = 3$, enquanto a média do conjunto $y_i = x_i + 10 = \{11, 12, 13, 14, 15\}$ é $\bar{y} = \bar{x} + 10 = 13$

2. A média do conjunto $x_i = \{1, 2, 3, 4, 5\}$ é $\bar{x} = 3$, enquanto a média do conjunto $y_i = 10\, x_i = \{10, 20, 30, 40, 50\}$ é $\bar{y} = 10\, \bar{x} + 10 = 30$

3. A média simples de uma vaca cuja produção ao longo de 7 dias é $10, 14, 13, 15, 16, 18, 12$ litros cada dia vale:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{10 + 14 + 13 + 15 + 16 + 18 + 12}{7} = \frac{98}{7} = 14.$$

Assim, uma outra vaca que produzisse 14 litros de leite em todos os 7 dias teria produzido, no final, o mesmo que esta vaca cuja produção tenha sido $10, 14, 13, 15, 16, 18, 12$.

4. O desvio da média dos valores $10, 14, 13, 15, 16, 18, 12$ da produção de leite de uma determinada vaca são:

$$\begin{aligned}
d_1 = x_1 - \bar{x} &= 10 - 14 = -4 \\
d_2 = x_2 - \bar{x} &= 14 - 14 = 0 \\
d_3 = x_3 - \bar{x} &= 13 - 14 = -1 \\
d_4 = x_4 - \bar{x} &= 15 - 14 = 1 \\
d_5 = x_5 - \bar{x} &= 16 - 14 = 2 \\
d_6 = x_6 - \bar{x} &= 18 - 14 = 4 \\
d_7 = x_7 - \bar{x} &= 12 - 14 = -2.
\end{aligned}$$

Somando todos os desvios:

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} (x_i - \bar{x}) = -4 + 0 - 1 + 1 + 2 + 4 - 2 = 0,$$

conforme esperado pela propriedade.

5. A média ponderada é aplicada quando os dados já estão agrupados. Considere a tabela

[3]Dividir equivale a multiplicar pelo inverso de uma constante ($\frac{1}{C}$)

Núm. alunos	0	1	2	3	4	
f_i	2	6	10	12	4	$\sum = 34$

A média vale:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} f_i\, x_i = \frac{2 \cdot 0 + 6 \cdot 1 + 10 \cdot 2 + 12 \cdot 3 + 4 \cdot 4}{34} = \; = \frac{78}{34} = 2.29.$$

Note que $n = \sum_i f_i$.

7.2 Moda

A **moda** (Mo) é o valor que ocorre com mais frequência na distribuição.

Quando os dados estão agrupados em classes, a moda correponde a frequência simples mais alta e o valor da moda é tomado como o ponto médio do intervalo da classe. Se os limites inferior e superior da classe mais frequente são l_* e L_*, a moda será $(l_* + L_*)/2$

EXEMPLO: Considere os seguintes salários: 100, 90, 110, 100, 100, 2500. A moda é o valor que mais ocorre, $Mo = 100$. Neste caso a média é $\bar{x} = 500$.

Distribuição modal: é aquela que possui uma só moda.
$x_i = \{100, 90, 110, 100, 100, 2500\} \rightarrow Mo = 100$.

Distribuição bimodal: possui duas modas.
$x_i = \{100, 200, 100, 100, 150, 210, 200, 120, 200\} \rightarrow Mo = 100$ e $Mo = 200$.

Distribuição amodal: não possui moda.
$x_i = \{1, 2, 3, 6, 7, 22, 300\} \rightarrow \nexists\, Mo$

7.3 Mediana

A **mediana** ou **valor mediano** (Md) é o valor que divide a série ordenada em dois conjuntos com o mesmo número de valores. Se a série tem um número ímpar de valores, a mediana é o valor que está no meio (ponto mediano) da série. Se a série tem um número par de valores, então utiliza-se como mediana o valor médio entre os dois valores que estão no meio da série.

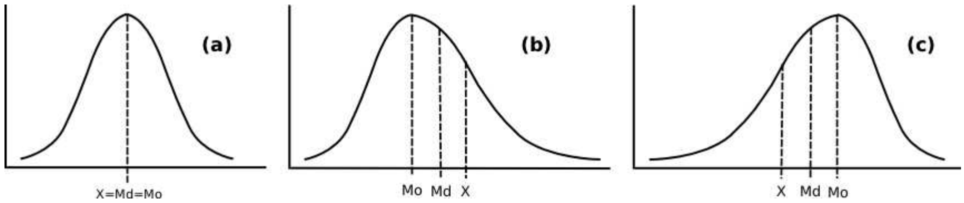
EXEMPLO:

Na série ordenada $\{2, 5, 6, 8, 10, 13, 15, 16, 18\}$, temos que $Md = 10$ pois abaixo de 10 temos 4 números (2, 5, 6, 8) e acima de 10 também 4 (13, 15, 16, 18).

Na série ordenada $\{1, 3, 6, 8, 9, 10\}$, temos que $Md = (6 + 8)/2 = 7$, pois não há um só número no centro da série, assim utilizamos o valor médio dois números centrais.

7.4 Posição relativa da média, mediana e moda

No caso de um distribuição simétrica (caso **(a)** na figura abaixo), as média, a mediana e a moda tem o mesmo valor. Entretanto, se a distribuição apresenta alguma tendência para valores positivos ou negativos, as medidas de posição poderão diferir. No caso de uma distribuição assimétrica positiva (caso **b** abaixo) tem-se que $Mo < Md < \bar{x}$. No caso de um distribuição assimétrica negativa, tem-se que $\bar{x} < Md < Mo$.



Distribuições: **(a)** simétrica, **(b)** assimétrica positiva e **(c)** assimétrica negativa.

8 Medidas de dispersão

8.1 Variância e desvio padrão

Para determinar a dispersão de uma série de medidas poder-se-ia usar a soma de todos os desvios $d_i = x_i - \bar{x}$ dos valores com relação a média dividido pelo número de valores, assim obtendo uma média dos desvios. Entretanto, como esta soma é nula ($\sum_i d_i = 0$), usa-se a soma dos desvios ao quadrado, pois elevando-se ao quadrado, perde-se a informação do sinal. Deste modo, define-se a **variância** como

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

Além disso, como a variância é uma medida que envolve o quadrado das quantidades, é comum usar a raiz quadrada da variância, chamado de **desvio padrão**:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Como maneiras alternativas de calculas o desvio padrão ou a variância podemos usar a equação:

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

que em algumas situações evita arredondamentos usados nas outras equações produzindo resultados mais precisos.

EXEMPLO:

Considere a série a seguir: $x_i = \{8, 10, 12, 9, 11, 7, 13\}$, para a qual $\bar{x} = 10$. Para calcular a variância e o desvio padrão é útil construirmos uma tabela com os desvios:

x_i	$d_i = x_i - \bar{x}$	d_i^2
8	-2	4
10	0	0
12	2	4
9	-1	1
11	1	1
7	-3	9
13	3	9
	$\sum d_i = 0$	$\sum d_i^2 = 28$

Deste modo, o desvio padrão pode ser calculado:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} = \sqrt{\frac{1}{6}\, 28} \simeq 2.16$$

média: $\bar{x} = 10$

desvio padrão: $s \simeq 2.16$

Do mesmo modo, considere a série $y_i = \{10, 11, 9, 10, 10, 9, 11\}$, que tem a mesma média $\bar{y} = 10$. e para cuja tabela de desvios tem-se:

x_i	$d_i = x_i - \bar{x}$	d_i^2
10	0	0
11	1	1
9	-1	1
10	0	0
10	0	0
9	-1	1
11	1	1
	$\sum d_i = 0$	$\sum d_i^2 = 4$

Procedendo da mesma forma, calcula-se o desvio padrão

$$s = \sqrt{\frac{1}{6}\, 4} \simeq 0.81$$

média: $\bar{x} = 10$

desvio padrão: $s \simeq 0.81$

8.2 Interpretação do desvio padrão

O desvio padrão indica a dispersão dos dados dentro da amostra, isto é, o quanto os dados em geral diferem da média. Quanto menor o desvio padrão, mais parecidos são os valores da série estatística. Nos exemplos acima, nota-se que tanto a série x_i quanto y_i têm o mesmo número de dados e ambas tem a mesma média, entretanto o desvio padrão de x_i é bem maior que de y_i, que indica que os dados em x_i estão mais afastados da média que em y_i. De fato, se examinamos as séries, vemos que em x_i há valores que estão até 3 unidades afastadas da média (7 e 13), enquanto na série y_i o maior afastamento é de 1 unidade (9 e 11).

Numa distribuição normal e simétrica, o desvio padrão é calculado dá uma ideia de onde estão localizados os valores da amostra, em torno da média, da seguinte maneira:

- 68% dos valores da série estão até 1 desvio padrão de distância da média, isto é, estão entre $\bar{x} - s$ e $\bar{x} + s$.
- 95% dos valores da série estão até 2 desvios padrão de distância da média, isto é, estão entre $\bar{x} - 2s$ e $\bar{x} + 2s$.
- 99.7% dos valores da série estão até 3 desvios padrão de distância da média, isto é, estão entre $\bar{x} - 3s$ e $\bar{x} + 3s$.

Assim, para simplificar, assuma uma série estatística relativa a alguma medida de uma população e cujos valores tem média $\bar{x} = 100$ e desvio padrão $s = 10$. De acordo com as afirmações acima, podemos dizer que 68% da amostra tem valores entre 90 (100-10) e 110 (100+10); da mesma forma, podemos dizer que 95% da amostra tem valores que se situam entre 80 ($100 - 2 \cdot 10$) e 120 ($100 + 2 \cdot 10$); finalmente, 99.7% situa-se entre 70 ($100 - 3 \cdot 10$) e 130 ($100 - 3 \cdot 10$).

A média de uma série estatística frequentemente é especificada mostrando-se o desvio padrão junto, da seguinte forma:

$$\bar{x} \pm s$$

que indica a dispersão da amostra. Nos exemplos acima, ter-se-ia especificado 10 ± 2.16 para x_i e 10 ± 0.81 para y_i.

No caso de uma série de medidas de uma mesma quantidade, o desvio padrão indica a incerteza nas medidas, ou o erro associado. Por isso, pode-se usar o desvio padrão para determinar os algarismos significativos de uma série de medidas. Por exemplo, se para várias medidas de uma mesma quantidade em laboratório obteve-se para valor de média $\bar{x} = 15.943$ e para desvio padrão $s = 2$. Um trabalhador descuidado escreveria $\bar{x} \pm s = 15.943 \pm 2$, entretanto o significado deste desvio padrão é que não temos certeza se a média é na verdade 13 (15-2) ou 17 (15+2), então como poderíamos saber sobre as três casas decimais mostradas? Realmente, a parte decimal 0.943 deveria ser desprezada e escrever-se-ia somente $\bar{x} \pm s = 15 \pm 2$. No mesmo caso, se o desvio padrão fosse $s = 0.2$ então poderíamos escrever a média até a mesma casa do desvio padrão, isto é, $\bar{x} \pm s = 15.9 \pm 0.2$, e assim por diante.

8.3 Coeficiente de variação

Para comparar a variação do desvio padrão com a média, usa-se a razão entre o desvio padrão e a média, chamado de **coeficiente de variação**, que muitas vezes é multiplicado por 100 para dar o resultado em porcentagem:

$$CV = \frac{s}{\bar{x}} \cdot 100$$

Por exemplo, se a média vale $\bar{x} = 980$ e o desvio padrão $s = 56$, temos

$$CV = \frac{56}{980} \cdot 100 = 5.7\%,$$

que indica a dispersão da amostra.

Apêndices

A Exercícios Parte I

POPULAÇÃO E AMOSTRA

1. Uma escola abriga 124 alunos. Obtenha uma amostra representativa correspondendo a 15% da população. Descreva o seu método de escolha da amostra.

2. O diretor de uma escola, na qual estão matriculados 280 meninos e 320 meninas, deseja fazer um levantamento das condições de vida dos estudantes. Para diminuir os custos, resolve fazer um levantamento com 10% dos estudantes. Obtenhas os componentes da amostra para esta população.

3. Mostre como seria possível retirar uma amostra de 32 elementos de uma população ordenada formada por 1920 elementos.

4. Os seguintes bairros de uma cidade apresentam o quadro de eleitores abaixo: Obtenha uma amostra proporcional estratificada de 120 eleitores.

Bairro	Homens	Mulheres
A	80	95
B	102	120
C	110	92
D	134	228
E	150	130
F	300	290
Total	876	955

SÉRIES ESTATÍSTICAS

5. Considere a seguinte série estatística: Complete-a, determinando as porcen-

Série	Alunos matriculados	%
1^a	546	
2^a	328	
3^a	280	
4^a	120	
Total	1274	

tagens com uma casa decimal e fazendo a compensação, se necessário.

6. Considerando que Minas Gerais, em 1992, apresentou os seguintes dados (IBGE):

população: 15.957,6 mil habitantes;
superfície: 586.624 km^2;
nascimentos: 292.036;
óbitos: 99.281;

Calcule:
a) o índice de densidade demográfica;
b) a taxa de natalidade;
c) a taxa de mortalidade;

7. Uma escola apresenta, no final do ano, o seguinte quadro:

	Matrículas	
Séries	**Março**	**Novembro**
1^a	480	475
2^a	458	456
3^a	436	430
4^a	420	420
Total	1794	1781

Calcule:
a) a taxa de evasão por série;
b) a taxa de evasão da escola;

8. Considere que um rebanho de ovelhas é constituído por animais que possuem os seguintes pesos em kg:

```
21 14 21 7 25 13 12 27 19 26 15 14 6 27 11 7 26 24 27 12
29 21 20 9 3 12 28 21 9 21 28 13 20 15 25 23 9 26 13 6 4
23 17 13 17 19 19 26 10 4 28 6 22 5 11 17 8 23 9 24
```

Faça os procedimentos estatísticos nos itens abaixo, explicando suas escolhas em cada passo detalhadamente. Para cada um dos itens, apresente os indivíduos escolhidos ordenados, determine o menor e maior valor e a média da amostra.
a) Uma amostra de 10% da população
b) Uma amostra de 50% da população
c) Uma amostra de 75% da população
d) A população inteira.
Refaça os cálculos para toda a população e compare os seus resultados com aqueles determinados para as amostras.

B Exercícios Parte II

DISTRIBUIÇÃO DE FREQUÊNCIA

1. As notas obtidas por 50 alunos em uma classe foram

1	2	3	4	5	6	6	7	7	8
2	3	3	4	5	6	6	7	8	8
2	3	4	4	5	6	6	7	8	9
2	3	4	5	5	6	6	7	8	9
2	3	4	5	5	6	7	7	8	9

a) Complete a distribuição de frequência abaixo:

i	**Notas**	x_i	f_i	fr_i	F_i	Fr_i
1	$0 \vdash 2$	1	1			
2	$2 \vdash 4$					
3	$4 \vdash 6$					
4	$6 \vdash 8$					
5	$8 \vdash 10$					
			$\sum f_i =$	$\sum fr_i =$	$\sum F_i =$	$\sum Fr_i =$

b) Responda:

1. Qual a amplitude amostral?
2. Qual a amplitude da distribuição?
3. Qual o número de classes da distribuição?
4. Qual o limite inferior da quarta classe?
5. Qual o limite superior da classe 2?
6. Qual a amplitude do segundo intervalo de classe?

c) Complete (mostrando os cálculos):

1. $h_3 =$
2. $n =$
3. $l_1 =$
4. $L_3 =$
5. $x_2 =$
6. $f_5 =$
7. $k =$
8. $\sum_{i=1}^{5} f_i =$

2. Complete a tabela abaixo:

Número de filhos para uma amostra de famílias

i	x_i	f_i	fr_i	F_i	Fr_i
1	0	1	$\ldots$	$\ldots$	$\ldots$
2	1	$\ldots$	0.15	4	$\ldots$
3	2	4	$\ldots$	$\ldots$	$\ldots$
4	3	$\ldots$	0.25	13	$\ldots$
5	4	3	0.15	$\ldots$	$\ldots$
6	5	2	$\ldots$	18	$\ldots$
7	6	$\ldots$	$\ldots$	19	$\ldots$
8	7	$\ldots$	$\ldots$	$\ldots$	$\ldots$
		$\sum f_i = 20$	$\sum fr_i = \ldots$		

Baseando-se nesta tabela responda as perguntas:

a. Quantos famílias tem 2 filhos?

b. Qual a fração de famílias com 4 filhos? E a porcentagem?

c. Quantas famílias têm até 3 filhos?

d. Quantas famílias têm mais de 5 filhos?

e. Quantas famílias têm até 7 filhos?

C Exercícios Parte III

MOMENTOS DA DISTRIBUIÇÃO

1. Para as séries abaixo, calcule a moda, a média, a mediana, o desvio padrão e o coeficiente de variacão:

- 1, 3, 3 , 5, 7, 9, 11
- 20, 14, 15, 19, 21, 22, 20
- 17.9, 22.5, 13.3, 16.8, 15.4, 14.2
- -10, -6, 2, 3, 7, 9, 10, 8, -2, 0, 8, 2, 3, 2
- 87, 82 ,81, 93, 94, 78, 99, 80, 82, 88, 82, 83

2. Um experimento de laboratório é realizado para medir a viscosidade do azeite, obtendo-se os seguintes valores: 0,040; 0,041; 0,042; 0,039; 0,041 e 0,039 mš/s. Calcule o valor médio, a variança e o desvio-padrão.

3. Dois torneiros, Paulo e João, concorrendo a uma vaga em uma metalúrgica, submeteram-se ao seguinte teste de precisão: cada um deles construiu quatro rodas de ferro, que deveriam ter 5 cm de diâmetro. A tabela abaixo descreve o desempenho de cada um.

	Diâmetro (roda 1)	Diâmetro (roda 2)	Diâmetro (roda 3)	Diâmetro (roda 4)
Paulo	4,8	5,2	5,0	5,0
João	4,7	5,3	5,0	5,0

Qual foi o concorrente mais regular?

4. Um atirador de ferraduras localiza-se a 30m de seu alvo. Os resultados dos lançamentos são:

Lançamento	1	2	3	4	5	6	7	8	9	10
Desvio do alvo (m)	0,0	3,0	-4,2	0,0	1,5	2,4	-2,6	3,5	2,7	0

a) qual é a distância média ao alvo atingida pelo jogador?
b) qual é o desvio padrão?
c) O que pode dizer a respeito da qualidade do jogador?

5. A seguir, apresentam-se algumas estimativas para a velocidade da luz, determinadas por Michelson em 1882 (Statistics and Data Analysis, Siegel):

```
299.96 299.88 299.90 299.94 299.88
299.96 299.85 299.94 299.80 299.84
```

Utilizando uma máquina que só admite números até 6 digitos **a)** Determine a média **b)** Determine o desvio padrão, utilizando a expressão da definição, abaixo:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

c) Determine o desvio padrão, utilizando a fórmula deduzida para efeitos de cálculo, abaixo, e compare o resultado com o obtido no item anterior. Qual a resposta correta ?

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

d) Subtraia 299 de cada um dos dados e determine o desvio padrão, dos resultados obtidos, utilizando a fórmula utilizada na alínea anterior. Comente os resultados obtidos.
e) Calcule a média dos valores com que trabalhou no item d). Adicione à média obtida 299. Compare-a com a obtida no item a).

6.

Considere os seguintes dados de diâmetro de laranjas (em mm)

40	42	45	45	48	49	50	50	50	51
51	52	55	55	57	58	59	59	60	60
60	61	62	62	64	64	64	64	64	65
65	66	67	68	68	68	69	71	71	72
72	73	75	75	78	78	79	80	80	81
83	85	87	88	89	91	92	93	96	96
98	100	101	101	101	102				

Calcular:
a) média
b) mediana
c) moda
d) variância
e) desvio padrão.

D Tabela de Números Aleatórios

Tabela construída de modo que 10 algarismos (0 a 9) são distribuídos ao acaso nas linhas e colunas. Gerada pelo seguinte programa Perl (a cada invocação do programa, a sequência de números será diferente):

```
#!/usr/bin/perl -w

for ($j=0; $j<50; $j++){
        for ($i=0; $i<50; $i++){
                print int(rand(10)), " ";
        }
        print "\n";
}
```

```
6 5 0 2 1 2 6 5 7 6 5 1 4 3 7 5 2 7 7 0 5 7 3 0 1 2 3 0 7 3 4 3 3 0 6 5 4 0 1 6 4 6 7 6 2 4 4 0 5 3
5 6 5 3 0 9 1 9 0 9 7 9 6 4 5 8 7 0 5 1 3 3 7 8 7 1 9 3 4 7 8 1 9 8 1 5 2 5 2 3 2 7 9 7 8 8 5 3 1 5
5 0 6 8 3 4 3 0 2 4 3 9 6 8 8 8 7 7 5 8 1 1 8 8 0 0 6 3 4 1 2 2 5 0 1 3 0 8 7 5 0 0 7 5 4 2 6 7 7 0
9 5 3 2 7 4 8 1 4 1 5 3 5 1 8 5 6 0 2 4 0 1 1 6 1 6 6 8 4 6 3 0 8 4 8 6 6 0 6 6 9 7 7 5 8 4 9 9 3 2
1 4 3 5 3 0 1 9 6 6 2 7 7 7 4 5 3 8 4 6 3 9 8 3 9 2 9 1 6 8 2 2 0 0 5 5 0 8 6 4 9 0 9 1 1 2 0 1 4 5
5 6 6 0 9 3 3 0 4 0 4 5 0 9 0 1 8 5 5 3 9 5 6 8 5 7 7 4 0 1 7 8 6 5 1 3 4 5 6 5 1 3 4 8 0 4 0 1 4 3
5 7 1 1 5 6 5 6 8 5 9 4 0 9 0 8 0 7 4 9 7 5 8 4 4 5 0 4 8 3 6 9 0 9 0 1 8 4 0 0 0 0 0 7 1 0 1 3 5 4
3 4 9 1 0 9 2 6 6 0 0 2 8 9 5 5 6 0 4 2 9 4 7 5 6 3 4 7 4 8 2 8 7 7 5 7 0 3 1 5 8 0 3 1 1 6 0 3 2 7
8 9 3 9 9 7 2 1 2 3 4 6 8 1 5 4 2 8 2 3 0 5 4 0 7 3 1 7 9 2 4 9 0 2 1 7 4 6 0 8 6 2 7 6 9 2 1 2 7 3
6 0 9 9 2 1 9 1 7 4 0 2 2 0 2 8 6 6 0 9 9 4 1 1 9 8 0 6 0 2 7 0 3 0 6 6 8 9 1 9 7 8 3 9 5 1 5 1 1 5
7 1 5 5 8 1 5 5 6 5 6 3 3 4 8 2 2 3 7 5 4 9 7 4 9 2 0 3 8 4 3 6 3 0 5 4 5 4 4 0 3 3 1 8 7 3 1 9 1 0
2 7 8 8 2 2 8 4 1 0 1 9 1 7 2 6 5 0 7 3 9 8 5 9 9 5 2 5 4 3 2 7 7 3 7 0 2 2 6 4 5 9 4 5 5 5 1 2 0 4
4 7 9 6 2 5 1 8 8 3 9 8 8 1 6 3 0 5 8 6 1 0 7 7 9 8 7 1 3 7 1 8 0 1 1 1 5 8 8 7 5 0 2 2 1 9 4 6 6 0
5 5 7 8 4 6 5 6 8 6 3 2 0 0 1 5 6 5 2 8 9 2 1 5 6 9 0 1 8 3 8 2 6 5 7 7 0 0 3 3 6 8 3 3 2 4 9 2 0 1
1 6 4 8 1 8 3 0 3 1 2 3 8 0 1 5 8 7 6 0 8 6 0 7 9 7 2 4 7 1 3 2 7 4 6 4 9 2 6 0 2 9 4 7 7 2 4 3 1 5
8 3 9 1 7 2 3 1 0 4 4 6 4 3 4 4 1 0 7 9 1 3 8 1 5 3 1 4 1 9 0 1 0 1 4 4 1 3 2 9 6 7 7 3 8 2 4 1 4 3
2 3 0 6 6 1 4 1 3 3 9 8 3 1 3 7 2 3 3 7 8 9 5 1 1 8 5 7 8 1 3 0 3 5 6 2 6 1 6 3 2 6 5 9 6 9 5 3 7 4
3 1 2 1 9 6 4 5 0 7 8 7 3 0 7 9 1 3 7 4 4 5 2 3 6 5 2 6 1 8 5 4 7 9 6 8 2 6 1 8 0 0 5 0 1 4 0 5 2 1
8 5 4 5 9 3 2 8 9 8 4 4 9 4 7 5 4 9 4 4 7 5 2 5 9 0 4 1 2 0 6 0 4 1 5 3 6 4 5 5 0 5 3 1 8 7 7 9 0 3
5 2 3 8 4 1 2 6 3 9 5 9 8 5 4 2 7 1 0 1 6 3 7 9 1 1 8 1 0 7 2 4 0 1 0 5 3 4 6 5 3 1 1 9 8 0 1 0 5 5
1 9 4 8 0 4 3 8 3 4 8 8 6 4 7 7 7 2 1 7 7 9 3 3 2 5 1 5 8 4 9 0 5 5 0 6 7 0 2 4 7 4 3 5 7 5 7 0 8 7
7 2 5 4 0 2 9 6 6 1 3 5 7 1 3 0 8 2 1 5 9 7 8 4 8 4 0 5 7 6 6 7 1 0 0 5 4 5 6 2 2 2 4 5 4 1 4 9 1 5
3 1 0 1 4 6 2 3 6 2 7 7 5 8 6 1 9 9 5 1 4 4 1 4 0 5 0 2 1 3 9 4 7 1 9 3 5 2 5 5 5 3 9 4 7 0 3 2 3 0
5 4 0 2 5 2 2 8 6 7 0 8 7 9 2 9 5 2 1 6 2 2 0 9 6 3 0 1 2 6 1 8 2 4 2 9 3 6 9 9 1 4 2 2 4 6 5 3 6 3
3 2 1 7 6 8 3 3 1 0 0 7 3 7 1 7 2 3 5 0 0 3 5 7 5 4 7 1 3 5 9 6 4 4 5 2 4 1 5 2 2 3 1 9 2 0 5 8 5 1
7 3 5 1 1 5 5 3 5 3 6 7 7 3 7 7 9 4 8 6 5 4 5 5 9 6 3 3 6 9 0 3 4 3 0 2 9 2 9 0 3 8 5 1 1 0 9 5 2 1
6 0 5 4 3 4 8 2 3 3 1 2 6 1 3 0 7 6 3 0 3 8 1 2 9 7 6 9 8 2 0 3 7 0 9 8 8 5 6 0 5 5 5 6 8 5 4 2 8 7
0 8 5 1 3 6 0 9 6 5 9 2 1 4 2 9 3 9 2 6 3 5 3 6 3 1 2 3 6 9 1 9 9 3 3 6 9 5 5 5 0 2 7 6 8 7 6 2 9 8
3 1 4 8 7 0 0 1 3 2 1 8 8 9 6 7 0 6 3 7 6 7 1 5 1 0 2 4 4 3 6 5 1 3 3 3 5 4 2 3 4 8 6 0 6 1 7 5 7 2
5 4 2 2 0 3 0 7 2 5 3 9 5 3 5 7 4 0 6 4 5 4 2 3 4 2 9 9 5 5 1 9 1 2 0 4 5 9 2 9 1 4 6 5 6 2 0 0 7 1
7 0 7 1 6 3 5 7 0 0 5 9 7 6 3 7 3 3 9 1 6 9 2 7 2 7 2 8 3 9 2 9 2 9 6 0 6 0 4 2 9 5 3 9 3 9 2 3 0 5
7 8 9 9 7 8 8 7 1 7 3 7 6 5 5 9 6 2 4 6 9 7 7 8 6 9 4 9 3 3 9 7 3 4 9 5 1 6 5 5 0 4 8 1 2 5 0 8 9 4
1 1 5 2 3 5 9 1 8 0 4 2 2 7 1 0 9 2 7 7 6 0 7 5 7 0 5 3 8 7 9 2 6 4 6 7 3 0 2 5 7 6 9 7 2 6 5 8 7 3
6 4 3 4 7 0 0 6 9 6 2 7 2 6 4 1 1 5 8 0 2 0 1 8 4 8 6 8 2 0 6 6 4 0 3 1 3 5 3 2 3 0 9 2 9 7 8 0 5 7
8 0 3 6 3 2 7 1 8 4 7 0 4 7 6 7 2 0 5 0 7 3 7 7 8 9 5 1 5 5 3 1 5 5 0 5 9 0 6 6 9 1 0 2 5 3 3 0 7 1
8 9 6 4 4 3 8 4 7 4 4 7 9 4 6 1 7 0 5 9 5 0 1 2 9 0 3 2 2 9 6 8 2 1 9 1 6 4 6 1 5 8 4 2 6 8 4 2 0 0
7 1 6 7 7 9 6 0 6 3 4 0 9 5 8 5 6 0 5 6 4 6 6 4 3 0 7 1 9 5 2 6 1 5 0 9 0 3 3 9 8 3 3 2 2 6 6 2 9 4
6 6 0 5 3 3 7 2 7 7 7 9 4 0 0 5 4 7 0 1 2 9 7 0 6 2 0 2 2 5 4 9 9 1 4 5 6 6 6 6 4 6 6 5 2 0 5 4 1 1
3 2 5 2 6 1 4 3 7 6 4 6 6 8 9 0 0 3 1 6 4 7 4 6 8 8 2 6 3 7 5 1 0 2 5 7 1 2 1 9 7 3 5 6 2 3 3 5 4 8
2 2 9 2 9 3 9 3 4 5 1 5 5 8 5 4 4 8 8 2 1 1 3 4 4 1 4 3 7 5 9 4 8 9 3 4 5 2 3 9 7 1 1 5 6 3 7 4 0 9
9 2 7 8 0 8 4 7 7 9 3 9 7 2 6 4 8 2 5 9 9 6 7 0 2 5 6 8 5 2 9 5 8 0 3 7 7 4 8 9 3 5 0 0 9 2 9 7 7 3
2 5 3 0 0 0 0 9 7 7 2 6 4 9 3 9 5 5 7 6 8 9 8 9 6 0 2 0 2 2 7 1 6 7 9 8 0 2 4 3 0 8 6 0 3 0 1 5 5 9
5 6 7 8 7 0 0 6 9 9 2 1 2 5 7 0 2 0 7 0 5 9 0 7 2 5 5 8 4 8 5 1 2 6 3 6 5 3 2 9 7 9 0 5 3 2 4 2 1 4
9 4 2 5 6 5 3 1 5 9 6 4 3 3 8 3 0 2 3 3 1 6 7 7 2 3 8 4 0 7 1 3 6 2 7 5 1 5 3 7 5 5 9 6 3 4 0 6 0 9
5 1 9 5 4 0 5 5 9 4 7 7 4 2 3 5 3 0 2 1 4 0 9 3 4 9 8 6 8 2 4 8 4 8 0 7 6 7 9 5 3 9 9 1 3 5 6 2 0 7
5 9 2 5 6 1 0 4 7 1 6 4 1 5 3 2 1 9 3 2 0 6 8 6 3 3 7 5 6 4 2 4 0 0 9 6 1 6 5 0 2 3 7 1 9 8 3 1 4 8
6 8 8 2 9 5 4 3 5 4 5 2 9 5 5 2 0 5 3 3 6 2 1 9 4 7 1 3 1 0 2 5 7 8 5 5 7 5 8 9 7 3 0 8 8 9 1 8 8 1
3 2 5 1 4 0 3 0 1 8 2 5 0 1 0 7 6 0 3 4 1 5 1 9 8 0 3 0 6 3 2 0 9 8 3 3 8 0 9 3 3 7 1 8 2 3 9 9 1 1
5 6 5 2 9 8 7 1 4 1 3 9 3 2 6 3 8 5 2 1 2 6 4 7 9 7 1 9 0 2 2 8 4 6 4 3 5 9 4 5 9 9 6 1 3 6 5 2 3 1
6 5 9 8 2 2 3 1 8 1 7 8 0 8 5 4 0 9 1 3 0 2 1 6 7 4 5 1 7 8 2 6 4 3 8 9 9 7 1 0 6 3 4 5 9 0 1 6 1 8
```