

ESTATÍSTICA BÁSICA

USANDO O R

Augusto Filho

augustofilho@cpdee.ufmg.br

<http://geocities.yahoo.com.br/augustofilho>

Belo Horizonte - MG

Versão 2.0 - 05-07-06

Sumário

Prefácio	v
1 Tabelas e Gráficos	1
1.1 Coleta de Dados	1
1.2 Crítica dos Dados	3
1.3 Apresentação dos Dados	3
1.4 Tabelas	3
1.5 Gráficos	4
1.6 Distribuição de Frequências	5
1.7 Interporlação Linear da Ogiva	5
1.7.1 Introdução	5
1.7.2 O Problema	6
1.7.3 Outro exemplo	8
1.7.4 Variação Importante	10
1.7.5 Exemplos de Interpolação da Ogiva	12
1.8 Exercícios	14
1.9 Referências Bibliográficas	18
2 Medidas de Tendência Central	19
2.1 Média Aritmética - Dados Não Agrupados.	19
2.2 Média Aritmética - Dados Agrupados.	20
2.3 Mediana - \tilde{X}	27
2.3.1 Mediana - Tabela Pontual	30
2.3.2 Mediana - Tabela Intervalar	31
2.4 Moda - Valores que mais se repetem.	34
2.4.1 Resumo e Propriedades	35
2.5 Exercícios	35
3 Principais Separatrizes	47
3.1 Determinação do Quartil	48
3.2 Calculando o primeiro quartil - Q_1	48
3.3 O segundo e o terceiro quartil	51

3.4	O primeiro decil - D_1	54
3.5	Calculando os outros decis - D_2 a D_9	56
3.6	Calculando os percentis	56
3.7	Exercícios	57
4	Medidas de Variabilidade	59
4.1	Amplitude e Desvio Médio	59
4.2	Variância e Desvio Padrão	60
4.2.1	Desvio padrão para dados não ordenados.	60
4.2.2	Desvio padrão - Dados Agrupados	62
4.3	Coefficiente de Variação	63
4.4	Medidas de Assimetria	64
4.4.1	Coefficiente de Assimetria de Pearson	65
4.5	Exercícios	65
4.6	Referências Bibliográficas	66
5	Introdução à Probabilidade Básica	67
5.1	Espaço Amostral - introdução	67
5.2	Frequência Relativa	68
5.3	Espaço Amostral	68
5.3.1	Operações com eventos aleatórios	69
5.4	Probabilidade	71
5.4.1	Função de Probabilidade	71
5.4.2	Teoremas	71
5.4.3	Espaços amostrais equiprováveis	71
5.4.4	Probabilidade Condicional	78
5.5	Exercícios	80
5.6	Teorema de Bayes	83
5.7	Referências Bibliográficas	85

Prefácio

Esta apostila é o resultado de alguns anos de trabalho ministrando estatística em Instituições Federais e Particulares de ensino superior. Sem sombra de dúvida, foi escrita, a partir das experiências e descobertas vividas em sala de aula.

É uma obra voltada, àqueles que procuram aprender a Estatística Básica, no intuito de aplicar tal conhecimento à resoluções de questões do dia-a-dia. Para tanto, a linguagem empregada foi tão próxima da coloquial, quanto próximo foram os exemplos e comentários, aqui presentes, daqueles utilizados em minhas aulas.

No endereço (<http://geocities.yahoo.com.br/augustofilho>) existe toda a base de dados para a resolução dos exercícios computacionais desta apostila.

Esta apostila não se propõe a ser uma apostila técnica, tampouco apresenta a complexidade normalmente inerente a "trabalhos científicos". Revela, antes, a feição de um manual, um material auxiliar, cuja meta é, facilitar ao estudante a compreensão da matéria, por intermédio de exemplos, de analogias e de elementos capazes de despertar seu entendimento.

Coloco-me a inteira disposição para acolher quaisquer sugestões e críticas que possam vir a possibilitar o aperfeiçoamento desta obra. A todos o meu muito obrigado.

Atenciosamente.

Augusto Filho

augustofilho@cpdee.ufmg.br

Capítulo 1

Tabelas e Gráficos

A finalidade das aulas iniciais é apresentar os principais conceitos sobre o levantamento de dados; destacar as técnicas de apresentação, por meio de tabelas e gráficos; oferecer as medidas estatísticas próprias para análises e as técnicas usadas para a interpretação dos dados numéricos, em resumo, minha preocupação é explorar cada uma das partes que compõem a seguinte definição: "A estatística é coleta, apresentação, análise e interpretação de dados numéricos". Particularmente, neste capítulo, destacaremos a Estatística descritiva que pode ser resumida no diagrama:

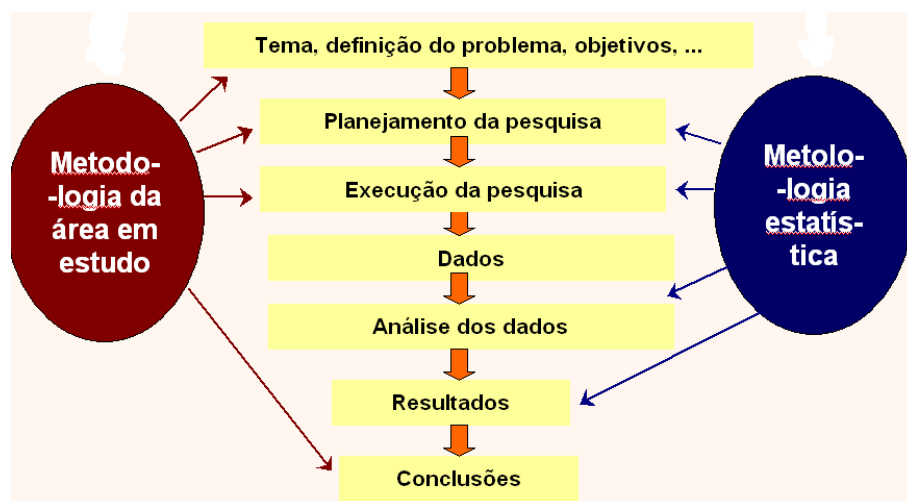


Figura 1.1: Descrição da Estatística Descritiva

Quanto à Estatística Indutiva ou Inferência Estatística (conclusões e interpretações sobre parâmetros populacionais, a partir de dados amostrais), será tratado nas próximas aulas.

1.1 Coleta de Dados

Após a definição do problema a ser estudado e o estabelecimento do planejamento da pesquisa (forma pela qual os dados serão coletados; cronograma das atividades; custos

envolvidos; exame das informações disponíveis; delineamento da amostra etc.), o passo seguinte é a coleta de dados, que consiste na busca ou compilação das variáveis, componentes do fenômeno a ser estudado. A coleta de dados é direta quando os dados são obtidos na forma originária. Os valores assim compilados são chamados de dados primários, como por exemplo, nascimentos, casamentos e óbitos, registrados no Cartório de registros Civil; opiniões obtidas em pesquisas de opinião pública; vendas registradas em notas fiscais da empresa, etc. O conjunto de informações disponíveis, após a tabulação do questionário ou pesquisa de campo, é denominado de tabela de dados brutos e contém os dados da maneira que foram coletados inicialmente. Por exemplo, imaginemos um questionário elaborado para uma turma de alunos, desta forma, cada uma das características perguntadas aos alunos, tais como o peso, a idade e a altura, entre outras, é denominada de variável. Assim, a variável Altura assume os valores (em metros) 1,60; 1,58;... e a variável Turma assume os valores A ou B. Claramente tais variáveis têm naturezas diferentes no que tange aos possíveis valores que podem assumir. Tal fato deve ser levado em conta nas análises e, para fixar idéias, vamos considerar dois grandes tipos de variáveis: numéricas e não numéricas. As numéricas serão denominadas quantitativas, ao passo que as não numéricas, qualitativas. A variável é qualitativa quando os possíveis valores que assume representam atributos e/ou qualidades. Se tais variáveis têm uma ordenação natural, indicando intensidades crescentes de realização, então elas serão classificadas como qualitativas ordinais. Caso contrário, quando não é possível estabelecer uma ordem natural entre seus valores, elas são classificadas como qualitativas nominais. Variáveis tais como Turma (A ou B), Sexo (feminino ou masculino) e Fuma (Sim, não) são variáveis qualitativas nominais. Por outro lado, variáveis como Tamanho (pequeno, médio ou grande), Classe Social (baixa, média ou alta) são variáveis qualitativas ordinais. Variáveis quantitativas, isto é, variáveis de natureza numérica, podem ser subdivididas em discretas e contínuas. A grosso modo, variáveis quantitativas discretas podem ser vistas como resultantes de contagens, assumindo assim, em geral, valores inteiros. De uma maneira mais formal, o conjunto dos valores assumidos é finito ou enumerável. Já as variáveis quantitativas contínuas assumem valores em intervalos dos números reais e, geralmente, são provenientes de uma mensuração.

Resumimos a classificação das variáveis no esquema apresentado a seguir:

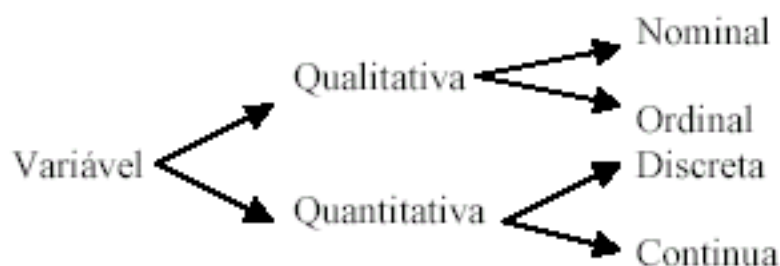


Figura 1.2: Classificação das Variáveis

Vale ressaltar que, em muitas situações práticas, a classificação depende de certas particularidades. Por exemplo, a variável idade, medida em números de anos, pode ser vista como discreta, entretanto, se levarmos em conta os dias, não é absurdo falar que a idade é 2,5 ou 2,85 anos, dando assim respaldo para classifica-la como contínua.

1.2 Crítica dos Dados

Objetivando a eliminação de erros capazes de provocar futuros enganos de apresentação e análise, procede-se a uma revisão crítica dos dados, suprimindo os valores estranhos ao levantamento.

1.3 Apresentação dos Dados

Após a crítica, convém organizarmos os dados de maneira prática e racional, para melhor entendimento do fenômeno que se está estudando. A organização dos dados denomina-se Série Estatística. Sua apresentação pode ocorrer por meio de tabelas e gráficos.

1.4 Tabelas

A elaboração de tabelas obedece à Resolução do conselho nacional de Estatística. Uma tabela e mesmo um gráfico deve apresentar o cabeçalho; o corpo; e o rodapé.

O cabeçalho deve conter o suficiente para que sejam respondidas as seguintes questões:

- O quê? (referente ao fato);
- Onde? (relativo ao lugar);
- Quando (correspondente à época).

O corpo é reservado para as observações pertinentes, bem como a identificação da fonte dos dados. Conforme o critério de agrupamento, as séries classificam-se em:

1. Série Cronológica: É a série estatística em que os dados são observados segundo a época de ocorrência.

Vendas da Companhia C&A 1980-1982.

2. Distribuição de frequências. É a série estatística em que os dados são agrupados com suas respectivas frequências absolutas.

Exemplos:

Ano	Venda
1980	2181
1981	3949
1982	5642
Total	48404

Tabela 1.1: Tabela Pontual

Números de Acidentes por dia na Rodovia X em Belo Horizonte em 2001.

N. de Acidentes por dia na Av. Antônio Carlos	Frequências Dias
0	10
1	7
2	4
3	5
4	3
5	2

Fonte: DNER.

1.5 Gráficos

A representação gráfica das series estatísticas tem por finalidade dar uma idéia, a mais imediata possível, dos resultados obtidos, permitindo chegar-se a conclusões sobre a evolução do fenômeno ou sobre como se relacionam os valores da serie. Não há apenas uma maneira de representar graficamente uma serie estatística. A escolha do gráfico mais apropriado ficará a critério do analista. Contudo, os elementos simplicidade, clareza e veracidade devem ser considerados quando da elaboração de um gráfico. Eis os principais tipos de gráficos.

1) Gráfico em Colunas.

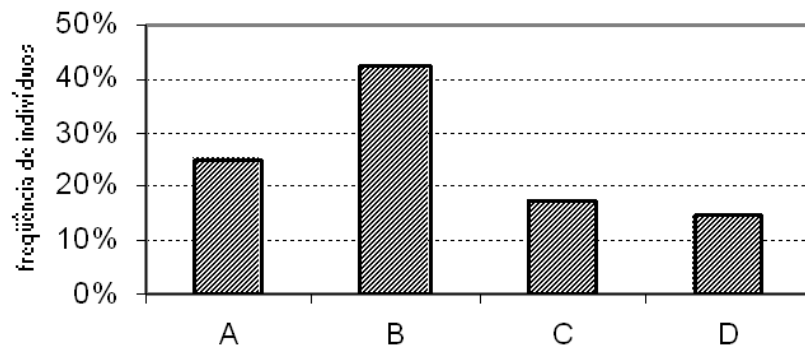


Figura 1.3: Gráfico em Colunas

2) Gráfico em Barras. É semelhante ao gráfico em colunas, porém os retângulos são dispostos horizontalmente. Eis uma configuração:

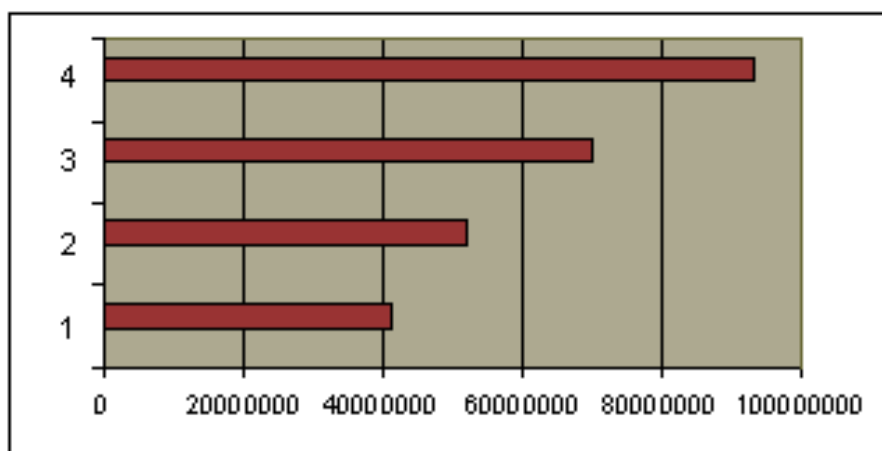


Figura 1.4: Gráfico em Barras

1.6 Distribuição de Frequências

- **POPULAÇÃO:** É um conjunto de indivíduos ou objetos que apresentam pelo menos uma característica em comum. A população pode ser finita ou infinita, dependendo de o número de elementos ser finito ou infinito. Na prática, quando uma população é finita, com um número grande de elementos, considera-se como população infinita.
- **AMOSTRA:** Considerando-se a impossibilidade, na maioria das vezes, do tratamento de todos os elementos da população, retiraremos uma amostra. Para nossos propósitos, admite-se que uma amostra já tenha sido escolhida de conformidade com alguma técnica de amostragem.

1.7 Interporlação Linear da Ogiva

1.7.1 Introdução

O assunto que veremos agora passou a fazer parte das provas de Estatísticas da ESAF já há alguns anos, mais ou menos desde o AFRF de 2001, e desde então não mais deixou de ser cobrado. Trata-se de uma questão fácil, embora o nome do assunto possa assustar um pouco.

Começaremos com um exemplo bem simples. Vejamos a *distribuição de frequência abaixo*:

Classes	Frequência
0 † 10	5
10 † 20	8
20 † 30	13
30 † 40	11
40 † 50	7
50 † 60	3

Se a questão da prova perguntasse, por exemplo, "*quantos elementos deste conjunto têm valor abaixo de 30*", como responderíamos? Ora, observando as classes desta distribuição, vemos facilmente que "participam desta resposta" os elementos das três primeiras classes. Desta forma, teríamos 5 elementos na primeira classes (abaixo de 10), mais 8 elementos na segunda classe (de 10 a 20) e finalmente 13 elementos na terceira classe (valores de 20 a 30). Somando tudo, nossa resposta seria 26. Sem problemas!

Mais uma vez: a pergunta agora é "*quantos elementos deste conjunto têm valor acima de 40?*" Também sem grandes dificuldades, percebemos que "participam desta resposta" os elementos das duas últimas classes, ou seja, elementos com valor de 40 a 50 (quinta classe) e de 50 a 60 (sexta classe).

Logo, como temos 7 elementos na penúltima, e 3 elementos na última classe, nossa resposta seria a soma, ou seja, 10 elementos. Até aqui, tudo muito bem.

1.7.2 O Problema

A nova pergunta é: *quantos elementos deste mesmo conjunto têm valor menor ou igual a 28?* Observando os limites das classes apresentadas, percebemos que 28 não é nem limite superior, nem inferior de qualquer destas classes. Na verdade, o valor 28 encontra-se dentro da terceira classe. Para completar o enunciado, a questão vai pedir ainda que determinemos esta resposta, utilizando-nos da *interpolação linear da ogiva*.

Embora ainda nem tenhamos falado de ogiva (ou de outros gráficos estatísticos), teremos já total condição de resolver este problema, fazendo uso de uma regra de três simples, a mais fácil possível.

Percebamos que é fácil deduzir que a primeira e a segunda classes participarão da resposta *integralmente*, porém a terceira classe (20 † 30) participará apenas *parcialmente* do resultado.

Ou seja

Classes	Frequência	
0 † 10	5	participa intergalmente da resposta !
10 † 20	8	participa integralmente da resposta !
20 † 30	13	participa parcialmente da resposta
30 † 40	11	
40 † 50	7	
50 † 60	3	

O segredo, então, é trabalharmos com esta classe que participa apenas *parcialmente* da resposta. Daí, faremos:

→ a terceira classe tem amplitude **h=10** e frequência simples, **Freq = 13**.

Assim, a primeira linha da regra de três está formada:

10 → 13 (dez está para treze)

Traduzindo: nesta amplitude de 10, temos 13 elementos.

Para o complemento da regra de três, pensaremos o seguinte: a questão quer saber "menor ou igual a 28". Ora, menor ou igual a 28, nesta classe, nós teremos desde o limite inferior da classe (20) até o próprio 28. Ou seja, a *amplitude desejada* para esta classe, neste momento, será apenas esta diferença: $(28 - 20) = 8$. Daí, a segunda linha da regra de três será:

8 → X (oito está para X)

Ou seja, nesta amplitude de apenas 8, quantos elementos teremos? ($X = ?$).

Agora, nossa regra de três completa será:

10 → 13
8 → X

Multiplicamos cruzando e chegaremos a:

$$X = (8 \cdot 13) / 10 \implies E : X = 104 / 10 \implies \text{Daí: } \mathbf{X=10,4}.$$

Observemos que este valor encontrado (10,4) é apenas a participação da terceira classe em nossa resposta. o valor que de fato procuramos reunirá também as frequência das duas primeiras classes deste conjunto, as quais, como vimos, participam integralmente do resultado.

Daí, teremos:

- * primeira classe: $(0 \vdash 10) \rightarrow 5$ elementos ($Freq = 5$).
- * segunda classe: $(10 \vdash 20) \rightarrow 8$ elementos ($Freq = 8$).
- * terceira classe: $(20 \vdash 30) \rightarrow 10,4$ elementos ($X = 10,4$).

Total de elementos: 23,4 elementos \rightarrow **Resposta!**

Obviamente que este resultado reflete apenas uma aproximação, ou seja, uma estimativa, uma vez que, quando trabalhamos com a *distribuição de frequência*, teremos efetivamente uma perda de informação. Mas não nos preocupemos: embora essa resposta seja o reflexo de uma aproximação, ela é a resposta correta.

1.7.3 Outro exemplo

Uma nova questão agora pergunta, para aquela mesma distribuição de frequências: *quantos elementos deste conjunto têm valor maior ou igual a 34?*

Aqui está novamente o nosso conjunto:

Classes	Frequência
0 \vdash 10	5
10 \vdash 20	8
20 \vdash 30	13
30 \vdash 40	11
40 \vdash 50	7
50 \vdash 60	3

Observamos que este valor, 34, não é limite inferior ou superior de nenhuma das classes; ao contrário, está dentro da quarta classe.

Constatamos, ainda, pela mera observação, que, se a questão pede elementos com valores acima de 34, esta quarta classe participará da resposta apenas de forma parcial. Enquanto isso, as duas últimas classes participarão integralmente do resultado. Ou seja:

Classes	Frequência	
0 \vdash 10	5	
10 \vdash 20	8	
20 \vdash 30	13	
30 \vdash 40	11	participa parcialmente da resposta!
40 \vdash 50	7	participa integralmente da resposta!
50 \vdash 60	3	participa integralmente da resposta!

Ficou fácil perceber que teremos que trabalhar a regra de três com a quarta classe, para descobrir quantos de seus elementos participarão da resposta.

Para compor a regra de três, inicialmente trabalhamos com a classe inteira. E, nesta quarta classe, temos amplitude $h=10$ e frequência simples $Freq=11$. Portanto, a primeira linha da regra de três será a seguinte:

$$10 \longrightarrow 11 \text{ (dez está para onze)}$$

Ora, para esta mesma quarta classe, maiores ou iguais a 34 serão os elementos 34 a 40. Ou seja, a *amplitude desejada* na resposta para essa classe será apenas esta diferença: $40 - 34 = 6$. Daí, a segunda linha da regra de três será:

$$6 \longrightarrow x \text{ (seis está para X)}$$

Ou seja, na amplitude de 6, teremos X elementos.

Portanto, nossa regra de três completa será a seguinte:

$10 \longrightarrow 11$ $6 \longrightarrow X$

Resolvendo, teremos: $10X = 6 \cdot 11 \Rightarrow E : X = 66/10 = 6,6$

Ou seja, em relação à quarta classe, participam da resposta apenas 6,6 elementos! Para chegarmos ao resultado da questão, todavia, temos de nos lembrar que as frequências das duas derradeiras classes terão participação integral. Daí, teremos:

\Rightarrow quarta classe: $(30 \vdash 40) \longrightarrow 6,6$ elementos ($X = 6,6$)

\Rightarrow quinta classe: $(40 \vdash 50) \longrightarrow 7$ elementos ($Freq = 7$)

\Rightarrow sexta classe: $(50 \vdash 60) \longrightarrow 3$ elementos ($Freq = 3$)

Total de elementos: 16,6 **elementos** \Rightarrow **Resposta!**

A questão é basicamente isso. Há algumas variações possíveis, como por exemplo, em vez de a questão perguntar "quantos elementos", ela perguntaria *qual o percentual de elementos*, ou seja, em vez de trabalharmos com a frequência absoluta simples (Freq), trabalharíamos com a frequência percentual (%).

Outra variação é aquela em que a questão pergunta "quantos elementos do conjunto têm valor acima de X e abaixo de Y?", de forma que X e Y são valores não-coincidentes com os limites inferiores ou superiores das classes da distribuição. Neste caso, teríamos duas classes participando parcialmente da resposta; logo, teríamos que fazer duas regras de três: uma para a classe em que o X estivesse inserido, outra para a classe a qual pertence o Y.

1.7.4 Variação Importante

Existe, todavia, uma variação desta questão digna de nota. Seria um enunciado do tipo que se segue:

Classes	%
0 † 10	5%
10 † 20	22%
20 † 30	33%
30 † 40	12%
40 † 50	8%

Considerando a distribuição de frequência acima, em que % representa a frequência percentual, determine, *via interpolação linear da ogiva*, qual o elemento deste conjunto que não é superado por 45% das observações.

Temos aí uma coluna com as frequências percentuais, e a questão pergunta, em outras palavras, qual o número que está abaixo de 45% do total de elementos do conjunto.

Vejam: a primeira classe tem 5% dos elementos; a segunda classe tem 22%. Somando estas duas primeiras frequências percentuais, teremos já 27% do total dos elementos. Agora: de 27% para chegarmos a 45%, quanto falta? Obviamente que faltam ainda 18%. É a diferença ($45\% - 27\% = 18\%$).

Seguindo: se precisamos avançar mais 18% a partir da segunda classe (para chegar aos 45% desejados), e a próxima classe, que é a terceira, já tem 33% dos elementos do conjunto, isso significa que a resposta que estamos procurando estará exatamente dentro desta terceira classe.

Senão, vejamos: já tínhamos 27% dos elementos acumulados nas duas primeiras classes. Se somássemos a esses 27% os 33% da terceira classe, passaríamos a 60% dos elementos do conjunto. E o nosso objetivo é chegar aos 45%.

Daí, trabalharemos, formando uma regra de três simples para a terceira classe, cuja frequência percentual participa apenas parcialmente na busca do resultado. De antemão, já sabemos que nossa resposta estará dentro da terceira classe, ou seja, será um valor no intervalo de 20 a 30.

A nossa situação é a seguinte:

Classes	%	
0 † 10	5%	5% acumulados!
10 † 20	22%	27% acumulados!
20 † 30	33%	Faltam 18% para chegarmos aos 45%
30 † 40	12%	
40 † 50	8%	

Assim, faremos nossa regra de três com o seguinte raciocínio: na terceira classe, temos amplitude $h=10$ e frequência percentual ($Freq\% = 33\%$). Logo, a primeira linha da regra de três será:

$$10 \longrightarrow 33\% \text{ (dez está para trinta e três por cento)}$$

Ou seja, em uma amplitude de 10, temos 33% dos elementos do conjunto.

Para construir a segunda linha da regra de três, pensaremos assim: interessam-nos, nesta terceira classe, apenas 18% dos elementos, que serão necessários para acumularmos os 45% desejados. Daí, faremos:

$$X \longrightarrow 18\% \text{ (X está para dezoito por cento)}$$

Ou seja: qual será a amplitude ($X = ?$) desta terceira classe, que abrangerá apenas 18% dos seus elementos?

A regra de três completa é a seguinte:

$10 \longrightarrow 33\%$ $X \longrightarrow 18\%$

Multiplicando em cruz, teremos:

$$X = (18\% \cdot 10) / 33\% \longrightarrow E : \mathbf{X=5,45}$$

Agora o mais importante: como usar esse X encontrado? *Somando-o ao limite inferior da terceira classe.*

Vamos entender: se estivermos no limite inferior da terceira classe ($lin_{inf} = 20$) e somarmos a este a amplitude da classe inteira ($h = 10$), chegaríamos ao limite superior ($lin_{sup} = 30$). Todavia, não nos interessa somar o limite inferior com a amplitude da classe, pois, assim, "avançaríamos", mais 33% dos elementos.

Queremos avançar apenas 18% dos elementos, o que corresponde a uma amplitude de $X = 5,45$, conforme calculamos acima.

Logo, para chegarmos ao resultado solicitado pela questão, faremos:

$$\mathbf{20+5,45=25,45 \longrightarrow Resposta da questão!}$$

A seguir, será apresentado alguns exemplos e aplicações.

1.7.5 Exemplos de Interpolação da Ogiva

Exemplo 1.1. *Em um ensaio para o estudo da distribuição de um atributo financeiro (X), foram examinados 200 itens de natureza contábil do balanço de uma empresa. Esse exercício produziu a tabela de frequência abaixo. A coluna Classes representa intervalos de valores de X em reais e a coluna P representa a frequência relativa acumulada. Não existem observações coincidentes com os extremos das classes.*

Classes	$P\%$
70 - 90	5
90 - 110	15
110 - 130	40
130 - 150	70
150 - 170	85
170 - 190	95
190 - 210	100

Encontre o valor que corresponde à estimativa da frequência relativa de observações de X menores ou iguais a 145.

Solução

Esta questão pede a resposta em valores percentuais, ou seja, ela quer que trabalhem com frequência relativas, mais especificamente com a frequência relativa simples (Freq). Essa constatação foi fácil. Resta agora verificar se a coluna fornecida foi já a *Freq*, ou se foi alguma outra.

Ora, o enunciado foi explícito, afirmando que a coluna P "representa a frequência relativa acumulada". Já aprendemos, neste caso, o que fazer para chegarmos à coluna da **Frequência relativa simples (Freq. Relativa)**.

Classes	$Fac \downarrow$	Freq.
70 - 90	5%	5%
90 - 110	15% (15%-5%=)	10%
110 - 130	40% (40% -15%=)	25%
130 - 150	70% (70% - 40%=)	30%
150 - 170	85% (85% - 70%=)	15%
170 - 190	95% (95% - 85%=)	10%
190 - 210	100% (100%-95%=)	5%

A questão quer saber valores "menores ou iguais a 145". É fácil verificar que este valor (145) está inserido na quarta classe (130 + 150). Logo, trabalharemos a regra de três exatamente aí, tendo em vista que as frequências relativas das três primeiras classes participarão integralmente da resposta.

Ou seja, a situação será a seguinte:

Classes	Freq.	
70 - 90	5%	→ participa integralmente da resposta!
90 - 110	10%	→ participa integralmente da resposta!
110 - 130	25%	→ participa integralmente da resposta!
130 - 150	30%	→ participa parcialmente da resposta!
150 - 170	15%	
170 - 190	10%	
190 - 210	5%	

A primeira parte desta regra de três levará em conta a quarta classe completa. Temos uma amplitude de $h = 20$ e uma frequência relativa de ($Freq_{relat} = 30\%$). Daí:

20 → 30% (**vinte está para trinta por cento**)

Na segunda parte da regra de três, trabalhamos com a classe "quebrada". Ora, menores ou iguais a 145, nesta classe, nós temos de 130 até 145. Logo, para este enunciado, a amplitude aqui desejada será esta diferença: $(145 - 130) = 15$. Daí, teremos:

15 → $X\%$ (**quinze está para $X\%$**)

Nossa regra de três completa ficará assim:

20 → 30%
15 → $X\%$

Resolvendo, ficaremos com:

$$X = (15 \times 30\%) / 20 \rightarrow X = 450\% / 20 \rightarrow X = 22,5\%$$

Logo, este valor encontrado será a parcela de participação da quarta classe na resposta. Contudo, é evidente que as frequências relativas das três primeiras classes também **participarão** do resultado, e de forma integral, como vimos acima.

Assim, teremos:

⇒ primeira classe: $(70 \vdash 90) \rightarrow 5\%$ dos elementos ($Freq = 5\%$)

⇒ segunda classe: $(90 \vdash 110) \rightarrow 10\%$ dos elementos ($Freq = 10\%$)

⇒ terceira classe: $(110 \vdash 130) \rightarrow 25\%$ dos elementos ($Freq = 25\%$)

⇒ quarta classe: $(130 \vdash 150) \rightarrow 22,5\%$ dos elementos ($Freq = 22,5\%$)

Total: **62,5%** dos elementos! ⇒ **Resposta!**

1.8 Exercícios

Exercício 1. Dada a amostra: 3, 4, 4, 5, 7, 6, 6, 7, 7, 4, 5, 5, 6, 6, 7, 5, 8, 5, 6, 6, pede-se:

- Construir a distribuição de freqüência.
- Construir o gráfico de freqüências;
- Determinar as freqüências relativas;
- Determinar as freqüências acumuladas
- Qual é a amplitude amostral;
- Qual é a porcentagem de elementos maiores que 5.

Exercício 2. Considere os dados obtidos pelas medidas das alturas de 100 indivíduos (dadas em cm);

151	152	154	155	158	159	159	160	161	161
161	162	163	163	163	164	165	165	165	166
166	166	166	167	167	167	167	167	168	168
168	168	168	168	168	168	168	168	169	169
169	169	169	169	169	170	170	170	170	170
170	170	171	171	171	171	172	172	172	173
173	173	174	174	174	175	176	175	175	176
176	176	176	177	177	177	177	178	178	178
179	179	180	180	180	180	181	181	181	182
182	182	183	184	185	186	187	188	190	190

Calcule:

- amplitude amostral;
- o numero de classes;
- a amplitude das classes;
- os limites das classes;
- as freqüências absolutas das classes;
- as freqüências relativas;
- Os pontos médios das classes;
- A freqüência acumulada;
- O histograma - polígono de freqüência;

j) O gráfico de frequência acumulada.

Exercício 3. As notas de 32 estudantes de uma classe estão descritas a seguir :

6,0 0,0 2,0 6,5 5,0 3,5 4,0 7,0
 8,0 7,0 8,5 6,0 4,5 0,0 6,5 6,0
 2,0 5,0 5,5 5,0 7,0 1,5 5,0 5,0
 4,0 4,5 4,0 1,0 5,5 3,5 2,5 4,5

Determine:

- O rol;
- As distribuições de frequências (variável contínua);
- O maior e o menor graus;
- A amplitude total;
- Qual a porcentagem dos alunos que tiveram nota menor do que 4;
- Qual o limite superior da segunda classe;
- Qual o ponto médio da quarta classe;
- Qual o ponto médio da terceira classe;
- Os gráficos (histograma e gráfico da F_{ac}).

Exercício 4. Os pesos de 40 alunos estão relacionados a seguir:

69 57 72 54 93 68 72 58 64 62
 65 76 60 49 74 59 66 83 70 45
 60 81 71 67 63 64 53 73 81 50
 67 68 53 75 65 58 80 60 63 53

- Construir a tabela da distribuição de frequência;
- Construir os gráficos da distribuição.

Exercício 5. Vinte e uma pacientes de uma clínica médica tiveram o seu nível de potássio no plasma medido. Os resultados foram os seguintes.

Nível de Frequência	Frequência
2,25 - 2,55	1
2,55 - 2,75	3
2,75 - 2,95	2
2,95 - 3,15	4
3,15 - 3,35	5
3,35 - 3,65	6
Total	

- a) Construa o histograma;
- b) Qual a porcentagem de valores que estão acima do nível 3 ?

Exercício 6. *A tabela a seguir apresenta as freqüências relativas de ocorrências de faixas de altura (em cm) para uma amostra de 100 crianças de 12 anos de idade.*

- a) Construa o histograma;
- b) Desejando-se separar os 15% mais altos, qual seria o ponto de corte?
- c) Qual a porcentagem de valores que estão acima da faixa 115 ?

Faixas	Frequência Relativa
100 – 100	0,10
100 – 120	0,25
120 – 130	0,30
130 – 140	0,25
140 – 160	0,10

Exercício 7. *Responda:*

- a) Quais são os limites (inferior e superior) da primeira classe?
- b) Quais são as fronteiras (inferior e superior) da primeira classe?
- c) A amplitude dos intervalos é a mesma para todas as classes da distribuição;
- d) Qual é a amplitude?
- e) Qual é o ponto médio da primeira classe?
- f) Quais as fronteiras da classe de aluguéis na qual foi observado o maior número de apartamentos?
- g) Suponha um aluguel mensal de \$239,50. Identificar os limites inferiores e superior da classe na qual esta observação seria registrada.

Distribuição de freqüência de aluguéis mensais para 200

Aluguel	Frequência
150 – 179	3
180 – 209	8
210 – 239	10
240 – 269	13
270 – 299	33
300 – 329	40
330 – 359	35
360 – 389	30
390 – 419	16
420 – 449	12
Total	200

Exercício 8. Um pesquisador de radio XY aborda 30 transeuntes ao acaso e pergunta-lhes a idade. O resultado é dado pela tabela:

35	26	39	25	39	22
42	40	39	22	21	40
16	32	39	21	28	39
18	37	23	14	27	44
30	32	21	15	26	43

- Resuma as informações sob forma de uma distribuição de frequência;
- Apresente os dados na forma de um histograma;
- Qual a porcentagem de valores que estão acima do valor 33?

Exercício 9. Dada a amostra de 60 rendas (em milhares) de dada região geográfica.

10	7	8	5	4	3	2	9	9	6
3	15	1	13	14	4	3	6	6	8
10	11	12	13	14	2	15	5	4	10
2	1	3	8	10	11	13	14	15	16
8	9	5	3	2	3	3	4	4	4
5	6	7	8	9	1	12	13	14	16

- Agrupar os elementos em classes. Sendo $k = 6$ e $h = 3$.
- Construir o histograma e o polígono de frequência.
- Qual a porcentagem de valores que estão acima do valor 9,5?

Exercício 10. Foi pedido aos alunos de uma classe de 40 alunos que escolhessem um dentre os números 0, 1, 2, 3, 4, 5, 6, 7, 8 e 9. Obteve-se o seguinte resultado:

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

a) montar a distribuição de frequência pontual.

Exercício 11. *Abaixo estão dadas as notas (em créditos) de 50 alunos:*

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

Pede-se:

- a) Determinar a amplitude total da amostra;
- b) Numero de classes pela fórmula de Sturges. Dado $\log 50 = 1,7$;
- c) Amplitude das classes;
- d) Quais as classes (inicie pelo 30);
- e) Frequências absolutas das classes;
- f) Frequências relativas;
- g) Pontos médios das classes;
- h) Frequência acumulada;
- i) Histograma.

1.9 Referências Bibliográficas

1. CARVALHO, Sérgio. Estatística Básica - Teoria e 150 questões. Editora IMPETUS, 2004
2. FURTADO, Daniel Ferreira. Estatística Básica. Editora UFLA. 2005
3. LEVINE, David M. Estatística: Teoria e Aplicações usando o microsoft excel em português. Editora LTC, 2003.
4. MAGALHÃES, Marcos Nascimento. Noções de Probabilidade e Estatística / São Paulo, 4ª edição, Editora USP, 2002

Capítulo 2

Medidas de Tendência Central

Vimos nas aulas anteriores a sintetização dos dados sob a forma de tabelas, gráficos e distribuições de frequências. Agora, vamos aprender o cálculo de medidas que possibilitem representar um conjunto de dados relativos à observação de determinado fenômeno de forma resumida. São as medidas de posição. Tais medidas orientam-nos quanto à posição da distribuição no eixo x (eixo dos números reais), possibilitam que comparemos séries de dados entre si pelo confronto desses números. São chamadas medidas da tendência central, pois representam os fenômenos pelo seus valores médios, em torno dos quais tendem a concentrar-se os dados.

2.1 Média Aritmética - Dados Não Agrupados.

Sejam x_1, x_2, \dots, x_n portanto, “n” valores da variável X . A média aritmética simples de X representado por \bar{x} é definida por:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}, \text{ ou simplesmente } \bar{x} = \frac{\sum x}{n}, \text{ em que } n \text{ é o número de elementos do conjunto.}$$

Exemplo 2.1. - *Determinar a média aritmética simples dos valores: 1, 3, 5, 7*

Desta forma, temos: $\bar{x} = \frac{\sum x_i}{n} = \frac{1+3+5+7}{4} = \frac{16}{4} = 4,0$. Desta forma $\bar{x} = 4,0$.

Utilizando-se o Programa R, que veremos ao longo desta apostila, o cálculo utilizado seria:

```
> a<-scan()
1: 1
2: 3
3: 5
4: 7
5:
Read 4 items
> mean(a)
[1] 4
```

2.2 Média Aritmética - Dados Agrupados.

Quando os dados estiverem agrupados numa distribuição de frequência usaremos a média aritmética dos valores x_1, x_2, \dots, x_n ponderados pelas respectivas frequências absolutas: F_1, F_2, \dots, F_n . Assim:

$$\bar{x} = \frac{\sum x_i \cdot F_i}{n}, \text{ onde } x_i \text{ é o ponto médio e o } F_i \text{ é a frequência simples.}$$

Exemplo 2.2.

a) Dada a seguinte distribuição

Renda familiar em milhares de reais.			
Classes	F_i (Freq)	x_i (Ponto Médio)	$x_i \cdot F_i$
2 - 4	5	3	15
4 - 6	10	5	50
6 - 8	14	7	98
8 - 10	8	9	72
10 - 12	3	11	33
<i>Total</i>	40	—	268

Logo, a média para valores agrupados em uma tabela de distribuição de frequência, é igual a:

$$\bar{x} = \frac{\sum x_i F_i}{n} = \frac{268}{40} = 6,7$$

Conclusão: Como a renda familiar foi dada em milhares de reais, podemos afirmar que a renda média desse grupo de 40 famílias é de R\$6.700,00.

O mesmo cálculo utilizando-se o Programa R, seria dado por:

```
> classes <- seq(3, 11, by = 2)
> Freq <- c(5,10,14,8,3)
> dados <- rep(semanas, Freq)
> mean(dados)
[1] 6.7
```

b) Para uma tabela com valores discretos, temos:

x_i	F_i
1	1
2	3
3	5
4	1

Resolução:

Encontraremos a média da seguinte forma:

x_i	F_i	$x_i F_i$
1	1	1
2	3	6
3	5	15
4	1	4
Total	10	26

Logo, temos:

$$\bar{x} = \frac{\sum x_i F_i}{n} = \frac{26}{10} = 2,6$$

Desenvolvendo este cálculo no Programa R, teríamos os seguintes passos:

```
> xi <- c(1,2,3,4)
> Freq <- c(1,3,5,1)
> mean(rep(xi, Freq))
[1] 2.6
```

Exemplo 2.3. *Quer se estudar o número de erros de impressão de um livro. Para isso escolheu-se uma amostra de 50 páginas, encontrando-se o número de erros por páginas da tabela abaixo.*

a) Qual o número médio de erros por página?

Resolução:

Erros	Frequência
0	25
1	20
2	3
3	1
4	1

Aqui utilizaremos o Programa R como se fosse apenas uma calculadora.

```
> media <- (0 * 25 + 1 * 20 + 2 * 3 + 3 * 1 + 4 * 1)/50
> media
[1] 0.66
```

Logo, o R encontrou o seguinte número médio de erros por página: 0,66.

Exemplo 2.4. *As taxas de juros recebidas por 10 ações durante um certo período foram (medidas em porcentagem) 2,59; 2,64; 2,60; 2,62; 2,57; 2,55; 2,61; 2,50; 2,63; 2,64. Calcule a média das taxas apresentadas.*

Resolução:

Utilizou-se o programa R para encontrar a média dos dados acima.

```
> a <- scan()
```

```
> a
```

```
[1] 2.59 2.64 2.60 2.62 2.57 2.55 2.61 2.50 2.63 2.64
```

```
> mean(a)
```

```
[1] 2.595
```

Logo, o resultado obtido foi 2.595, como encontrado anteriormente. Portanto, a taxa de juros média recebidas por 10 ações durante um certo período foi de 2.595.

Exemplo 2.5. *Para facilitar um projeto de ampliação da rede de esgoto de uma certa região de uma cidade, as autoridades tomaram uma amostra de tamanho 50 dos 270 quarteirões que compõem a região, e foram encontrados os seguintes números de casa por quarteirão.*

2	2	3	10	13	14	15	15	16	16
18	18	20	21	22	22	23	24	25	25
26	27	29	29	30	31	36	42	44	45
45	46	48	52	58	59	61	61	61	65
66	66	68	75	78	80	89	90	92	97

- a) Use cinco intervalos e construa um histograma;
- b) Qual a porcentagem de valores que estão acima do valor 83?
- c) Qual o ponto de corte para 20% dos maiores valores?
- d) Calcule a média para os dados agrupados em uma tabela;
- e) Calcule a média para os dados não agrupados em uma tabela e compare o resultado encontrado com a letra "d";

Exemplo 2.6. *Determine a média aritmética das seguintes séries:*

a) 3, 4, 1, 3, 6, 5, 6

b) 7, 8, 8, 10, 12

c) 3, 2; 4; 0, 75; 5; 2, 13; 4, 75

Exemplo 2.7. A média mínima para aprovação em determinada disciplina é 5,0. Se um estudante obtém as notas 7,5; 8,0; 3,5; 6,0; 2,5; 2,0; 5,5; 4,0 nos trabalhos mensais da disciplina em questão, pergunta-se ele foi ou não aprovado.

Exemplo 2.8. A tabela dada a seguir apresenta uma parte dos resultados de uma pesquisa realizada por alunos do curso de Especialização em centro cirúrgico e centro de material, no hospital Universitário de Belo Horizonte, no ano de 1996, com o objetivo de traçar o perfil dos partos ocorridos. Estes dados foram apresentados em relatório final de pesquisa apresentado na disciplina de Estatística.

Número de Semanas	Freq.	%
26 – 28	1	1
28 – 30	1	1
30 – 32	3	4
32 – 34	8	10
34 – 36	3	4
36 – 38	14	18
38 – 40	28	36
40 – 42	13	17
42 – 44	7	9
Total	78	100

Identifique qual é o número médio de semanas de gestação?

Resolução:

Desenvolveremos este item utilizando o R. Note que precisamos encontrar o ponto médio (x_i) e multiplicarmos pela Frequência simples. Desta forma, temos:

```
> semanas <- seq(27, 43, by = 2)
> Freq <- c(1, 1, 3, 8, 3, 14, 28, 13, 7)
> dados <- rep(semanas, Freq)
> mean(dados)
```

```
[1] 37.97436
```

Logo, o número médio de semanas de gestações foi de aproximadamente 38 semanas.

Exemplo 2.9. Calcule para cada uma das distribuições abaixo sua respectiva média.

a)	x_i	F_i
	3	2
	4	5
	7	8
	8	4
12	3	

b)	x_i	F_i
	10	5
	11	8
	12	10
	13	6

Resolução:

Utilizando o R para encontrarmos o resultado, para a tabela pontual, temos:

```
> xi <- c(3, 4, 7, 8, 12)
> Freq <- c(2, 5, 8, 4, 3)
> mean(rep(xi, Freq))
```

```
[1] 6.818182
```

E o mesmo procedimento, para a letra (b).

```
> xi <- c(10, 11, 12, 13)
> Freq <- c(5, 8, 10, 6)
> mean(rep(xi, Freq))
```

```
[1] 11.58621
```

Logo, as médias são respectivamente: 6,81 e 11,58.

Exercício 12. Dadas as estaturas de 140 alunos, conseguiu-se a distribuição abaixo. Calcular a média.

Estatura (cm)	N.de alunos
145 – 150	2
150 – 155	10
155 – 160	27
160 – 165	38
165 – 170	27
170 – 175	21
175 – 180	8
180 – 185	7

Resolução:

Utilizaremos novamente o aplicativo R para encontrarmos a média da tabela intervalar acima.

```
> Estatura <- seq(147.5, 182.5, by = 5)
> n.alunos <- c(2, 10, 27, 38, 27, 21, 8, 7)
> dados <- rep(Estatura, n.alunos)
> mean(dados)
```

```
[1] 164.9286
```

Logo a estatura média deste grupo de alunos é de 164,92 cm.

Exemplo 2.10. *Abaixo temos a distribuição dos aluguéis de 65 casas. Determine sua média.*

Aluguel	Frequência
1,5 - 3,5	12
3,5 - 5,5	18
5,5 - 7,5	20
7,5 - 9,5	10
9,5 - 11,5	5

Resolução:

Utilizando o R, para encontrarmos a média, deveremos trabalhar com o ponto médio de cada classe:

```
> aluguel <- seq(2.5, 10.5, by = 2)
> freq <- c(12, 18, 20, 10, 5)
> mean(rep(aluguel, freq))
```

```
[1] 5.823077
```

Logo, a média para a tabela intervalar acima é de 5,82.

Exemplo 2.11. *Dados os seguintes números:*

```
1 3 5 7 9 2 4 6 8 10 15 20 25 0 1
2 3 4 5 6 7 8 9 9 8 7 8 6 5 4
3 2 1 0 10 15 20 25 12 11 8 6 4 2 1
3 5 7 9 11
```

a) Construa a distribuição de frequência para os dados acima (intervalar);

Resolução:

```
> dados <- scan()
> range(dados)
```

```
[1] 0 25
> nclass.Sturges(dados)

[1] 7
> dadostb <- table(cut(dados, seq(-0.5, 28, l = 8)))
> dadostb

(-0.5,3.57] (3.57,7.64] (7.64,11.7] (11.7,15.8] (15.8,19.9]
(19.9,23.9]
          14          16          13           3           0           2
(23.9,28]
          2
```

b) Determine sua média.

Resolução:

O cálculo para a média da tabela obtida em (a) é dado a seguir:

```
> dados.me <- mean(dados, na.rm = T)
> dados.me

[1] 7.24
```

Logo, a média encontrada para o conjunto de dados acima é 7,24.

Exemplo 2.12. Foi pedido aos alunos de uma classe de 40 alunos que escolhessem um dentre os números 0, 1, 2, 3, 4, 5, 6, 7, 8, e 9. Obteve-se o seguinte resultado:

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

a) Montar a distribuição de frequência (pontual);

Resolução:

```
> a <- scan()
> table(a)

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


Logo, a tabela acima mostra que o valor zero ocorreu duas vezes, o valor 1 ocorreu três vezes, o valor 2 ocorreu quatro vezes, e idéia análoga para o restante.

b) Determinar a média;

```
> a.me <- mean(a, na.rm = T)
> a.me

[1] 4.625
```

Logo, o valor médio para a tabela pontual encontrada em (a) foi de 4,625.

c) Qual foi o número mais repetido? o que ele representa?

O valor três possui a maior frequência (8), isso mostra que ele é o valor com maior ocorrência, caracterizando a moda. Logo, $Mo = 3$.

d) Calcule a mediana.

```
> a.md <- median(a, na.rm = T)
> a.md

[1] 5
```

Logo, a mediana para a tabela pontual encontrada acima é o número 5, mostrando que existe 50% das informações acima e abaixo deste valor.

2.3 Mediana - \tilde{X}

Colocados os valores em ordem crescente, mediana é o elemento que ocupa a posição central.

Vamos considerar, em primeiro lugar, a determinação da mediana para o caso de variável discreta, isto é, para distribuição de frequência simples.

Assim, para a série:

- 5, 7, 8, 10, 14, a mediana será o 8. Indica-se $\tilde{x} = 8$.

Para a série:

- 5, 7, 8, 10, 14, 15, a mediana será o 9, ou seja $\tilde{x} = 9$.

Você já deve ter percebido que precisamos considerar os dois casos: para “n” (numero de elementos da amostra) ímpar o 1º exemplo e para “n” para o 2º exemplo.

Então:

Se n for ímpar, a mediana será o elemento central (de ordem $\frac{n+1}{2}$). Caso “n” seja par, a mediana será a média entre os elementos centrais (de ordem $\frac{n}{2}$ e $\frac{n}{2} + 1$).

Exemplo 2.13. Para cada série, determine a mediana:

a) 1, 3, 3, 4, 5, 6, 6;

Resolução:

- O primeiro passo é ordenarmos os valores de forma crescente ou decrescente e verificarmos se o número de informações (n) é par ou ímpar.

Ordenados: 1, 3, 3, 4, 5, 6, 6

No caso acima, temos “ $n=7$ ”, logo “ $n=\text{ímpar}$ ”.

Desta forma, temos:

$$\tilde{x} = \frac{n+1}{2}$$

Aplicando a fórmula dada, obtemos os seguintes valores:

$$\tilde{x} = \frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4a.$$

É importante destacar que o cálculo feito acima não é a mediana e SIM a localização que a mediana ocupa.

$$\tilde{x} = \frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4a = 4.$$

Logo a mediana (\tilde{x}) é o quarto elemento 4^a , portanto procuramos o número 4 que ocupa a quarta posição nos valores dados.

- Utilizando o R, temos:

```
> a <- scan()
> median(a)

[1] 4
```

Logo, o elemento que corta 50% das informações é o valor 4. O mesmo procedimento para as outras letras.

b) 1, 3, 3, 4, 6, 8, 8, 9;

Resolução:

```
> b <- scan()
> median(b)

[1] 5
```

A mediana é o número 5.

c) 12, 7, 10, 8, 8;

Resolução:

```
> c <- scan()
```

```
> median(c)
```

```
[1] 8
```

A mediana é o número 8, ou seja, 8 corta exatamente 50% das informações.

d) 82, 86, 88, 84, 91, 93;

Resolução:

```
> d <- scan()
```

```
> median(d)
```

```
[1] 87
```

O valor que separa 50% das informações é o número 87.

Exemplo 2.14. *Seja a série: 9, 15, 3, 7, 6, 16, 4, 19, 1, determine a mediana.*

Resolução:

Utilizaremos o R para obtermos a mediana.

```
> a <- scan()
```

```
> median(a)
```

```
[1] 7
```

Exemplo 2.15. *Seja a série: 3, 7, 4, 12, 15, 10, 18, 14, determine a mediana.*

Resolução:

```
> b <- scan()
```

```
> median(b)
```

```
[1] 11
```

Logo a mediana para a série acima é 11.

2.3.1 Mediana - Tabela Pontual

1. Para dados ordenados em uma distribuição pontual.

a)

X_i	F_i	F_{AC}
1	1	1
2	2	4
3	5	9
4	2	11
Total	11	-

→ Contém o 6° elemento

Figura 2.1: Tabela Pontual - “n” - ímpar

Neste caso o número de informações é $n = 11$, onde n é ímpar, logo \tilde{x} será o elemento de ordem $\frac{n+1}{2}$, ou seja, $\frac{11+1}{2} = 6^a$.

Portanto, o 6^a elemento será identificado pela frequência acumulada.

Desta forma, a frequência acumulada será importante para localizarmos a *posição da mediana*, onde, o 6^a elemento é o número 3. Logo $\tilde{x} = 3$.

Exemplo 2.16. Dada uma outra tabela de frequência pontual, temos:

X_i	F_i	F_{AC}
82	5	5
85	10	15
87	15	30
89	8	38
90	4	42
Total	42	-

→ 21° e 22°

Figura 2.2: Tabela Pontual - “n” - par

Logo, temos $n = 42$, n é par, logo \tilde{x} será a média entre os elementos de ordem $\frac{n}{2}$ e $\frac{n}{2} + 1$, ou seja $\frac{42}{2} = 21^a$ e $\frac{42}{2} + 1 = 22^a$. Portanto, como no exemplo anterior, identifica-se os elementos de ordem 21^a e 22^a pela F_{ac}

Assim, temos: 21^a corresponde a 87.

22ª corresponde a 87, logo temos:

$$\tilde{x} = \frac{21^a + 22^a}{2} = \frac{87 + 87}{2} = 87.$$

Portanto, o valor 87 corta exatamente 50% das informações. Em outras palavras, temos 50% das informações acima e abaixo do valor 87.

2.3.2 Mediana - Tabela Intervalar

1ª Passo Calcula-se a ordem $\frac{n}{2}$. Como a variável é contínua, não se preocupe se n é par ou ímpar.

2ª Passo Pela F_{ac} identifica-se a classe que contém a mediana (classe Mediânica);

3ª Passo Utiliza-se a fórmula:

$$\tilde{x} = l_i + \frac{(50\% \text{ de } n - \sum f) h}{F_{\tilde{x}}}$$

em que:

l_i = limite inferior da classe Mediânica;

n = tamanho da amostra ou número de elementos;

$\sum f$ = soma das frequências anteriores à classe Mediânica;

$F_{\tilde{x}}$ = frequência da classe mediânica.

Exemplo 2.17. *Encontre a mediana para a tabela intervalar abaixo:*

Classe	Freq	F_{ac}
35 - 45	5	5
45 - 55	12	17
55 - 65	18	35
65 - 75	14	49
75 - 85	6	55
85 - 95	3	58
Total	58	-

1ª Passo Calcula-se $\frac{n}{2}$. Como $n = 58$, temos $\frac{58}{2} = 29^a$;

2ª Passo Identifica-se a classe Medianica pela F_{ac} . Neste caso, a classe Md

3ª Passo Aplica-se a fórmula:

$$\tilde{x} = l_i + \frac{(50\% \text{ de } n - \sum f) h}{F_{\tilde{x}}}$$

onde:

$$l_i = 55; \quad n = 58; \quad \sum f = 17; \quad h = 10; \quad F_{\tilde{x}} = 18.$$

Logo:

$$\tilde{x} = 55 + \frac{\left(\frac{58}{2} - 17\right) 10}{18} = 61,67$$

Para encontrarmos este resultado utilizando o programa R, deveremos observar o fato de que pode haver uma diferença em relação a resposta original, pois a tabela acima foi composta com os dados originais. Este erro é conhecido como *erro de agrupamento*.

```
> classes <- c(40, 50, 60, 70, 80, 90)
> freq <- c(5, 12, 18, 14, 6, 3)
> median(rep(classes, freq))
```

```
[1] 60
```

O valor da mediana é aproximado, pois aqui existe um erro de agrupamento.

Exemplo 2.18. Para cada distribuição, determine a mediana:

a)

x_i	F_i
2	3
3	5
4	8
5	4
7	2

Resolução:

O cálculo no R é feito de maneira semelhante ao feito acima. No entanto, neste caso, o resultado é exato.

```
> obs <- c(2, 3, 4, 5, 7)
> freq <- c(3, 5, 8, 4, 2)
> median(rep(obs, freq))
```

```
[1] 4
```

Logo, a mediana é 4. Ou seja, existem 50% das informações acima e abaixo deste valor.

b)

x_i	F_i
73	2
75	10
77	15
79	5
81	2

Resolução:

Da mesma forma, de como foi feito acima, poderemos proceder:

```
> obs <- c(73, 75, 77, 79, 81)
> freq <- c(2, 10, 15, 5, 2)
> median(rep(obs, freq))
```

[1] 77

Portanto, temos que o valor 77 corta exatamente 50% das informações.

Exemplo 2.19. *Determine a mediana:*

a)

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

b)

Classes	22 - 25	25 - 28	28 - 31	31 - 34
Freq	18	25	30	20

Exemplo 2.20. *Calcular a média de idades dos pacientes atendidos no Hospital das clínicas.*

Idade de uma amostra de Pacientes atendidos pelo hospital das clínicas da UFMG - 1999

Classes	Freq.
5 - 20	13
20 - 35	15
35 - 50	11
50 - 65	8
65 - 80	10
80 - 95	2
95 - 110	1
Total	60

Fonte: Relatório de pesquisa de alunos

A seguir, temos os cálculos para determinarmos a média, mediana e moda.

Para a média, temos:

```
> Idade <- seq(12.5, 102.5, by = 15)
> freq <- c(13, 15, 11, 8, 10, 2, 1)
> tabela <- rep(Idade, freq)
> mean(tabela)

[1] 41.75
```

Portanto, a idade média dos pacientes atendidos pelo Hospital das Clínicas da UFMG é 41,75 anos.

2.4 Moda - Valores que mais se repetem.

Dentre as principais medidas de posição, destaca-se a Moda. É o valor mais freqüente da distribuição. Para distribuições simples (sem agrupamento em classes), a identificação da Moda é facilitada pela simples observação do elemento que apresenta maior freqüência. Assim, para a distribuição.

x_i	F_i
243	7
245	17
248	23
251	20
307	8

A Moda será 248. Indica-se $Mo = 248$. Notem que esse número é o mais comum nesta distribuição (aparece mais vezes). Para dados Agrupados em classes, temos diversas fórmulas para o cálculo da Moda. Apresentarei o MÉTODO de CZUBER.

1ª Passo Identifica-se a classe Modal (aquela que possuir maior freqüência).

2ª Passo Aplica-se a fórmula:

$$Mo = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} h$$

Onde:

l_i = limite inferior da classe modal;

Δ_1 = diferença entre a freqüência da classe modal e a imediatamente anterior;

Δ_2 = diferença entre a freqüência da classe modal e a imediatamente posterior;

h = amplitude da classe modal.

Exemplo 2.21. Determine a moda para a distribuição.

Classe	Freq
0 - 1	3
1 - 2	10
2 - 3	17
3 - 4	8
4 - 5	5
Total	43

1ª Passo Indica-se a classe Modal. No caso, trata-se da 3ª classe 2| - 3.

2ª Passo Aplica-se a fórmula:

$$Mo = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} h$$

Em que:

$$l_i = 2 \quad \Delta_1 = 17 - 10 = 7 \quad \Delta_2 = 17 - 8 = 9 \quad h = 1$$

Logo:

$$Mo = 2 + \frac{7}{7 + 9} 1 = 2,44$$

Desta forma, existe uma relação muito importante entre a média, a mediana e a moda.

Em uma distribuição simétrica, observa-se que a *média* \simeq *mediana* \simeq *Moda*.

Logo uma distribuição assimétrica positiva observa-se: que a *Média* $>$ *Mediana* $>$ *Moda*. Em uma distribuição com assimetria negativa, observa-se que a *Média* $<$ *Mediana* $<$ *Moda*.

2.4.1 Resumo e Propriedades

2.5 Exercícios

Exercício 13. Os valores a seguir são os pagamentos (em dólares) feitos aos executantes de um concerto de rock. A média é \$8900. Calcule a mediana.

500 600 800 50.000 1.000 500

Resolução:

Utilizaremos o R para resolvermos o problema.

```
> a <- scan()
```

```
> a
```

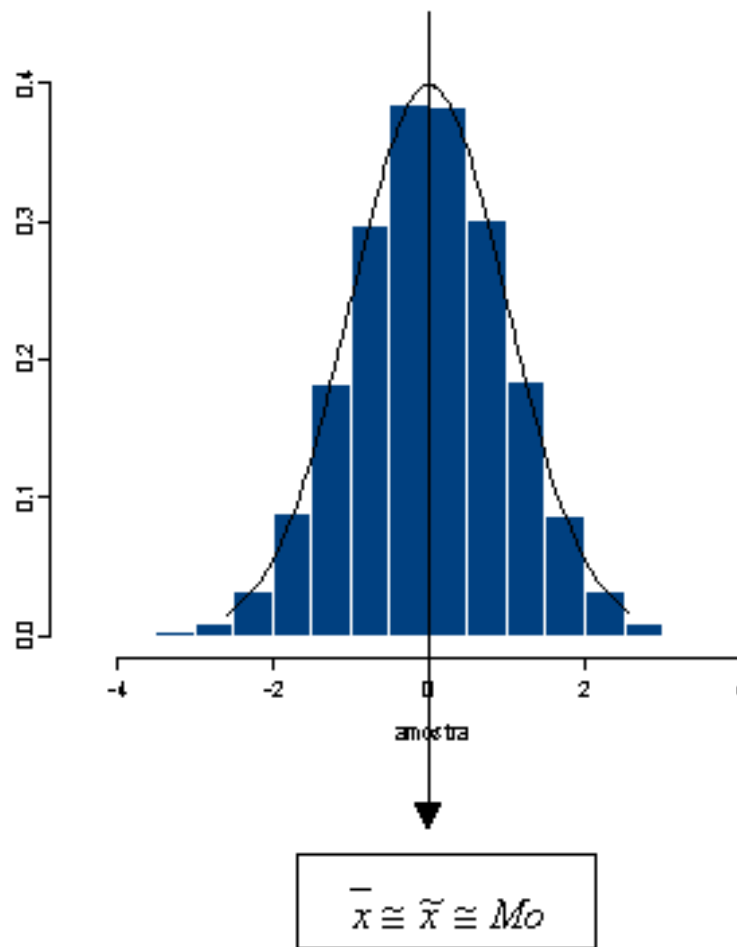


Figura 2.3: Gráfico de uma distribuição Simétrica

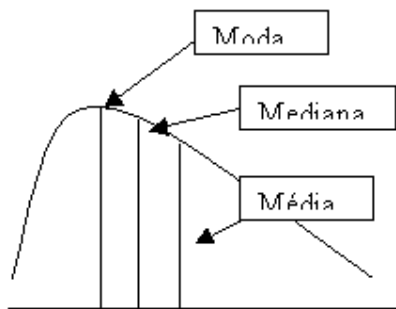


Figura 2.4: Gráfico Assimétrico

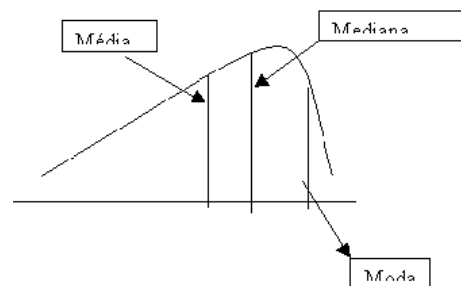


Figura 2.5: Gráfico Assimétrico

```
[1] 500 600 800 50000 1000 500
```

```
> median(a)
```

```
[1] 700
```

Logo, 50% das informações se encontram abaixo e acima do valor 700.

Medida	Definição	Quão Freqüente	Existência	Levam em conta todos os valores?	Afetada pelos valores Extremos?	Vantagens e Desvantagens
Média	$\bar{x} = \frac{\sum x}{n}$	“média” mais familiar	Existe sempre	Sim	Sim	Usada em todo este curso; funciona bem com muitos métodos estatísticos.
Mediana	Valor do meio	Usada comumente	Existe sempre	Não	Não	Costuma ser uma boa escolha se há alguns valores extremos
Moda	Valor mais freqüente	Usada às vezes	Pode não existir; pode haver mais de uma moda.	Não	Não	Apropriada para dados ao nível nominal
Ponto Médio	$\frac{alto + baixo}{2}$	Raramente usada	Existe sempre	Não	Sim	Muito sensível a valores extremos.

Figura 2.6: Resumo e Propriedades

Exercício 14. Para uma amostra de 16 clientes de um pequeno mercado, foram observados os seguintes montantes de vendas, ordenados em ordem crescente:

0,10 0,10 0,25 0,25 0,35 0,40 0,53 0,90
1,25 1,35 2,45 2,71 3,09 3,09 4,00 4,10

Determine:

a) A média;

Utilizando o R, encontramos a média para o conjunto de dados acima.

Resolução:

```
> a <- scan()
> a.tabela <- table(a)
> a.tabela

a
 0.1 0.25 0.35 0.4 0.53 0.9 1.25 1.35 2.45 2.71 3.09 4 4.1
  2   2   1   1   1   1   1   1   1   1   2   1   1

> mean(a)

[1] 1.5575
```

b) A mediana;

Resolução:

```
> median(a)

[1] 1.075
```

Logo a mediana para os dados acima foi de 1,075.

c) A moda para esses valores de vendas.

Resolução:

```
> a.moda <- names(a.tabela)[a.tabela == max(a.tabela)]  
> a.moda  
  
[1] "0.1" "0.25" "3.09"
```

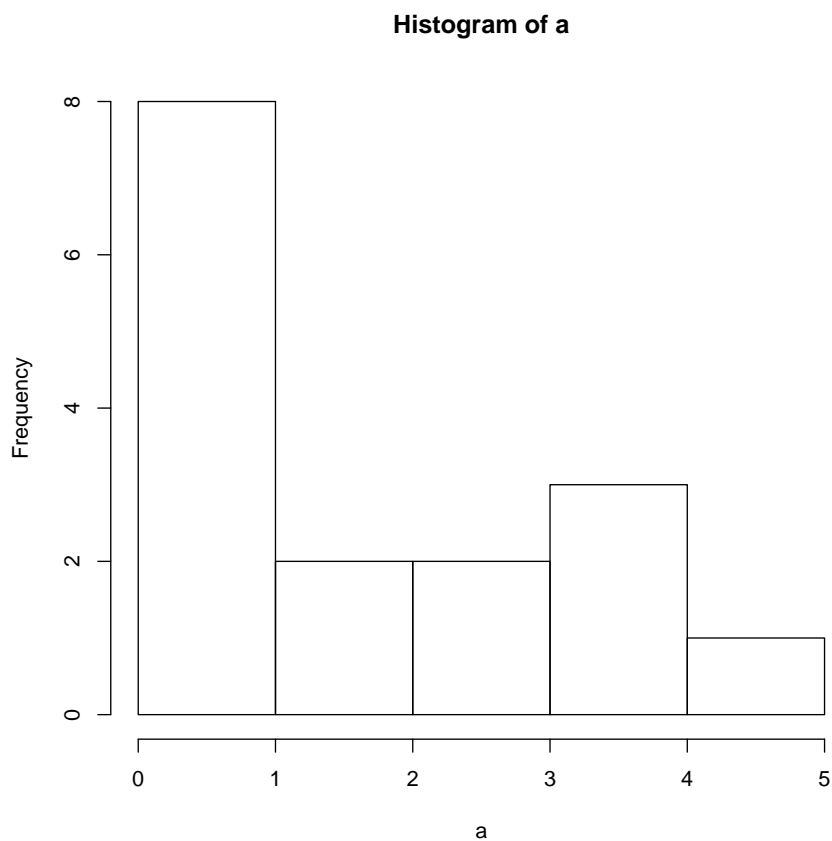
Portanto, temos uma situação trimodal.

Exercício 15. Como você descreveria a distribuição dos dados do problema anterior do ponto de vista da assimetria?

Resolução:

Através da análise do histograma e dos valores de tendência central, é possível concluir que o conjunto de dados anterior tem um comportamento assimétrico, com uma média não representativa.

```
> hist(a)
```



Exercício 16. *Se lhe pedissem uma descrição dos dados do Problema (2) que envolvesse a informação da quantidade "típica" de compra por cliente da amostra, qual medida de tendência central, ou promédio, você utilizaria? Por quê?*

Resolução:

Como a média sofre a influência de valores atípicos, deve-se trabalhar com a mediana ou a moda, como valores representativos de tendência central.

Exercício 17. *Uma amostra de 20 operários de uma companhia apresentou os seguintes salários recebidos durante certa semana, arredondados para o dólar mais próximo e apresentados em ordem crescente:*

140	140	140	140	140	140	140	140	155	155
165	165	180	190	200	205	225	225	230	240

Encontre:

a) A média;

Resolução:

Utilizaremos o programa R, para encontrarmos as medidas de tendências centrais. Logo, a média é encontrada como:

```
> a <- scan()
> mean(a)

[1] 172.75
```

b) A mediana;

Resolução:

```
> median(a)

[1] 160
```

c) A moda para este grupo de salários.

Resolução:

Primeiramente, colocaremos o conjunto de dados em uma tabela pontual, e só depois encontraremos a moda.

```
> a.tabela <- table(a)
> a.tabela

a 140 155 165 180 190 200 205 225 230 240
  8   2   2   1   1   1   1   2   1   1
```

```
> a.moda <- names(a.tabela)[a.tabela == max(a.tabela)]
> a.moda

[1] "140"
```

Exercício 18. Um especialista em padrões de trabalho observa, em um escritório, a quantidade de tempo requerida para a digitação de uma amostra de 9 cartas, com os seguintes resultados enumerados em ordem crescente, arredondados para o minuto mais próximo: 5, 5, 5, 7, 9, 14, 15, 16, 18. Determinar:

a) A média;

Resolução:

```
> a <- scan()
> a

[1] 5 5 5 7 9 14 15 16 18

> mean(a)

[1] 10.44444
```

b) A mediana;

Resolução:

A mediana é o elemento que ocupa a posição central, logo:

```
> median(a)

[1] 9
```

c) A moda para este grupo de valores.

Resolução:

A moda é o elemento que ocorre com maior frequência. Desta forma, a tabela abaixo mostra o elemento com maior ocorrência:

```
> a.tabela <- table(a)
> a.tabela

a
5 7 9 14 15 16 18
3 1 1 1 1 1 1

> a.moda <- names(a.tabela)[a.tabela == max(a.tabela)]
> a.moda
```

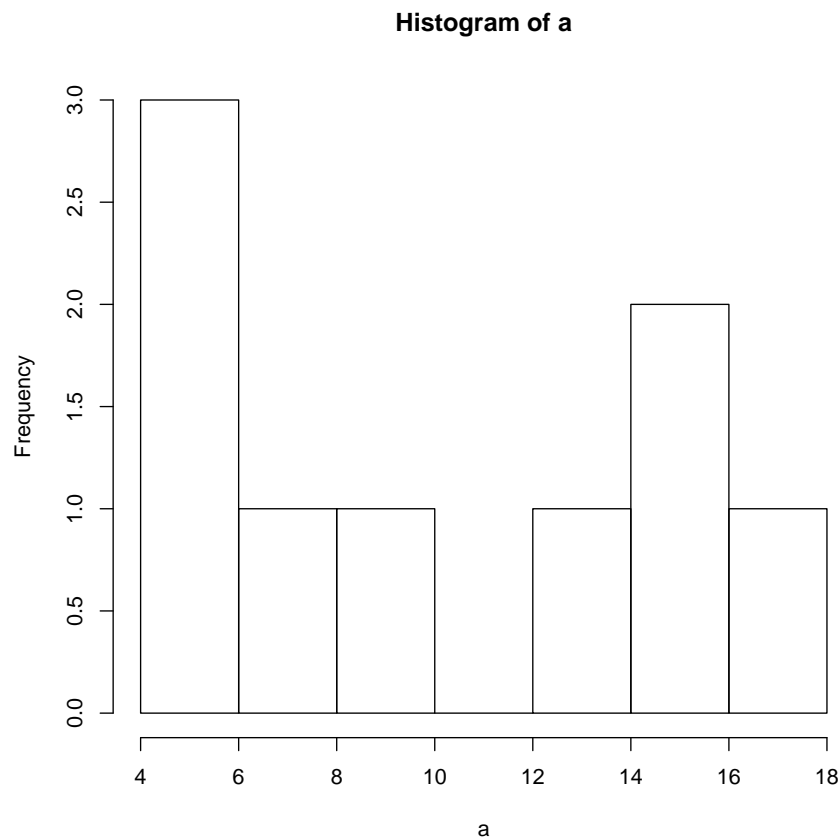
[1] "5"

Portanto, o valor 5 é o valor que ocorre mais vezes, logo é a moda da distribuição.

Exercício 19. *Comparar os valores da média, da mediana e da moda do Problema anterior e comentar a forma da distribuição.*

Resolução:

```
> hist(a)
```



Através da análise do gráfico acima e da média, mediana e moda é possível perceber que a distribuição dos tempos de digitação possui um comportamento assimétrico.

Exercício 20. *Determine a média, a mediana e a moda. Supor que estes são todos os apartamentos de determinada área geográfica.*

Resolução:

Mais uma vez utilizaremos o R para resolvermos o problema abaixo. É interessante deixar claro que o problema não deve ser resolvido única e exclusivamente utilizando os recursos computacionais. O uso de calculadoras financeiras também poderá resultar nos mesmos resultados.

Distribuição de frequência de taxas mensais de aluguel de apartamentos.

Aluguel	Ponto Médio da Classe (X_i)	Número de Apartamentos (F_i)
150 – 179	164,50	3
180 – 209	194,50	8
210 – 239	224,50	10
240 – 269	254,50	13
270 – 299	284,50	33
300 – 329	314,50	40
330 – 359	344,50	35
360 – 389	374,50	30
390 – 419	404,50	16
420 – 449	434,50	12
<i>Total</i>	–	200

Primeiramente, entraremos com a tabela no R.

```
> aluguel <- seq(165, 435, by = 30)
> n.aparta <- c(3, 8, 10, 13, 33, 40, 35, 30, 16, 12)
> dados <- rep(aluguel, n.aparta)
```

Logo a média é encontrada com o comando (mean):

```
> mean(dados)
```

```
[1] 323.25
```

Portanto, a distribuição média das taxas mensais de aluguel é de R\$323,25.

O cálculo da mediana é determinado pelo seguinte procedimento.

Exercício 21. *Um canal de comunicações está sendo monitorado pelo registro do número de erros em um conjunto de caracteres (string) de 1.000 bits. Dados para 20 desses conjuntos são visto a seguir. Leia os dados da esquerda para a direita.*

3	1	0	1	3	2	4	1	3	1
1	1	2	3	3	2	0	2	0	1

- (a) *Construa um diagrama de ramo e folhas dos dados;*
- (b) *Encontre a média, mediana e moda para os dados acima.*

Resolução:

Para encontrarmos o diagrama de ramo e folhas o comando utilizado no R é o :

stem.leaf(basededados).

Logo, temos:


```

> a <- scan()
> stem.leaf(a)

1 | 2: represents 1.2
  leaf unit: 0.1
      n: 20

      3   0* | 000
        0. |
     10   1* | 0000000
        1. |
     10   2* | 0000
        2. |
      6   3* | 00000
        3. |
      1   4* | 0

```

Por efeito didático, construiremos uma tabela pontual para encontrarmos a média, mediana e moda. Mais a utilização de comandos diretos como (mean e median), encontrariam as medidas com maior rapidez.

Desta forma, temos:

```

> b <- scan()
> table(b)

b
0 1 2 3 4
3 7 4 5 1

```

Acima, temos a tabela pontual para o conjunto de dados.

Logo, para calcularmos a média, mediana e moda da tabela pontual, iremos proceder da seguinte forma:

```

> mediab <- mean(b, na.rm = T)
> mediab

[1] 1.7

```

O valor da media foi 1.7.

```
> medianab <- median(b, na.rm = T)
> medianab
```

```
[1] 1.5
```

O valor da mediana foi 1.5.

```
> moda <- names(table(b))[table(b) == max(table(b))]
> moda
```

```
[1] "1"
```

E o valor com maior ocorrência foi o 1, sendo chamado de moda.

Exercício 22. *Uma amostra de vinte empresas, de porte médio, foi escolhida para um estudo sobre o nível educacional dos funcionários do setor de vendas. Os dados coletados, quanto ao número de empregados com curso superior completo, são apresentados abaixo.*

<i>Empresa</i>	1	2	3	4	5	6	7	8	9	10	11
<i>N. Funcionários</i>	1	0	0	3	0	1	1	2	2	2	0

<i>Empresa</i>	12	13	14	15	16	17	18	19	20
<i>N. Funcionários</i>	2	0	2	0	1	1	2	3	2

(a) organize uma tabela de frequência pontual;

Resolução:

```
> n.funcion <- scan()
> tabela <- table(n.funcion)
> tabela
```

```
n.funcion
0 1 2 3
6 5 7 2
```

Acima a tabela pontual. Para encontrarmos a média, procederemos da seguinte forma:

(b) Calcule a média, mediana e moda.

Resolução:

```
> media <- mean(n.funcion, na.rm = T)
> media
```

```
[1] 1.25
```

Logo, a média para o número de trabalhadores nas empresas com nível superior é 1,25.

```
> mediana <- median(n.funcion, na.rm = T)
> mediana
```

```
[1] 1
```

A mediana para o número de trabalhadores com nível superior é 1.

E a moda é encontrada da seguinte forma:

```
> moda <- names(tabela)[tabela == max(tabela)]
> moda
```

```
[1] "2"
```

Exercício 23. Para o conjunto de dados abaixo, entre a média e mediana.

2	3	5	7
8	4	55	25
32	548	2	1
0	1	2	1

O que se pode dizer, em relação a média neste exercício.

Resolução:

```
> valores <- scan()
```

O valor médio encontrado é:

```
> mean(valores)
```

```
[1] 43.5
```

E o valor mediano é:

```
> median(valores)
```

```
[1] 3.5
```

Desta forma, é possível observar que a média sofre influência dos valores atípicos no conjunto de dados acima, o que não ocorre com a mediana.

Capítulo 3

Principais Separatrizes

Passamos, agora, a uma análise mais pormenorizada das medidas separatrizes - último passo antes de adentrarmos no estudo das medias de dispersão.

Em um momento anterior, quando iniciamos o estudo da mediana, já havíamos feito as primeiras considerações acerca das medidas separatrizes, afirmando que são também medidas de posição (assim como as medidas de tendência central - media, moda e mediana), Vimos também que a mediana classifica-se tanto como medida de tendência central quanto como medida separatriz, e que as separatrizes - como o próprio nome sugere - são aquelas medidas que "separam" ou que dividem o conjunto em um certo numero de partes iguais.

No caso da mediana, vimos que ela divide o conjunto em duas metades. Já o quartil, separa o conjunto em quatro partes iguais; o decil, em dez partes e, finalmente, o centil (ou percentil), em cem partes iguais.

Recordando disso, lembraremos também que aprendemos uma relação importantíssima entre as quatro medidas separatrizes. Na verdade, é uma relação ate visual, que não precisamos fazer esforço para "decorar", bastando traçar uma reta (que representará o conjunto), e depois fazer as divisões, exatamente como mostramos anteriormente quando estudamos a mediana) e transcrevemos abaixo:

<i>Md</i>								
<i>Q₂</i>								
<i>D₁</i>	<i>D₂</i>	<i>D₃</i>	<i>D₄</i>	<i>D₅</i>	<i>D₆</i>	<i>D₇</i>	<i>D₈</i>	<i>D₉</i>
<i>C₁₀</i>	<i>C₂₀</i>	<i>C₃₀</i>	<i>C₄₀</i>	<i>C₅₀</i>	<i>C₆₀</i>	<i>C₇₀</i>	<i>C₈₀</i>	<i>C₉₀</i>

Daí, concluímos sem maiores dificuldades que:

$$Md = Q_2 = D_5 = C_{50}$$

A mediana já sabemos calcular. Aprenderemos, agora, como determinar o valor das demais medidas separatrizes.

3.1 Determinação do Quartil

Já sabemos que, para dividir um conjunto em quatro partes iguais, precisamos marcar três pontos apenas (como vimos no desenho acima). Portanto, já sabemos que existem três quartis, os quais designaremos por Q_1 (primeiro quartil), Q_2 (segundo quartil) e Q_3 (terceiro quartil).

Quando estudamos a *mediana*, vimos que as questões que exigiam o cálculo desta medida costumavam dizer apenas algo como "determine o valor da *mediana* deste conjunto" (e só). Isso porque existem somente uma *mediana*. Porém, em se tratando do *quartil*, um enunciado jamais poderia dizer apenas "determine o valor do quartil". Se assim o fizesse, ficaria no ar a pergunta: "Qual deles?". Se existem três quartis, uma questão de prova teria, logicamente, que explicitar qual deles está exigindo.

Ocorre que, normalmente, as provas não contemplam as *medidas separatrizes* como uma questão exclusiva. Explicando melhor: não costuma cair uma questão exigindo que se calcule este ou aquele quartil, este ou aquele decil... O que se pede é que se determine, por exemplo, o *coeficiente quartílico de assimetria*, ou o *coeficiente percentílico de curtose*. Ainda não estudamos esses assuntos - *assimetria e curtose* -, mas já podemos adiantar que, na determinação desses referidos coeficientes, se fará necessário o conhecimento das medidas separatrizes.

Em suma: os quartis, decis e percentis serão, normalmente, calculados como um meio para se chegar ao fim desejado pelo enunciado. Este fim será, provavelmente, um coeficiente de assimetria ou de curtose (assuntos que veremos em capítulos seguintes).

Outra coisa importante: quem sabe calcular a *mediana*, fatalmente não terá dificuldades em aprender a determinar as outras medidas separatrizes. Daremos ênfase à determinação do quartil, decil e percentil no âmbito das *distribuições de frequências*, que é a forma comumente exigida em prova.

Lembremos como se acha a *mediana* para uma *distribuição de frequência*. Por primeiro, temos que encontrar a classe mediana. Para isso, fazemos a conta $(n/2)$ - independentemente de n ser um valor par ou ímpar - e depois comparamos este valor $(n/2)$ com os valores da coluna de frequência acumulada (*fac*), fazendo a pergunta de praxe que aprendemos: esta *fac* é maior ou igual a $(n/2)$?. Repetiremos a pergunta até que a resposta seja afirmativa. Daí, a classe correspondente será a *classe medianica*.

3.2 Calculando o primeiro quartil - Q_1

Para calcular o primeiro quartil, temos antes que determinar qual será a *classe do primeiro quartil*.

Lembremos que, no caso da *mediana*, a primeira conta que fazíamos era $(n/2)$. Dividíamos o n por 2, exatamente porque a *mediana* divide o conjunto em duas partes. Agora,

sabemos que o *quartil* divide o conjunto em quatro partes. Portanto, a conta que faremos (para o primeiro quartil) é a seguinte:

$$25\% \text{ de } n.$$

Para fazer esta conta, também não nos preocuparemos se n é um valor par ou ímpar (da mesma forma da mediana). Feita esta conta, passaremos a comparar seu resultado com os valores de fac , exatamente da mesma forma que fizemos para achar a *classe medianica*. A pergunta, agora adaptada ao *quartil*, será a seguinte:

Esta fac é maior ou igual a 25% de n ?

Enquanto a resposta for negativa, passaremos para a classe seguinte, e repetiremos a pergunta, até o momento em que a resposta for **SIM!** Ao chegarmos à resposta afirmativa, pararemos e procuraremos a classe correspondente. Esta será a *classe do primeiro quartil*. Ou seja, será desta classe que iremos extrair os dados para usar na fórmula do Q_1 .

Vejam que, até aqui, a única diferença observada nos passos para achar o *quartil* e a *mediana*, foi que agora fazemos (25% de n) – em vez de (50% de n) – e comparamos este (25% de n) com a coluna da fac .

Uma vez constatado qual é a classe do primeiro quartil, só nos restará aplicar a fórmula. A facilidade em se memorizar a fórmula do Q_1 é absoluta. Vamos recordar a fórmula da mediana:

$$\tilde{X} = l_{\text{inf}} + \frac{(50\% \text{ de } n - fac_{\text{ant}})}{F_{\tilde{x}}} \cdot h$$

Agora é só pensar o seguinte: o que mudou até aqui para o quartil foi que (50% de n) passou a ser (25% de n). Então também será apenas isso que irá mudar na fórmula. Daí, o primeiro será determinado por:

$$Q_1 = l_{\text{inf}} + \frac{((25\% \text{ de } n) - fac_{\text{ant}})}{F_i} \cdot h$$

Ora, esta fórmula nos fala em limite inferior (l_{inf}), em amplitude da classe (h), além de duas frequências - F_i e fac_{ant} . A única coisa que teremos de lembrar é que todos esses dados serão retirados, tomando como referência a *classe do primeiro quartil*.

Em suma, os passos para determinação do Q_1 de um conjunto de dados serão os seguintes:

- (a) determinamos o n (somando a coluna da F_i);
- (b) calculamos o valor de (25% de n) (independentemente de n ser par ou ímpar);
- (c) construímos a coluna da fac ;

- (d) comparamos o valor do (25% de n) com os valores da fac , iniciando da fac da primeira classe (a mais de cima) e fazendo a seguinte pergunta: "esta fac é maior ou igual a (25% de n)?". Se a resposta for NÃO, passamos à fac da classe seguinte. Quando a resposta for SIM, pararemos e procuraremos a classe correspondente. Esta será a nossa classe do primeiro quartil;
- (e) finalmente, aplicaremos a fórmula do Q_1 , extraíndo os dados desta classe do Q_1 , que acabamos de encontrar. Novamente a fórmula:

$$Q_1 = l_{\text{inf}} + \frac{((25\% \text{ de } n) - fac_{\text{ant}})}{F_i} \cdot h$$

Vamos a um exemplo. Para o conjunto abaixo, determinemos o valor do primeiro quartil.

Classe	F_i
0 † 10	2
10 † 20	5
20 † 30	8
30 † 40	6
40 † 50	3

Solução:

1a. Passo: encontraremos n e calcularemos (25% de n);

Classe	F_i
0 † 10	2
10 † 20	5
20 † 30	8
30 † 40	6
40 † 50	3
Total	24

Daí, achamos que $n = 24$ e, portanto, (25% de n) = $0,25 \times 24 = 6$.

2a. Passo: construímos a fac :

Classe	F_i	fac
0 † 10	2	2
10 † 20	5	7
20 † 30	8	15
30 † 40	6	21
40 † 50	3	24
Total	24	-

3a. Passo: comparamos os valores da *fac* com o valor de (25% de n), fazendo a pergunta de praxe, adaptada ao primeiro quartil.

Classe	F_i	<i>fac</i>	
0 † 10	2	2	2 é maior ou igual a 6? NÃO!
10 † 20	5	7	7 é maior ou igual a 6? SIM!
20 † 30	8	15	
30 † 40	6	21	
40 † 50	3	24	
Total	24	-	

Como a resposta foi afirmativa na segunda *fac*, procuramos a classe correspondente (10 † 20) e dizemos que esta será nossa classe do primeiro quartil.

4a. Passo: só nos resta agora aplicar a fórmula do primeiro quartil, tornando como referência a classe do Q_1 , que acabamos de encontrar. Teremos:

$$Q_1 = l_{inf} + \frac{(25\% \text{ de } n - fac_{ant})}{F_i} \cdot h$$

$$Q_1 = 10 + \frac{(6 - 2)}{5} \cdot 10$$

$$Q_1 = 18$$

3.3 O segundo e o terceiro quartil

A determinação do Q_2 e do Q_3 é semelhante à do Q_1 , com uma pequena diferença. É preciso sabermos do seguinte:

o que irá ser alterado na determinação do cálculo destas medidas separatrizes é exatamente aquela fração que aparece no numerador da fórmula

No caso da mediana, a fração é (50% de n); no caso do primeiro quartil, é (25% de n); nos demais quartis, como será?

Para o segundo quartil, teremos um acúmulo de (25% de n) do quartil anterior, logo o segundo quartil ficará (50% de n). Daí, a fórmula do segundo quartil - Q_2 - é a seguinte:

$$Q_2 = l_{inf} + \frac{((50\% \text{ de } n) - fac_{ant})}{F_i} \cdot h$$

- Ou seja, o segundo quartil é igual a Mediana.

E disso já sabíamos: o segundo quartil é a própria mediana.

Portanto, não vacilaremos na prova. Se o enunciado da questão fornecer um conjunto de dados e solicitar que determinemos o Q_2 , não nos restará qualquer dúvida: calcularemos a mediana.

Já no caso do terceiro quartil, repete-se (50% de n) e acumula-se com mais 25%, logo teremos (75% de n). Logo teremos a seguinte fórmula para determinar o terceiro quartil:

$$Q_3 = l_{\text{inf}} + \frac{((75\% \text{ de } n) - fac_{\text{ant}})}{F_i} \cdot h$$

Ora, conhecer a fração que consta na fórmula da medida separatriz implica conhecer também o primeiro passo para encontrá-la.

Senão vejamos: no cálculo da mediana, calculávamos o valor de (50% de n); no cálculo do primeiro quartil, calculávamos o valor de (25% de n).

Por mera dedução, o primeiro passo para encontrarmos o valor do terceiro quartil será exatamente calcularmos o valor de (75% de n).

Os passos para determinação do Q_3 serão, portanto, os seguintes:

- (a) determinamos o n (somando a coluna da F_i);
- (b) calculamos o valor de (75% de n) (independentemente de n ser par ou ímpar);
- (c) construímos a coluna da fac ;
- (d) comparamos o valor do (75% de n) com os valores da fac , iniciando da fac da primeira classe (a mais de cima!) e fazendo a seguinte pergunta: "esta fac é maior ou igual a (75% de n)?". Se a resposta for **NÃO**, passamos à fac da classe seguinte. Quando a resposta for **SIM**, pararemos e procuraremos a classe correspondente. Esta será a nossa classe do terceiro quartil;
- (e) finalmente, aplicaremos a fórmula do Q_3 , extraíndo os dados desta classe do Q_3 , que acabamos de encontrar. Novamente a fórmula:

$$Q_3 = l_{\text{inf}} + \frac{((75\% \text{ de } n) - fac_{\text{ant}})}{F_i} \cdot h$$

Neste momento, os bons observadores já perceberam que a única diferença verificada nos passos descritos para calcularmos o primeiro e o terceiro quartil consiste naquela fração presente no numerador da fórmula de cada *medida separatriz*.

Já perceberam também que esta fração é quem define tudo. Ela será o valor de referência, que utilizaremos para realizar a comparação com a coluna da frequência absoluta (fac), para efeitos de encontrarmos a classe da *medida separatriz*, ou seja, a classe que usaremos para lançar os dados na fórmula.

Façamos um exemplo para cálculo do Q_3 .

Exemplo. Para o conjunto de dados abaixo, determinemos o valor do terceiro quartil.

Classe	F_i
0 † 10	2
10 † 20	5
20 † 30	8
30 † 40	6
40 † 50	3

Solução:

1a. Passo: encontraremos n e calcularemos (75% de n);

Classe	F_i
0 † 10	2
10 † 20	5
20 † 30	8
30 † 40	6
40 † 50	3
Total	24

Daí, achamos que $n = 24$ e, portanto, (75% de n) ou $(0,75 \times 24) = 18$.

2a. Passo: construímos a fac .

Classe	F_i	fac
0 † 10	2	2
10 † 20	5	7
20 † 30	8	15
30 † 40	6	21
40 † 50	3	24
Total	24	-

3a. Passo: comparamos os valores da fac com o valor de (75% de n), fazendo a pergunta de praxe, adaptada ao terceiro quartil.

Classe	F_i	fac	
0 † 10	2	2	2 é maior ou igual a 18? NÃO!
10 † 20	5	7	7 é maior ou igual a 18? NÃO!
20 † 30	8	15	15 é maior ou igual a 18? NÃO!
30 † 40	6	21	21 é maior ou igual a 18? SIM!
40 † 50	3	24	
Total	24	-	

Como a resposta **SIM** surgiu na *fac* da quarta classe (30 † 40), diremos que esta será nossa *classe do terceiro quartil*.

4a. Passo: aplicaremos a fórmula do Q_3 , usando os dados da classe do Q_3 , que acabamos de identificar.

Teremos:

$$\begin{aligned} Q_3 &= l_{inf} + \frac{(75\% \text{ de } n - fac_{ant})}{F_i} \cdot h \\ Q_3 &= 30 + \frac{(18 - 15)}{6} \cdot 10 \\ Q_3 &= 35 \end{aligned}$$

3.4 O primeiro decil - D_1

Como já aprendemos aqui, o decil dividirá o conjunto em dez partes iguais. Assim, a fração que constará no numerador da fórmula do primeiro *decil* será justamente (10% de n).

Daí, faremos o seguinte: independentemente de n ser um valor par ou ímpar, calcularemos o valor de (10% de n) e compararemos este valor com a coluna da *fac*. a nossa pergunta de praxe, agora adaptada ao *primeiro decil* será: "**esta fac é maior ou igual a (10% de n)?**"

E por que faremos isso? Porque precisamos encontrar a *classe do primeiro decil*, ou seja, precisamos identificar a classe da qual extrairemos os dados para utilizarmos na fórmula do D_1 .

Quando encontrarmos a classe do D_1 , só teremos que aplicar a fórmula do D_1 . A fórmula do D_1 será igual à da *mediana*, com uma única diferença. Qual? Em lugar de (50% de n), aparecerá a fração (10% de n), uma vez que o *decil* divide o conjunto de dados em dez partes iguais.

Estamos percebendo que os passos todos se identificam, quando se trata de determinarmos as *medidas separatrizes*.

Serão, portanto, os seguintes os passos adotados para o cálculo do *primeiro decil*:

- (a) determinamos o n (somando a coluna da F_i);
- (b) calculamos o valor de (10% de n) (independentemente de n ser par ou ímpar);
- (c) construímos a coluna da *fac*;
- (d) comparamos o valor do (10% de n) com os valores da *fac*, iniciando da *fac* da primeira classe (a mais de cima) e fazendo a seguinte pergunta: "*esta fac é maior ou igual a (10% de n)?*" Se a resposta for **NÃO**, passamos à *fac* da classe seguinte. Quando a resposta for **SIM**, pararemos e procuraremos a classe correspondente. Esta será a nossa *classe do primeiro decil*;

- (e) finalmente, aplicaremos a fórmula do D_1 , extraindo os dados desta classe do D_1 , que acabamos de encontrar. Eis a fórmula:

$$D_1 = l_{\text{inf}} + \frac{((10\% \text{ de } n) - fac_{\text{ant}})}{F_i} \cdot h$$

Vamos a um exemplo. Para o conjunto abaixo, determine o valor do primeiro decil.

Classe	F_i
0 † 10	2
10 † 20	5
20 † 30	8
30 † 40	6
40 † 50	3
Total	24

Solução:

1a. Passo: Encontraremos n e calculamos (10% de n);

Classe	F_i
0 † 10	2
10 † 20	5
20 † 30	8
30 † 40	6
40 † 50	3
Total	24

Daí, achamos que $n = 24$ e, portanto, (10% de n) ou $(0,10 \times 24) = 2,4$.

2a. Passo: construímos a fac .

Classe	F_i	fac
0 † 10	2	2
10 † 20	5	7
20 † 30	8	15
30 † 40	6	21
40 † 50	3	24
Total	24	-

3a. Passo: comparamos os valores da fac com o valor de (10% de n), fazendo a pergunta de praxe, adaptada ao primeiro decil.

Classe	F_i	fac	
0 † 10	2	2	2 é maior ou igual a 2,4? NÃO!
10 † 20	5	7	7 é maior ou igual a 2,4? SIM!
20 † 30	8	15	
30 † 40	6	21	
40 † 50	3	24	
Total	24	-	

Como a resposta **SIM** surgiu na fac da quarta classe (10 † 20), diremos que esta será nossa *classe do primeiro decil*.

4a. Passo: aplicaremos a fórmula do D_1 , usando os dados da classe do D_1 , que acabamos de identificar.

Teremos:

$$D_1 = l_{inf} + \frac{(10\% \text{ de } n - fac_{ant})}{F_i} \cdot h$$

$$D_1 = 10 + \frac{(2,4 - 2)}{5} \cdot 10$$

$$D_1 = 10,8$$

3.5 Calculando os outros decis - D_2 a D_9

Estamos quase prontos para generalizar o nosso entendimento sobre as *medidas separatrizes*. Vejamos apenas o que haverá de novo na determinação dos demais *decis*.

Já sabemos que o que diferencia uma *medida separatriz* de outra, para fins de cálculo, é aquela fração que aparece no numerador da fórmula. Para o *primeiro decil* (D_1), essa fração é (10% de n), conforme vimos acima. E, para os demais *decis*, qual será a fração de cada um deles?

Para o *segundo decil*, teremos (20% de n), assim teremos:

$$D_2 = l_{inf} + \frac{((20\% \text{ de } n) - fac_{ant})}{F_i} \cdot h$$

Dai, concluímos que a fórmula do D_9 será a seguinte fórmula:

$$D_9 = l_{inf} + \frac{((90\% \text{ de } n) - fac_{ant})}{F_i} \cdot h$$

3.6 Calculando os percentis

Restaram agora os *percentis*. Lembraremos que o *percentis* (ou centil) dividirá o conjunto em cem partes iguais. Por analogia, já podemos concluir que a fração do numerador da fórmula para o primeiro centil será (1% de n).

Daí, a sequência de passos que usaremos para determinar os *percentis*, usando o mesmo artifício para encontrarmos o X -ésimo *percentil*, logo temos:

$$P_1 = l_{\text{inf}} + \frac{((1\% \text{ de } n) - fac_{\text{ant}})}{F_i} \cdot h$$

Para encontrarmos o primeiro percentil ou (1% de n).

Para calcularmos o P_{23} , temos:

$$P_{23} = l_{\text{inf}} + \frac{((23\% \text{ de } n) - fac_{\text{ant}})}{F_i} \cdot h$$

E assim, sucessivamente.

$$P_{83} = l_{\text{inf}} + \frac{((83\% \text{ de } n) - fac_{\text{ant}})}{F_i} \cdot h$$

3.7 Exercícios

Exercício 24. *Determine para o conjunto abaixo os valores do primeiro quartil, terceiro quartil, primeiro decil e nono decil.*

Classe	F_i
0 - 15	4
15 - 30	13
30 - 45	15
45 - 60	10
60 - 75	6
Total	

E interprete os resultados encontrados.

Exercício 25. *A tabela abaixo mostra a distribuição de frequência do número de filhos dos pais de alunos da FNH, considerando uma amostra de 212 estudantes, entrevistados pelos alunos do curso de Administração, 2001. Obtenha o primeiro quartil, segundo quartil, terceiro quartil, mediana e o septuagésimo sétimo percentil e interprete os resultados obtidos.*

N. de filhos	1	2	3	4	5	6	7	8	9	10	11	12
Frequência	10	45	32	50	23	23	9	7	6	2	3	2

Exercício 26. *Apresentamos a seguir os resultados da segunda prova de estatística da turma N2 do 1a. semestre de 2003 da disciplina de Estatística I.*

6	12	12	14	15	15	15	15	16
17	18	18	19	19	19	20	21	21
22	22	22	23	23	23	23	23	23
24	25	25	25	27	27	28	32	

- (a) Calcule a média e a mediana;
- (b) Calcule o primeiro e terceiro quartis. Explique o significado destes números.

Exercício 27. A tabela a seguir apresenta os dados de um teste de psico-analogia (um teste de inteligência em que um indivíduo tem que resolver uma série de analogias). A amostra contém 158 indivíduos que receberam notas de acordo com o rendimento no teste.

Escores	Freq. simples
33 † 36	1
36 † 39	3
39 † 42	1
42 † 45	4
45 † 48	3
48 † 51	7
51 † 54	15
54 † 57	23
57 † 60	16
60 † 63	24
63 † 66	21
66 † 69	21
69 † 72	15
72 † 75	4

- (a) Localize a primeiro quartil, a mediana e o percentil de ordem 90 (P_{90}). E interprete este resultados.
- (b) Acima de que nota encontram-se 80% dos indivíduos? A que percentil corresponde este valor?

Capítulo 4

Medidas de Variabilidade

Neste capítulo abordaremos a característica da variação, de grande importância para a estatística, sendo, por isso, uma das principais de todo o curso. O estudante deve dominar os seguintes conceitos-chaves: (1) a variação se refere a quanto os valores podem diferir entre si e pode ser medida por números específicos; (2) os números relativamente próximos uns dos outros têm baixas medidas de variação, enquanto os valores mais dispersos têm maior medida de variação; (3) o desvio padrão é uma medida de variação particularmente importante, e devemos saber calculá-lo para um conjunto de valores; (4) os valores dos desvios padrão devem ser interpretados corretamente.

Quase nunca uma única medida é suficiente para descrever de modo satisfatório um conjunto de dados. Tomemos como exemplo o caso da média aritmética, que é uma medida de localização, ou de tendência central, largamente empregada, e consideremos os dois conjuntos de observações:

$$A: \{25, 28, 31, 34, 37\} \quad B: \{17, 23, 30, 39, 46\}$$

Ambos têm a mesma média, $\bar{x} = 31$. No entanto, percebe-se, intuitivamente, que o conjunto B acusa dispersão muito maior que o conjunto A. torna-se então necessário estabelecer medidas que indiquem o grau de dispersão, ou variabilidade, em relação ao valor central.

4.1 Amplitude e Desvio Médio

A medida de dispersão mais simples é a *amplitude*. Define-se amplitude como a diferença entre o maior e o menor valor do conjunto. No exemplo anterior, a amplitude de A é $37 - 25 = 12$, enquanto a de B é $46 - 17 = 29$. A amplitude de B é quase 2,5 vezes a de A. Fácil de calcular a amplitude tem a desvantagem de levar em conta apenas dois valores, desprezando todos os outros. Poderíamos pensar também na soma das diferenças dos valores do conjunto em relação à sua média:

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})$$

Mas

$$\sum_{i=1}^n (x_i - \bar{x}) = x_1 + x_2 + \dots + x_n - n\bar{x} = n\bar{x} - n\bar{x} = 0$$

Então, a soma dos desvios em relação à média não serve como medida de dispersão, por ser identicamente nula. Por ela, todos os conjuntos teriam variabilidade nula. Entretanto, a idéia de considerar a soma dos desvios em relação à média é boa. Se retirarmos o efeito dos sinais da diferença, conseguimos uma boa medida de variabilidade. Isso pode ser feito de duas maneiras: tomando-se o módulo da diferença ou o quadrado da diferença. A primeira opção leva ao desvio médio DM.

$$DM = \frac{\sum |x_i - \bar{x}|}{n}$$

Embora intuitivamente atraente, essa medida é pouco utilizada.

4.2 Variância e Desvio Padrão

Consideremos então a soma dos quadrados dos desvios em relação à média. Com ela, estabeleceremos uma medida de variabilidade para um conjunto de dados, chamada variância, denotada por s^2 e definida como:

- $\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$ (Variância Populacional)
- $s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$ (Variância Amostral)

Por motivos associados à inferência estatística, é usual utilizar $n - 1$ em lugar de n na expressão acima. Adotaremos essa prática.

A raiz quadrada da variância é chamada desvio padrão; representa-se por s :

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \text{ ou } s = \sqrt{s^2}$$

- **Obs.: A unidade de medida do desvio padrão é a mesma dos dados originais. O cálculo do desvio padrão exige o cálculo prévio da variância.**

De modo geral, o desvio padrão é a mais importante e mais útil medida de variação. Ao contrario da amplitude, o desvio padrão leva em conta todos os valores, mas essa vantagem torna o cálculo mais difícil. Mostraremos a seguir aplicações do desvio padrão, mas para entender perfeitamente esse conceito, é preciso atenção aos exemplos.

4.2.1 Desvio padrão para dados não ordenados.

Muitos bancos costumavam exigir que os clientes formassem filas separadas para os diversos guichês, mas recentemente passaram a adotar fila única. Qual o motivo dessa modificação? O tempo médio de espera não se modifica, porque a fila de espera não afeta a eficiência dos caixas. A adoção de fila única se deveu ao fato de os clientes preferirem

tempos de espera mais consistentes com menor variação. Assim é que milhares de bancos efetuaram uma modificação que resultou em uma variação menor (e clientes mais satisfeitos), mesmo que a média não tenha sido afetada. Consideremos agora uma amostra de dados bancários usados em uma prova de mestrado da ANPAD. Os valores relacionados são tempos de espera (em minutos) de clientes.

Exemplo 4.1. *Os clientes do B.B entram em uma fila única que é atendida por três caixas. Os clientes da C.E podem entrar em qualquer uma de três filas que conduzem a três guichês. Se calcularmos a média de espera de ambos os bancos, veremos que possuem a mesma média, 7,15, a mesma mediana de 7,20, a mesma moda de 7,7. Com base apenas nestas medidas de tendência central, poderíamos admitir que os tempos de espera nos dois bancos fossem praticamente os mesmos. Todavia, esquadrinhado os tempos de espera originais, constataríamos uma diferença fundamental: O B.B tem tempos de espera com muito menos variação do que a C.E. Mantidas todas as outras características, os clientes provavelmente preferirão o B.B, onde não correm o risco de entrar em uma fila muito mais lenta do que as outras.*

BB	6,5	6,6	6,7	6,8	7,1	7,3	7,4	7,7	7,7	7,7
CE	4,2	5,4	5,8	6,2	6,7	7,7	7,7	8,5	9,3	10,0

Para obtermos o desvio padrão, temos:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} \text{ (Variância Populacional)}$$

$$s^2 = \sum (x_i - \bar{x})^2 = (6,5 - 7,15)^2 + (6,6 - 7,15)^2 + (6,7 - 7,15)^2 + \dots + (7,7 - 7,15)^2 = 2,0450$$

Como há $n = 10$ valores, divida, pois por 9, ou seja, $(n - 1 = 10 - 1 = 9)$;

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{2,0450}{9} = 0,2272 \text{min}^2.$$

Portanto, o desvio padrão é igual a:

$$s = \sqrt{s^2}$$

Logo, $s = \sqrt{0,2272} = 0,48 \text{ min}$

Teoricamente, deveríamos dar aqui uma interpretação do desvio padrão de 0,48 min, mas essa interpretação será dada mais adiante.

Exemplo 4.2. *Calcule o desvio padrão da C.E. do exemplo acima.*

4.2.2 Desvio padrão - Dados Agrupados

Exemplo 4.3. Dada a distribuição abaixo, encontrar a média e o desvio padrão.

Classes	2 - 4	4 - 6	6 - 8	8 - 10	10 - 12	Total
F_i	2	4	7	4	3	20

Exemplo 4.4. Calcule a variância amostral para a serie abaixo:

X_i	2	3	5	6	7
F_i	1	4	5	3	2

Exemplo 4.5. Dada a amostra: 2, 3, 4, 5, 7, 10, 12;

- Qual é a amplitude amostral?
- Determine o desvio padrão;
- Calcule a variância.

Exemplo 4.6. Para a série: 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 9, 9.

- Construir a distribuição simples de frequência;
- Determinar o desvio padrão da tabela construída em a;
- Calcular a variância.

Exemplo 4.7. Calcule o desvio padrão para a tabela abaixo:

Classes	2 - 4	4 - 6	6 - 8	8 - 10	10 - 12	Total
F_i	3	5	8	6	3	25

Exemplo 4.8. Lançado um dado 50 vezes, obteve-se a seguinte distribuição:

X_i	1	2	3	4	5	6
F_i	6	11	6	7	9	11

Calcular a variância e o desvio padrão para a tabela acima.

4.3 Coeficiente de Variação

Trata-se de uma medida relativa de dispersão, útil para a comparação em termos relativos do grau de concentração em torno da média de séries distintas. É dado por:

$$CV = \frac{\sigma}{\bar{x}} \text{ ou } CV = \frac{s}{\bar{x}}$$

Exemplo 4.9. *Numa empresa, o salário médio dos homens é de R\$4.000 com desvio padrão de R\$1.500, e o das mulheres é em média de R\$3.000 com desvio padrão de R\$1.200. O que podemos concluir em relação aos salários dos homens e das mulheres.*

- Para os homens $CV = \frac{s}{\bar{x}} = \frac{1.500}{4.000} = 0,375$
- Para as mulheres $CV = \frac{s}{\bar{x}} = \frac{1.200}{3.000} = 0,40$

Logo, podemos concluir que os salários das mulheres apresentam maior dispersão relativa que os dos homens. Para obtermos o resultado do CV em porcentagens, basta multiplicarmos o resultado por 100. No caso, temos:

Para efeitos práticos, costuma-se considerar que CV superior a 50% indica alto grau de dispersão e, conseqüentemente, pequena representatividade da média. Enquanto para valores inferiores a 50%, a média será tanto mais representativa do fato quanto menor for o valor de seu CV.

Deve ficar claro que para grupos diferentes, usa-se o Coeficiente de variação para se ter idéia da consistência do grupo em estudo... Caso, o estudo seja feito no mesmo grupo, o indicado é a variância amostral.

Exemplo 4.10. *A seguir são apresentados os resultados da segunda prova das turmas de Logística e Geral. Compare as notas das turmas quanto a sua homogeneidade.*

TURMA	MÉDIA	DESVIO PADRÃO
Logística	22,5	4,5
Geral	24,0	5,4

Qual a turma mais homogênea ?

Exemplo 4.11. *Ache a média, a variância e o desvio padrão de cada uma das seguintes distribuições:*

a)

X_i	2	3	11
F_i	1/3	1/2	1/6

b)

X_i	-5	-4	1	2
F_i	1/4	1/8	1/2	1/8

Exemplo 4.12. Num certo bairro da cidade de São Paulo, as companhias de seguro estabeleceram o seguinte modelo para o número de veículos furtados por semana:

<i>Frutos</i>	0	1	2	3	4
F_i	1/4	1/2	1/8	1/16	1/16

Calcule a média e a variância do número de furtos semanais desse bairro.

4.4 Medidas de Assimétria

Já foi acentuado que, em uma distribuição simétrica, coincidem a média, a moda e a mediana e que os quartis ficam equidistantes da mediana, o que não ocorre numa distribuição assimétrica.

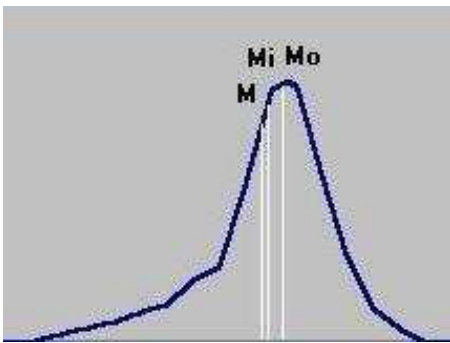


Figura 4.1: Assimétrica Negativa

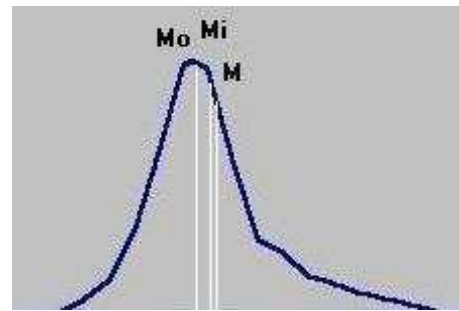


Figura 4.2: Assimétrica Positiva

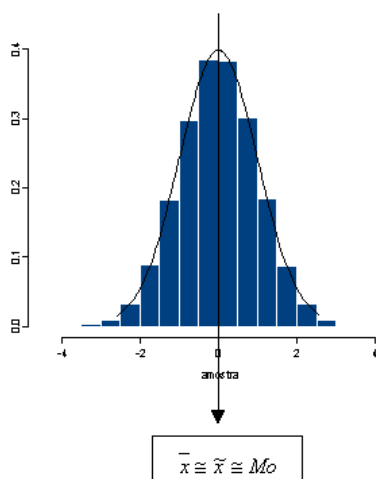


Figura 4.3: Simétrica

4.4.1 Coeficiente de Assimetria de Pearson

É uma medida usada para quantificar a assimetria da distribuição de um conjunto de dados. Pearson definiu um coeficiente de assimetria que é indicado por A_s e dado por:

$$A_s = \frac{\bar{x} - Mo}{S}$$

Se $|A_s| < 0,15$, considera-se a distribuição simétrica;

Se $0,15 \leq |A_s| \leq 1$, considera-se a distribuição moderadamente assimétrica;

Se $|A_s| > 1$, considera-se a distribuição fortemente assimétrica.

Em muitos casos já se considera a distribuição fortemente assimétrica se $|A_s| > 0,7$.

Caso a distribuição seja **amodal**, isto é, sem a presença da moda e a distribuição pareça levemente assimétrica, pode-se calcular A_s utilizando-se a mediana pela fórmula:

$$A_s = \frac{3(\bar{x} - \tilde{x})}{S}$$

4.5 Exercícios

Exercício 28. *Determinar o coeficiente de assimetria pelos dois processos para a distribuição:*

Classes	50 † 60	60 † 70	70 † 80	80 † 90	90 † 100
Freq.	15	20	30	20	15

Exercício 29. *Para a série 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 9, 9; determine:*

- Construir a distribuição de frequência pontual;
- Calcular a variância amostral;
- Determinar o desvio padrão;
- Calcule o coeficiente de variação;

Exercício 30. *Calcular a variância amostral:*

Classes	2 † 4	4 † 6	6 † 8	8 † 10	10 † 12
Freq.	3	5	8	6	3

Exercício 31. *Num teste aplicado a 20 alunos, obteve-se a seguinte distribuição de pontos:*

Classes	35 † 45	45 † 55	55 † 65	65 † 75	75 † 85	85 † 95
Freq.	1	3	8	3	3	2

- Calcule o desvio padrão;

- b) *Determine a variância amostral;*
- c) *Calcule o coeficiente de variação;*
- d) *Determinar o coeficiente de assimetria;*

Exercício 32. *Abaixo temos a distribuição de frequência dos pesos de uma amostra de 45 alunos:*

<i>Classes</i>	<i>40 † 45</i>	<i>45 † 50</i>	<i>50 † 55</i>	<i>55 † 60</i>	<i>60 † 65</i>	<i>65 † 70</i>
<i>Freq.</i>	<i>4</i>	<i>10</i>	<i>15</i>	<i>8</i>	<i>5</i>	<i>3</i>

- a) *Determinar a média amostral;*
- b) *Determinar a variância amostral;*
- c) *Qual o valor do coeficiente de variação?*
- d) *A distribuição é simétrica ?*

Exercício 33. *Sendo:*

<i>Classes</i>	<i>30 † 40</i>	<i>40 † 50</i>	<i>50 † 60</i>	<i>60 † 70</i>	<i>70 † 80</i>
<i>Freq.</i>	<i>10</i>	<i>20</i>	<i>35</i>	<i>25</i>	<i>10</i>

Calcular , S^2 , S , CV , AS .

4.6 Referências Bibliográficas

1. CARVALHO, Sérgio. Estatística Básica - Teoria e 150 questões. Editora IMPETUS, 2004
2. FURTADO, Daniel Ferreira. Estatística Básica. Editora UFLA. 2005
3. LEVINE, David M. Estatística: Teoria e Aplicações usando o microsoft excel em português. Editora LTC, 2003.
4. MAGALHÃES, Marcos Nascimento. Noções de Probabilidade e Estatística / São Paulo, 4ª edição, Editora USP, 2002

Capítulo 5

Introdução à Probabilidade Básica

5.1 Espaço Amostral - introdução

Encontramos na natureza dois tipos de fenômenos: determinísticos e aleatórios.

Os fenômenos determinísticos são aqueles em que os resultados são sempre os mesmos, qualquer que seja o número de ocorrências dos mesmos.

Se tomarmos um determinado sólido, sabemos que a uma certa temperatura haverá a passagem para o estado líquido. Este exemplo caracteriza um fenômeno determinístico.

Nos fenômenos aleatórios, os resultados não serão previsíveis, mesmo que haja um grande número de repetições do mesmo fenômeno.

Por exemplo: se considerarmos um pomar com centenas de laranjeiras, as produções de cada planta serão diferentes e não previsíveis, mesmo que as condições de temperatura, pressão, umidade, solo, etc., sejam as mesmas para todas as árvores.

Podemos considerar os experimentos aleatórios, que são fenômenos produzidos pelo homem.

Nos experimentos aleatórios, mesmo que as condições iniciais sejam sempre as mesmas, os resultados finais de cada tentativa do experimento, serão diferentes e não previsíveis.

- a) Lançamento de uma moeda honesta;
- b) Lançamento de um dado;
- c) Lançamento de duas moedas;
- d) Retirada de uma carta de um baralho completo de 52 cartas;
- e) Determinação da vida útil de um componente eletrônico.

A cada experimento aleatório está associado o resultado do mesmo, que não é previsível, chamado evento aleatório.

No exemplo *a* os ventos associados são cara (c) e coroa (r), no exemplo *b* poderá ocorrer uma das faces 1, 2, 3, 4, 5 ou 6.

5.2 Frequência Relativa

Consideremos o experimento que consiste em lançar uma mesma moeda n vezes.

Sejam m o número de vezes em que ocorre cara.

Definimos frequência relativa do evento cara como sendo:

$$f(c) = \frac{m}{n}$$

Notamos que $0 \leq f(c) \leq 1$.

Se aumentarmos o número de tentativas do experimento, a $f(c)$ tende a se estabilizar em torno de $\frac{1}{2}$.

Este fato é muito importante, pois quando dissermos que a probabilidade de um evento A é $P(A)$, estaremos dizendo que para um número bastante expressivo de tentativas de um experimento, a $f(a)$ tende a se estabilizar em torno de $P(A)$.

5.3 Espaço Amostral

Espaço amostral de um experimento aleatório é o conjunto dos resultados do experimento. Os elementos do espaço amostral serão chamados também de pontos amostrais.

Representaremos o espaço amostral por Ω .

Nos exemplos dados em 1.1, os espaços amostrais são:

- a) $\Omega = \{c, r\}$
- b) $\Omega = \{1, 2, 3, 4, 5, 6\}$
- c) $\Omega = \{(c, r), (c, c), (r, c), (r, r)\}$
- d) $\Omega = \{A_0, \dots, K_0, A_p, \dots, K_p, A_E, \dots, K_E, A_c, \dots, K_c\}$
- e) $\Omega = \{t \in \mathfrak{R} | t \geq 0\}$

O evento aleatório pode ser um único ponto amostral ou uma reunião deles, como veremos no exemplo:

Lançam-se dois dados. Enumerar os seguintes eventos:

A: saída de faces iguais.

B: saída de faces cuja soma seja igual a 10;

C: saída de faces cuja soma seja menor que 2;

D: saída de faces cuja soma seja menor que 15;

E: saída de faces onde uma face é o dobro da outra.

Determinação do espaço amostral: podemos determiná-lo por uma tabela de dupla entrada (produto cartesiano).

dados	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Os ventos pedidos são:

$$A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$$

$$B = \{(4, 6), (5, 5), (6, 4)\}$$

$$C = \phi \text{ (evento impossível)}$$

$$D = \Omega \text{ (evento certo)}$$

$$E = \{(1, 2), (2, 1), (2, 4), (3, 6), (4, 2), (6, 3)\}$$

5.3.1 Operações com eventos aleatórios

Consideremos um espaço amostral finito $\Omega = \{e_1, e_2, \dots, e_n\}$.

Sejam A e B dois eventos de $F(\Omega)$.

As seguintes operações são definidas:

a) REUNIÃO

Se $A \cup B = \{e_i \in \Omega | e_i \in A \text{ ou } e_i \in B\}$, $i = 1, 2, \dots, n$. O evento reunião é formado pelos pontos amostrais que pertencem a pelo menos a um dos eventos.

b) INTERSECÇÃO

Se $A \cap B = \{e_i \in \Omega | e_i \in A \text{ e } e_i \in B\}$, $i = 1, \dots, n$. O evento intersecção é formado pelos pontos amostrais que pertencem simultaneamente aos eventos A e B.

Obs. Se $A \cap B = \phi$, A e B são eventos mutuamente exclusivos.

c) COMPLEMENTAÇÃO

$$\text{Se } \Omega - A = \bar{A} = \{e_i \in \Omega | e_i \notin A\}$$

Exercício 34. Lança-se um dado. Sejam A: saída de uma face par e B: saída de uma face menor que 4. Determine os eventos:

a) $A \cup B$

b) $A \cap B$

c) \bar{A}

d) \bar{B}

e) $\overline{(A \cup B)}$

f) $(\overline{A \cap B})$

g) $(\overline{A} \cap \overline{B})$

h) $(\overline{A} \cup \overline{B})$

i) $B - A$

j) $A - B$

Exercício 35. *Sejam A, B e C três eventos de um espaço amostral. Expressar os eventos abaixo, usando as operações de reunião, intersecção e complementação.*

- a) somente A ocorrer;
- b) A e C ocorrerem, mas B não;
- c) A, B e C ocorrerem;
- d) pelo menos um ocorrer;
- e) exatamente um ocorrer;
- f) nenhum ocorrer;
- g) exatamente dois ocorrerem;
- h) pelo menos dois ocorrerem;
- i) no máximo dois ocorrerem.

Exercício 36. *Suponha que um conjunto fundamental seja formado pelos inteiros positivos de 1 a 10. Sejam $A = \{2, 3, 4\}$, $B = \{3, 4, 5\}$, $C = \{5, 6, 7\}$. Enumere os elementos dos seguintes conjuntos:*

a) $\overline{A} \cap B$

b) $\overline{A} \cup B$

c) $\overline{\overline{A} \cap \overline{B}}$

d) $\overline{A \cap (\overline{B \cap C})}$

e) $\overline{A \cap (B \cup C)}$

Exercício 37. *Mostre que a desigualdade é verdadeira.*

$$(\overline{A \cap B}) = \overline{A} \cup \overline{B}$$

$$(\overline{A \cup B}) = \overline{A} \cap \overline{B}$$

5.4 Probabilidade

5.4.1 Função de Probabilidade

É a função P que associa a cada evento de F um número real pertencente ao intervalo $[0, 1]$, satisfazendo aos axiomas de Kolmogorov.

1. $P(\Omega) = 1$
2. Para todo evento A , $0 \leq P(A) \leq 1$
3. Se A e B são eventos mutuamente exclusivos, então $P(A \cup B) = P(A) + P(B)$.

5.4.2 Teoremas

- Se ϕ é o conjunto vazio, então $P(\phi) = 0$
- Se A^C é o complemento de um evento A , então $P(A^C) = 1 - P(A)$
- Se $A \subset B$, então $P(A) \leq P(B)$
- Se A e B são dois eventos quaisquer, então $P(A \cap \overline{B}) = P(A) - P(A \cap B)$
- Se A e B são dois eventos quaisquer, então $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

5.4.3 Espaços amostrais equiprováveis

Quando nós associamos a cada ponto amostral a mesma probabilidade, o espaço amostral chama-se equiprovável ou uniforme. Em particular, se Ω contém " n " pontos, então, a probabilidade de cada ponto será $\frac{1}{n}$.

Por outro lado, se um evento A contém " r " pontos, então $P(A) = r \left(\frac{1}{n}\right) = \frac{r}{n}$.

Este método de avaliar $P(A)$ é frequentemente enunciado da seguinte maneira.

$$P(A) = \frac{\text{número de vezes em que o evento } A \text{ pode ocorrer}}{\text{número de vezes em que o Espaço amostral } \Omega \text{ ocorre.}}$$

Exercício 38. Se $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{4}$ e A e B mutuamente exclusivos, calcular:

- a) $P(\overline{A})$
- b) $P(\overline{B})$
- c) $P(A \cap B)$
- d) $P(A \cup B)$
- e) $P(\overline{A \cap B})$

Exercício 39. Se $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{3}$ e $P(A \cap B) = \frac{1}{4}$.

- a) $P(A \cup B)$
- b) $P(\overline{A} \cup \overline{B})$
- c) $P(\overline{A} \cap \overline{B})$

Exercício 40. Considere dois eventos: A e B , mutuamente exclusivos, com $P(A) = 0,3$ e $P(B) = 0,5$. Calcule:

- a) $P(A \cap B)$
- b) $P(A \cup B)$
- c) $P(A|B)$
- d) $P(A^c)$
- e) $P((A \cup B)^c)$

Exercício 41. Sendo $P(A) = x$, $P(B) = y$ e $P(A \cap B) = z$, calcular:

- a) $P(\overline{A} \cup \overline{B})$
- b) $P(\overline{A} \cap \overline{B})$
- c) $P(\overline{A} \cap B)$
- d) $P(\overline{A} \cup B)$

Exercício 42. Se A, B, C são eventos arbitrários, exprima em notação de conjuntos os seguintes eventos:

- a) ocorrem apenas 2;
- b) ocorrem não mais de 2;
- c) ocorrem A e B mas não C ;
- d) ocorre ao menos um;
- e) não ocorre nenhum;
- f) ocorre apenas um.

Exercício 43. Sejam $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $A = \{2, 4, 6, 8\}$, $B = \{1, 3, 5, 7, 9\}$, $C = \{2, 3, 4, 5\}$ e $D = \{1, 6, 7\}$. Encontre:

- a) $A \cup B$

- b) $A \cap B$
- c) \bar{C}
- d) $(C \cap D) \cap \bar{B}$
- e) $A \cap C \cap \bar{D}$

Exercício 44. Sendo $P(A) = x$, $P(B) = y$ e $P(A \cap B) = z$, Calcular:

- a) $P(\bar{A} \cup \bar{B})$
- b) $P(\bar{A} \cap \bar{B})$
- c) $P(\bar{A} \cap B)$
- d) $P(\bar{A} \cup B)$

Exercício 45. Sejam A e B eventos com $P(A) = \frac{3}{8}$, $P(B) = \frac{1}{2}$ e $P(A \cap B) = \frac{1}{4}$. Encontre:

- a) A e B são mutuamente excludentes?
- b) $P(A \cup B)$
- c) $P(\bar{A})$
- d) $P(\bar{B})$
- e) $P(\bar{A} \cap \bar{B})$
- f) $P(\bar{A} \cup \bar{B})$
- g) $P(A \cap \bar{B})$
- h) $P(\bar{A} \cap B)$

Exercício 46. Sejam A e B os eventos com $P(A \cup B) = 3/4$, $P(\bar{A}) = 2/3$ e $P(A \cap B) = 1/4$. Determine:

- a) A e B são mutuamente excludentes?
- b) $P(A)$
- c) $P(B)$
- d) $P(A \cap \bar{B})$

Exercício 47. Sejam A e B eventos tais que: $P(A) = 1/2$; $P(B) = 1/4$ e $P(A \cap B) = 1/5$. Calcule:

- a) A e B são disjuntos?

- b) $P(A \cup B)$
- c) $P(\bar{A})$
- d) $P(\bar{B})$
- e) $P(A \cap \bar{B})$
- f) $P(\bar{A} \cap B)$
- g) $P(\bar{A} \cap \bar{B})$
- h) $P(\bar{A} \cup \bar{B})$

Exercício 48. *Suponha que A e B sejam eventos tais que $P(A) = 2/5$, $P(B) = 2/5$ e $P(A \cup B) = 1/2$. Determine:*

- a) $P(A \cap B)$
- b) $P(\bar{A})$
- c) $P(\bar{B})$
- d) $P(\bar{A} \cap \bar{B})$
- e) $P(\overline{A \cap B})$
- f) $P(\bar{A} \cap B)$
- g) $P(\bar{A} \cup B)$
- i) A e B são disjuntos ?

Exercício 49. *Se $P(A) = 1/3$, $P(A \cup B) = 1/2$ e $P(A \cap B) = 1/4$, determine $P(B)$*

Exercício 50. *Se A e B são eventos disjuntos e $P(A) = 0,5$ e $P(A \cup B) = 0,6$, encontre $P(B)$*

Exercício 51. *Uma escola do ensino médio do interior de Minas Gerais tem 40% de estudantes do sexo masculino. Entre estes, 20% nunca viram o mar, ao passo que, entre as meninas, essa porcentagem é de 50%. Qual a probabilidade de que um aluno selecionado ao acaso seja:*

- a) Do sexo masculino e nunca tenha visto o mar?
- b) Do sexo feminino ou nunca tenha visto o mar?

Exercício 52. *Sendo A e B dois eventos em um mesmo espaço amostral "traduza" para a linguagem da teoria dos conjuntos, as seguintes situações:*

- a) Pelo menos um dos eventos ocorre;
- b) O evento A ocorre mas B não;
- c) Nenhum deles ocorre;
- d) Exatamente um dos eventos ocorre.

Exercício 53. *Uma universidade tem 10 mil alunos dos quais 4 mil são considerados esportistas. Temos ainda que 500 alunos são do curso de Biologia diurno, 700 da biologia noturno, 100 são esportistas e da biologia diurno e 200 são esportista e da biologia noturno. Um aluno é escolhido ao acaso e pergunta-se a probabilidade de:*

- a) Ser esportista;
- b) Ser esportista e aluno da Biologia Noturno;
- c) Não ser da Biologia;
- d) Ser esportista ou aluno da Biologia;
- e) Não ser esportista e nem aluno da Biologia.

Exercício 54. *Dois processadores tipos A e B são colocados em teste por 50 mil horas. A probabilidade de que um erro de cálculo acontecer em um processador do tipo A é de $1/30$, no tipo B, $1/80$ e em ambos, $1/1000$. Qual a probabilidade de que:*

- a) Pelo menos um dos processadores tenha apresentado erro?
- b) Nenhum processador tenha apresentado erro?
- c) Apenas o processador A tenha apresentado erro?

Exercício 55. *Sejam A e B dois eventos em um dado espaço amostral, tais que $P(A) = 0,2$, $P(B) = p$, $P(A \cup B) = 0,5$ e $P(A \cap B) = 0,1$. Determine o valor de p.*

Exercício 56. *Consideremos um experimento aleatório e os eventos A e B associados, tais que $P(A) = 1/2$, $P(B) = 1/3$ e $P(A \cap B) = 1/4$. Então, temos:*

- a) $P(\bar{A})$
- b) $P(\bar{B})$
- c) $P(A \cup B)$
- d) $P(\bar{A} \cap \bar{B})$
- e) $P(\bar{A} \cup \bar{B})$

f) $P(\bar{A} \cap B)$

Exercício 57. De 300 estudantes de administração, 100 estão matriculados em Contabilidade e 80 em Estatística. Estes dados incluem 30 que estão matriculados em ambas as disciplinas. Qual a probabilidade de que um estudante aleatoriamente escolhido esteja matriculado em Contabilidade (A) ou em Estatística (B)?

Exercício 58. De 100 pessoas que solicitaram emprego de programador de computadores, durante o ano passado, em uma grande empresa, 40 possuíam experiência anterior (W) e 30 possuíam um certificado profissional (C). Vinte dos candidatos possuíam tanto experiência anterior como certificado profissional e foram incluídos nas contagens dos dois grupos.

- a) Qual a probabilidade de que um candidato aleatoriamente escolhido tenha experiência ou certificado (ou ambos)?
- b) Qual a probabilidade de que um candidato aleatoriamente escolhido tenha experiência ou certificado, mas não ambos?

Exercício 59. Discos de plástico de policarbonato, provenientes de um fornecedor, são analisados com relação às resistências a arranhões e a choques. Os resultados de 100 discos são resumidos abaixo:

Resistência a arranhão	Resistência a choques	
	Alta	Baixa
Alta	80	9
Baixa	6	5

Faça A denotar o evento em que um disco tenha alta resistência a choque e faça B denotar o evento em que um disco tenha alta resistência a arranhões. Se um disco for selecionado aleatoriamente, determine as seguintes probabilidades:

- a) $P(A)$
- b) $P(B)$
- c) $P(\bar{A})$
- d) $P(A \cap B)$
- e) $P(A \cup B)$
- f) $P(\bar{A} \cap B)$

Exercício 60. Amostras de uma peça de alumínio fundido são classificadas com base no acabamento (em micropologadas) da superfície e nas medidas de comprimento. Os resultados de 100 peças são resumidos a seguir:

Acabamento da Superfície	Comprimento	
	excelente	bom
excelente	75	7
bom	10	8

Faça A denotar o evento em que uma amostra tenha excelente acabamento na superfície e faça B denotar o evento em que uma amostra tenha excelente comprimento. Se uma amostra for selecionada ao acaso, determine as seguintes probabilidades:

- $P(A)$
- $P(B)$
- $P(\bar{A})$
- $P(A \cap B)$
- $P(A \cup B)$
- $P(\bar{A} \cap B)$

Exercício 61. Amostras de uma espuma, provenientes de três fornecedores são classificados com relação a satisfazer ou não as especificações. Os resultados de 100 amostras são resumidas a seguir:

Fornecedor	Obdece	
	SIM	NÃO
1	18	2
2	17	3
3	50	10

Faça A denotar o evento em que uma amostra seja proveniente do fornecedor 1 e faça B denotar o evento em que uma amostra atenda às especificações. Se uma amostra de espuma for selecionada ao acaso, determine as seguintes probabilidades:

- $P(A)$
- $P(B)$
- $P(\bar{A})$

- d) $P(A \cap B)$
- e) $P(A \cup B)$
- f) $P(\bar{A} \cap B)$
- g) $P(A \cap \bar{B})$

Exercício 62. Se $P(A) = 0,3$, $P(B) = 0,2$ e $P(A \cap B) = 0,1$. Determine as seguintes probabilidades.

- a) A e B são disjuntos?
- b) $P(\bar{A})$
- c) $P(A \cup B)$
- d) $P(\bar{A} \cap B)$
- e) $P(A \cap \bar{B})$
- f) $P(\overline{A \cup B})$
- h) $P(\bar{A} \cup B)$

Exercício 63. Se A , B e C forem eventos mutuamente excludentes, com $P(A) = 0,2$, $P(B) = 0,3$ e $P(C) = 0,4$, determine as seguintes probabilidades:

- a) $P(A \cup B \cup C)$
- b) $P(A \cap B \cap C)$
- c) $P(A \cap B)$
- d) $P[(A \cup B) \cap C]$

Exercício 64. Um rebanho de cem bovinos está formado por 52 Hereford, 27 Angus, dez Shorthorn e os demais Zebu. Escolhido ao acaso um bovino do rebanho, qual é a probabilidade de que seja Hereford ou Angus?

5.4.4 Probabilidade Condicional

Veremos a noção de probabilidade condicional através do seguinte exemplo:

Consideremos 250 alunos que cursam o primeiro ciclo de uma faculdade. Destes alunos 100 são homes (H) e 150 são mulheres (M), 110 cursam matemática financeira (F) e 140 cursam direito trabalhista (Q). A distribuição dos alunos é a seguinte:

Sexo/Disciplina	F	Q	Total
H	40	60	100
M	70	80	150
Total	110	140	250

Um aluno é sorteado ao acaso. Qual a probabilidade de que esteja cursando direito trabalhista, dado que é mulher?

Pelo quadro vemos que esta probabilidade é $\frac{80}{150}$ e representamos:

$P(Q/M) = \frac{80}{150}$ (probabilidade de que o aluno curse direito trabalhista, condicionado ao fato de ser mulher).

Observamos porém, que $P(M \cap Q) = \frac{80}{250}$ e $P(M) = \frac{150}{250}$. Para obtermos o resultado do problema basta considerar que

$$P(Q/M) = \frac{\frac{80}{250}}{\frac{150}{250}} = \frac{80}{150}$$

logo

$$P(Q/M) = \frac{P(M \cap Q)}{P(M)}$$

Sejam $A \subset \Omega$ e $B \subset \Omega$. Definimos Probabilidade Condicional de A dado que B ocorre (A/B) como segue:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \text{ se } P(B) \neq 0.$$

Também:

$$P(B/A) = \frac{P(B \cap A)}{P(A)}, \text{ se } P(A) \neq 0.$$

Exemplo 5.1. Sendo $P(A) = \frac{1}{3}$, $P(B) = \frac{3}{4}$ e $P(A \cup B) = \frac{11}{12}$, calcular $P(A/B)$.

Solução: Como $P(A/B) = \frac{P(A \cap B)}{P(B)}$, devemos calcular $P(A \cap B)$.

Como $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, temos

$$\frac{11}{12} = \frac{1}{3} + \frac{3}{4} - P(A \cap B) \therefore P(A \cap B) = \frac{2}{12} = \frac{1}{6}$$

$$\text{Logo } P(A/B) = \frac{1/6}{3/4} = \frac{2}{9}$$

Tiramos da definição da probabilidade condicional, o chamado TEOREMA DO PRODUTO: Sejam $A \subset \Omega$ e $B \subset \Omega$. Então $P(A \cap B) = P(B) \cdot P(A/B)$ ou $P(A \cap B) = P(A) \cdot P(B/A)$.

Teorema do Produto

A partir da definição de probabilidade condicional, poderemos enunciar o teorema do produto:

”A probabilidade da ocorrência simultânea do dois eventos, A e B , do mesmo espaço amostral, é igual ao produto da probabilidade de um deles pela probabilidade condicional do outro, dado o primeiro.

Assim:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \implies P(A \cap B) = P(B)P(A/B)$$

ou

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \implies P(A \cap B) = P(A)P(B/A)$$

Exemplo 5.2. *Em um lote de 12 peças, 4 são defeituosas, 2 peças são retiradas uma após a outra sem reposição. Qual a probabilidade de que ambas sejam boas?*

Solução: $A = \{\text{a primeira peça é boa}\}$

$B = \{\text{a segunda peça é boa}\}$

$$P(A \cap B) = P(A)P(B/A) = \frac{8}{12} \cdot \frac{7}{11} = \frac{14}{33}$$

Independência Estatística

Um evento A é considerado independente de um outro evento B se a probabilidade de A é igual à probabilidade condicional de A dado b , isto é, se

$$P(A) = P(A/B)$$

É evidente que, se A é independente de b , B é independente de A ; assim:

$$P(B) = P(B/A) \tag{5.1}$$

Considerando o teorema do produto, poderemos afirmar que se A e B são independentes, então:

$$P(A \cap B) = P(A) \cdot P(B) \tag{5.2}$$

5.5 Exercícios

Exercício 65. *Seja $\Omega = \{1, 2, 3, 4\}$ um espaço amostral equiprovável e $A = \{1, 2\}$; $B = \{1, 3\}$; $C = \{1, 4\}$ três eventos de Ω . Verificar se os eventos A, B e C são independentes.*

Exercício 66. *Sejam A e B eventos tais que $P(A) = 0,2$, $P(B) = p$, $P(A \cup B) = 0,6$. Calcular p considerando A e B :*

a) mutuamente exclusivos;

b) independentes.

Exercício 67. *Se $P(A \cup B) = 0,8$; $P(A) = 0,5$ e $P(B) = x$, determine o valor de x no caso de:*

- a) A e B serem mutuamente exclusivos;
- b) A e B serem independentes;

Exercício 68. *Verifique se são válidas as afirmações:*

- a) Se $P(A) = 1/3$ e $P(B/A) = 3/5$ então A e B não podem ser disjuntos;
- b) Se $P(A) = 1/2$, $P(B/A) = 1$ e $P(A/B) = 1/2$ então A não pode estar contido em B.

Exercício 69. *A probabilidade de que um homem esteja vivo daqui a 30 anos é $\frac{2}{5}$; a de sua mulher é de $\frac{2}{3}$. Determinar a probabilidade de que daqui a 30 anos:*

- a) ambos estejam vivos;
- b) somente o homem esteja vivo;
- c) somente a mulher esteja viva;
- d) nenhum esteja vivo;
- e) pelo menos um esteja vivo.

Exercício 70. *Se $P(B) = 0,4$; $P(A) = 0,7$ e $P(A \cap B) = 0,3$; Calcule $P(A|B^C)$*

Exercício 71. *Comente a afirmação: se dois eventos são mutuamente exclusivos então eles não são independentes.*

Exercício 72. *O Atlético Mineiro ganha com probabilidade 0,7 se chove e com 0,8 se não chove. Em Setembro a probabilidade de chuva é de 0,3. O Atlético ganhou uma partida em Setembro, qual a probabilidade de ter chovido nesse dia?*

Exercício 73. *Uma classe de estatística teve a seguinte distribuição das notas finais: 4 do sexo masculino e 6 do feminino foram reprovados, 8 do sexo masculino e 14 do feminino foram aprovados. Para um aluno sorteado dessa classe, denote por M se o aluno escolhido for do sexo masculino e por A se o aluno foi aprovado.*

Calcule:

- a) $P(A \cup M^C)$
- b) $P(A^C \cap M^C)$
- c) $P(A|M)$
- d) $P(M^C|A)$
- e) $P(M|A)$

Exercício 74. *Numa cidade do interior de Minas Gerais, estima-se que cerca de 20% dos habitantes têm algum tipo de alergia. Sabe-se que 50% dos alérgicos praticam esporte, enquanto que essa porcentagem entre os não alérgicos é de 40%. Para um indivíduo escolhido aleatoriamente nessa cidade, obtenha a probabilidade de:*

- a) Não praticar esporte.
- b) Ser alérgico dado que não pratica esportes.

Exercício 75. *As preferências de homens e mulheres por cada gênero de filme alugado em uma locadora de vídeos, estão apresentadas na próxima tabela.*

Sexo/filme	Comédia	Romance	Policial
Homens	136	92	248
Mulheres	102	195	62

Sorteando-se ao acaso uma dessas locações de vídeo, pergunta-se a probabilidade de:

- a) Uma mulher ter alugado um filme policial?
- b) O filme alugado ser uma comédia?
- c) Um homem ter alugado ou o filme ser um romance?
- d) O filme ser policial dado que foi alugado por um homem?

Exercício 76. *Dois armários guardam as bolas de voleibol e basquete. O armário 1 tem 3 bolas de voleibol e 1 de basquete, enquanto o armário 2 tem 3 bolas de voleibol e 2 de basquete. Escolhendo-se ao acaso um armário e, em seguida, uma de suas bolas, calcule a probabilidade dela ser:*

- a) De voleibol, sabendo-se que o armário 1 foi escolhido.
- b) De basquete, sabendo-se que o armário 2 foi escolhido.
- c) De basquete.

Exercício 77. *Numa bolsa temos 5 moedas de Cr\$ e 4 de Cr\$ 0,50. Qual a probabilidade de, ao retirarmos duas moedas, obtermos Cr\$ 1,50?*

Exercício 78. *Uma urna contém 5 bolas pretas, três vermelhas e duas brancas. Foram extraídas 3 bolas com reposição. Qual a probabilidade de terem sido duas bolas pretas e uma vermelha?*

Exercício 79. *Uma urna contém 5 bolas brancas e 6 pretas. Três bolas são retiradas. Calcule a probabilidade de:*

- a) Todas pretas;
 b) exatamente um branca;
 c) ao menos uma preta.

Exercício 80. *A urna n.1 contém: 1 bola vermelha e 2 brancas. A urna n.2 contém: 2 bolas vermelhas e 1 branca. Tiramos aleatoriamente uma bola da urna n.1, colocamos na urna 2 e misturamos. Em seguida tiramos aleatoriamente uma bola da urna n.2. Qual é a probabilidade de tirarmos uma bola branca da urna n.2?*

5.6 Teorema de Bayes

Teorema da Probabilidade Total

”Sejam A_1, A_2, \dots, A_n eventos que formam uma partição do espaço amostral. Seja B um evento desse espaço. Então

$$P(B) = \sum P(A_i) \cdot P(B/A_i).$$

Os eventos $(B \cap A_i)$ e $(B \cap A_j)$, para $i \neq j$, $i = 1, 2, \dots, n$ e $j = 1, 2, 3, \dots, n$ são mutuamente exclusivos, pois

$$(B \cap A_i) \cap (B \cap A_j) = B \cap (A_i \cap A_j) = B \cap \phi = \phi$$

O evento B ocorre como segue:

$$B = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup \dots \cup (B \cap A_n) \therefore \\ \therefore P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) + \dots + P(B \cap A_n)$$

e usando o teorema do produto vem:

$$P(B) = P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + \dots + P(A_n) \cdot P(B/A_n)$$

$$\text{ou } P(B) = \sum P(A_i) P(B/A_i)$$

Teorema de Bayes

Sejam A_1, A_2, \dots, A_n eventos que formam uma partição do Ω . Seja $B \subset \Omega$. Sejam conhecidas $P(A_i)$ e $P(B/A_i)$, $i = 1, 2, \dots, n$.

$$P(A_j/B) = \frac{P(A_j)P(B/A_j)}{\sum P(A_i)P(B/A_i)}$$

Exercício 81. *Três candidatos disputam as eleições para o governo do Estado. O candidato do partido de direita tem 30% de preferência eleitoral, o de centro tem 30% e o da esquerda 40%. Em sendo eleito, a probabilidade de dar efetivamente prioridade para educação e saúde é de 0,4; 0,6 e 0,9 para os candidatos de direita, centro e esquerda, respectivamente.*

- a) Qual é a probabilidade de não ser dada prioridade a essas áreas no próximo governo?

- b) Se a área teve prioridade, qual a probabilidade do candidato de direita ter ganho a eleição?

Exercício 82. *Três máquinas A, B e C, produzem 50%, 30% e 20%, respectivamente, do total de peças de uma fábrica. As percentagens de produção defeituosa destas máquinas são 3%, 4% e 5%. Se uma peça é selecionada aleatoriamente, ache a probabilidade de ela ser defeituosa.*

Exercício 83. *Considere a fábrica do exercício anterior. Suponha que uma peça, selecionada aleatoriamente, seja considerada defeituosa. Encontre a probabilidade de ela ter sido produzida pela máquina A; ou seja, encontre a $P(A/X)$. Pelo teorema de Bayes.*

Exercício 84. *Três fábricas fornecem equipamentos de precisão para o laboratório de química de uma universidade. Apesar de serem aparelhos de precisão, existe uma pequena chance de subestimação ou superestimação das medias efetuadas. A tabela a seguir apresenta o comportamento do equipamento produzido em cada fábrica.*

Fabrica I	Subestima	Exata	Superestima
Probabilidade	0,01	0,98	0,01
Fabrica II	Subestima	Exata	Superestima
Probabilidade	0,005	0,98	0,015
Fabrica III	Subestima	Exata	Superestima
Probabilidade	0,00	0,99	0,01

As fábricas *I, II e III* fornecem, respectivamente, 20%, 30% e 50% dos aparelhos utilizados, Escolhemos, ao acaso, um desses aparelhos e perguntamos a probabilidade de:

- Haver superestimação de medidas?
- Não haver subestimação das medidas efetuadas?
- Dando medidas exatas, ter sido fabricado em *III*?
- Ter sido produzido por *I*, dado que não subestima as medidas?

5.7 Referências Bibliográficas

1. CARVALHO, Sérgio. Estatística Básica - Teoria e 150 questões. Editora IMPETUS, 2004
2. FURTADO, Daniel Ferreira. Estatística Básica. Editora UFLA. 2005
3. LEVINE, David M. Estatística: Teoria e Aplicações usando o microsoft excel em português. Editora LTC, 2003.
4. MAGALHÃES, Marcos Nascimento. Noções de Probabilidade e Estatística / São Paulo, 4ª edição, Editora USP, 2002