

ELSA NEGAS

Com a colaboração
ADELAIDE CARREIRA

ESTATÍSTICA DESCRITIVA

Explicação Teórica, Casos de Aplicações
e Exercícios Resolvidos



EDIÇÕES SÍLABO

ESTATÍSTICA DESCRITIVA

**Explicação Teórica, Casos de Aplicações
e Exercícios Resolvidos**

ELSA NEGAS

Com a colaboração

ADELAIDE CARREIRA

Doutora em Ciências – Área Matemática

EDIÇÕES SÍLABO

É expressamente proibido reproduzir, no todo ou em parte, sob qualquer forma ou meio gráfico, eletrónico ou mecânico, inclusive fotocópia, este livro.

As transgressões serão passíveis das penalizações previstas na legislação em vigor.

Não participe ou encoraje a pirataria eletrónica de materiais protegidos.

O seu apoio aos direitos dos autores será apreciado.

Visite a Sílabo na rede

www.silabo.pt

FICHA TÉCNICA:

Título: Estatística Descritiva – Explicação Teórica, Casos de Aplicações e Exercícios Resolvidos

Autora: Elsa Negas

© Edições Sílabo, Lda.

Capa: Pedro Mota

Imagem da capa: Monsit Jangariyawong | Dreamstime.com

1ª Edição – Lisboa, setembro de 2019

Impressão e acabamentos: Europress, Lda.

Depósito Legal: 459011/19

ISBN: 978-989-561-019-8



EDIÇÕES SÍLABO, Lda.

Publicamos conhecimento

Editor: Manuel Robalo

R. Cidade de Manchester, 2

1170-100 Lisboa

Tel.: 218130345

e-mail: silabo@silabo.pt

www.silabo.pt

Índice

Prefácio	9
-----------------	---

Capítulo 1

Recolha e classificação de dados

1. Introdução	13
2. Universo e amostra	14
3. Variáveis estatísticas – classificação	20
4. Escala de Likert	24
5. Organização dos dados em tabelas e gráficos	31
5.1. Variáveis qualitativas	34
5.2. Variáveis quantitativas discretas	39
5.3. Variáveis quantitativas contínuas	51
5.4. Organizar os dados em classes com amplitudes diferentes	65
6. Exercícios resolvidos	69
7. Enunciados de exames para autoavaliação	80

Capítulo 2

Medidas de localização, de dispersão e de concentração

1. Introdução	93
2. Medidas de localização	94
2.1. Medidas de localização central	95
2.1.1. Moda	95
2.1.2. Mediana	113
2.1.3. Média	125

2.2. Medidas de localização não central	147
2.2.1. Quartis	148
2.2.2. Decis, percentis	174
3. Medidas de dispersão	185
3.1. Amplitude ou intervalo de variação	191
3.2. Amplitude interquartil	195
3.3. Diagrama de extremos e quartis	198
3.4. <i>Outliers</i>	217
3.5. Diagrama de extremos e quartis quando existem <i>outliers</i>	223
3.6. Média aparada	225
3.7. Desvios	229
3.7.1. Desvio absoluto médio	232
3.7.2. Desvio quadrático médio	243
3.8. Variância	247
3.9. Desvio padrão	252
3.10. Coeficiente de variação	259
4. Medidas de assimetria	266
4.1. Distribuição simétrica	266
4.2. Distribuição assimétrica	271
4.2.1. Classificação da assimetria	274
4.2.2. Caso de estudo 1	281
4.2.3. Caso de estudo 2	283
4.2.4. Caso de estudo 3	286
5. Curtose	288
6. Regras de normalidade – <i>Score Z</i>	297
7. Medidas de concentração	302
7.1. Introdução	302
7.2. Curva de Lorenz	303
7.3. Cálculo da área de concentração	309
7.4. Índice de Gini	322
8. Exercícios resolvidos	329
9. Enunciados de exames para autoavaliação	342

Anexos

Anexo 1 – Questionário de avaliação do grau de satisfação dos clientes	363
Anexo 2 – Cálculo da moda para dados agrupados em classes	364
Anexo 3 – Cálculo da mediana para dados agrupados em classes	365
Anexo 4 – Média geométrica e média harmónica	368
Anexo 5 – Processo abreviado para o cálculo das médias	373
Anexo 6 – Média das médias	379

Bibliografia	381
---------------------	------------

Prefácio

Neste livro proponho uma exposição teórica dos conceitos básicos da Análise de Dados, acompanhada de exemplos ilustrativos e exercícios resolvidos e propostos. É meu propósito com este manual colocar à disposição dos interessados um instrumento prático de consulta e de apoio ao estudo autônomo, onde se apresentam de forma clara e concisa os conceitos, se enunciam e justificam propriedades e se deduzem fórmulas de cálculo em Estatística Descritiva.

A ideia de elaborar um manual com estas características partiu das muitas solicitações dos meus alunos, conjugada com a minha experiência no ensino e apoio ao ensino superior nas disciplinas de Estatística nas áreas de Ciências Humanas e de Ciências Exatas e no trabalho com alunos do ensino a distância. Procuo, nas páginas seguintes, responder às necessidades manifestadas e, de um modo conciso e claro, proporcionar aos leitores um instrumento de trabalho de apoio e acompanhamento ao estudo autônomo ou integrado ao nível das diferentes licenciaturas existentes e onde se aborde a Estatística nos seus planos curriculares. Professores do ensino secundário também poderão encontrar aqui elementos importantes para se capacitarem para lecionar de modo adequado as disciplinas cujos currículos contenham noções e princípios de Estatística Descritiva.

Partindo de exemplos práticos, que na sua maior parte descrevem circunstâncias da vida real, introduzem-se os conceitos teóricos e chama-se a atenção para alguns pontos em particular que, pela sua dificuldade ou pela sua importância, necessitam de um maior esclarecimento. Em cada capítulo, são propostos dois tipos de exercícios de aplicação: exercícios resolvidos e exercícios de autoavaliação. Nos exercícios resolvidos é apresentada a resolução detalhada, alguns com notas de esclarecimento, quando tal é considerado útil para a compreensão da resolução. Nos exercícios de autoavaliação, o leitor é convidado a desenvolver autonomamente a resolução e a aferir os conhecimentos adquiridos, podendo confrontar os resultados com os da solução.

A abordagem a que recorri é rigorosa sem ser exaustiva no seu tratamento matemático. Esta opção foi feita para que os conteúdos possam ser facilmente entendidos, mesmo para os leitores que não tenham uma preparação matemática mais avançada.

Faço votos para que os leitores deste manual possam confirmar o alcance dos propósitos que me levaram a escrevê-lo. Tal dar-me-ia muita satisfação e seria sinal que este livro cumpre a sua missão.

A autora

Capítulo 1

Recolha e classificação de dados

1. Introdução

A **estatística**, que se ocupa da obtenção de informação «válida», da sua organização e do seu tratamento com o objetivo de definir estratégias, orientar na tomada de decisão e fundamentar previsões em contexto de incerteza, exige uma cuidadosa recolha de informação.

Um processo estatístico começa pela identificação de um contexto e do estudo que se pretende realizar seguido da recolha, da descrição e da análise de um conjunto de observações.

Esta fase do processo é o objetivo da **Estatística Descritiva** ou **Análise Exploratória de Dados**. Como resultado desta fase é possível fazer inferências sobre uma população da qual apenas se analisou uma parte – a amostra – etapa esta que é designada por **Inferência Estatística** ou **Análise Inferencial**. A Inferência Estatística, com recurso a diversas técnicas, permite, então, perspetivar um eficaz conhecimento da população, com base no tratamento da amostra. A Estatística Descritiva utiliza processos numéricos e construção de gráficos para estabelecer padrões num conjunto de dados, para sintetizar informação revelada por esses dados e apresentar formas convenientes de explicitar as relações entre eles.

A fase inicial de um processo estatístico envolve as seguintes etapas:

- Identificação da população.
- Identificação do estudo pretendido.
- Identificação de uma amostra aleatória.¹
- Recolha de dados.
- Validação dos dados.
- Tratamento e organização dos dados.
- Análise e interpretação dos dados.

Neste capítulo analisam-se as quatro primeiras etapas.

Ao identificar o estudo estatístico que se pretende realizar é importante definir a especificidade do estudo pretendido, a que se chama **variável estatística**. Após o qual se deve identificar a população onde o estudo vai decorrer.

Posteriormente, identifica-se a forma de recolher uma amostra. Sob vários aspetos esta é uma etapa importante pois a amostra tem de representar «bem» a população e ter uma dimensão adequada para que a recolha dos dados seja compatível com os recursos disponíveis – humanos, técnicos e financeiros – e, ainda, que permita o desenvolvimento do estudo em tempo útil.

(1) Entende-se por amostra aleatória aquela em que a escolha de cada elemento é feita ao acaso sem qualquer interferência intrínseca ou extrínseca à própria recolha. No contexto deste texto considera-se a aleatoriedade na recolha dos elementos da amostra.

Sendo a representatividade da população uma das principais características de uma boa amostra, a metodologia utilizada na sua constituição é, naturalmente, relevante para que, a par das análises estatísticas posteriormente analisadas, viabilizar as conclusões que vierem a ser feitas.

2. Universo e amostra

O **universo** ou **população** é o conjunto de indivíduos (pessoas, objetos, empresas, equipamentos, entre outros) com uma ou mais características comuns, que variam em quantidade e/ou em qualidade, e que se pretende analisar.

Neste texto distingue-se a **população de estudo** – que corresponde a um universo global para a análise em investigação – da **população alvo** – que corresponde à população sobre a qual a investigação vai incidir de uma forma direta e de onde se vai extrair, eventualmente uma amostra – e, ainda, **população inquirida** – o conjunto dos indivíduos que vão ser observados no âmbito do estudo.

População é o conjunto dos elementos que traduzem a(s) característica(s) comum a todos os objetos sobre os quais incide a análise em causa.

EXERCÍCIO 1.1

Com o objetivo de decidir sobre as vantagens de implementar, ou não, uma campanha publicitária sobre pacotes de férias em cruzeiros para clientes seniores, uma agência de turismo analisa a sua base de dados de clientes nos últimos cinco anos. O objetivo é apurar a percentagem de clientes com mais de 65 anos que fizeram férias em cruzeiro. Qual a população de estudo e qual a população alvo?

Resolução

Neste caso a população de estudo é o conjunto de clientes da agência e a população alvo, que aqui consideramos ser também a população inquirida, é o conjunto de clientes da agência com mais de 65 anos.

EXERCÍCIO 1.2

Num centro de saúde pretende-se avaliar a aceitação por parte dos cidadãos inscritos nesse centro de uma possível alteração no horário das consultas de pediatria. Qual a população de estudo e qual a população inquirida?

Resolução

A população de estudo é constituída por todos os inscritos nesse centro de saúde, a população alvo é constituída por todos os potenciais interessados nas consultas de pediatria – por exemplo, as crianças inscritas com menos de 12 anos – e a população inquirida é o conjunto de todos os pais das crianças com menos de 12 anos que sejam contactados.

Uma população pode ser **finita** ou **não finita** para uma determinada análise estatística, consoante o número de indivíduos que a compõem (**elementos**) permitir que todos sejam observados, ou não. Em alguns estudos estatísticos, a natureza da análise a fazer não permite que todos os elementos da população, mesmo que em número finito, sejam observados. Consideramos, no primeiro caso, que a população não é finita para o estudo que está a ser realizado e, no segundo caso, que se trata de uma população finita, dizendo, então, que se tem um **censo** ou um **recenseamento** dessa população. Uma vantagem indiscutível dos censos é que os seus resultados são exatos.

EXERCÍCIO 1.3

Suponhamos que uma companhia de seguros pretende classificar os seus 10 000 clientes com seguro automóvel em seis níveis etários: com 21 anos ou menos, entre 21 anos e 30 anos (inclusive), entre 30 anos e 40 anos (inclusive), entre 40 anos e 50 anos (inclusive), entre 50 anos e 60 anos (inclusive) e com mais de 60 anos. Pretende ainda analisar, dentro desta carteira de clientes, o grau de satisfação destes em relação ao atendimento quando contactam a companhia. Utilizando para tal um inquérito que será realizado, por via telefónica, pelos funcionários da companhia. Classifique a população para cada um dos estudos em finita ou infinita?

Resolução

Note-se que no primeiro estudo, todos os clientes do ramo automóvel são analisados e classificados pelo que a população considera-se finita. No segundo estudo, atendendo ao número total de clientes e à natureza do próprio estudo, é aconselhável trabalhar sobre uma amostra, considerando-se a população não finita.

Identificar o universo ou população de acordo com o estudo que se pretende efetuar é, por vezes, um processo não trivial pois os elementos conhecidos para a elaboração do próprio estudo podem ser demasiado abrangentes e generalistas. Atenda-se aos exemplos seguintes.

Exemplo 1.1

Suponha-se que se pretende efetuar um estudo sobre a literacia em Portugal. A primeira questão que se põe é saber qual deve ser a variável a definir? Naturalmente, as habilitações académicas de cada português. Quais as dificuldades?

Resolução

Uma primeira dificuldade surge ao se estabelecer um limite inferior para essa escolaridade pois a escolaridade mínima obrigatória sofreu alterações ao longo do tempo. Uma pessoa mais velha, que pode ser incluída no estudo, poderá ter como escolaridade obrigatória o 4º ano e outra pode ter como referência a escolaridade obrigatória no 9º ano ou no 12º ano. Se não se tiver cuidado na definição prévia da variável o conhecimento recolhido pode ser pouco esclarecedor.

Também a definição de população apresenta as suas dificuldades como, por exemplo, qual deve ser a idade mínima pretendida para uma pessoa responder ao inquérito. Inclui-se os que residem em Portugal no momento do estudo ou os portugueses em geral mesmo que temporariamente estejam ausentes do país?

Exemplo 1.2

Identifique algumas dificuldades se se pretender efetuar um estudo sobre a obesidade infantil.

Resolução

Deve-se começar por decidir as idades abrangidas. Se o objetivo recair sobre crianças dos 6 aos 9 anos, supostamente a frequentar o 1º ciclo do ensino básico os meios envolvidos são uns, mas se se incluir no estudo crianças de uma faixa etária mais alargada, por exemplo, dos 4 aos 12 anos os recursos envolvidos são muitos mais e a articulação entre instituições é muito maior. Definida a população é importante identificar a, ou as, variáveis em estudo: se a resposta é simplesmente sim ou não e no caso de se haver obesidade, se se pretendem analisar fatores que possam contribuir para o estudo mais detalhado do problema da obesidade infantil como, por exemplo, a hereditariedade, os hábitos alimentares, o facto da criança poder ser portadora de doenças, etc. Neste caso a complexidade está associada à definição da população e da(s) variável(is) em estudo.

Uma vez identificada a população importa definir a necessidade de trabalhar, ou não, uma amostra. Sendo importante caso afirmativo, definir processos de amostragem que permitem recolher elementos representativos dessa população.

Define-se **unidade de análise** ou **unidade estatística** ou **unidade observacional** como sendo a menor parte distinta e identificável da população para fins seleção da amostra. O resultado da observação da variável sobre cada indivíduo é o **dado estatístico** ou, mais simplesmente, **dado**.

O elemento da população é a unidade de análise estatística, isto é, o indivíduo em relação ao qual se pretende obter uma determinada informação.

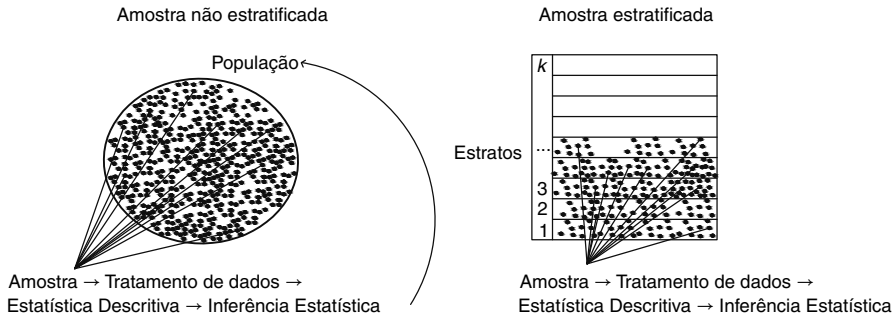
EXERCÍCIO 1.4

Num aglomerado de escolas, com cerca de 2 000 alunos com idades entre os 5 anos e os 20 anos, pretende-se fazer um estudo para perceber se esses alunos visitam regularmente museus. Dado o objetivo do estudo, um processo natural para o desenvolver é não inquirir todos os alunos mas apenas alguns deles, isto é, uma amostra. A população não é homogénea atendendo à diferença de idades contempladas pelo que a amostra deverá refletir a estratificação da população, isto é, por exemplo, envolver alunos de várias idades, de ambos os sexos e, eventualmente, de diferentes níveis socioeconómicos. Qual a unidade estatística?

Resolução

A unidade estatística é cada aluno do aglomerado de escolas considerado.

Quando a população tem características muito genéricas e abrangentes e é constituída por vários subconjuntos com características específicas que constituem estratos da população, é importante garantir que todos eles estão representados na amostra o que obriga a recolher elementos de cada um deles. Neste caso a amostra também é conhecida por diferentes **níveis** ou **estratos** e diz-se por isso, uma **amostra estratificada**.

Figura 2. Tipos de amostra**EXERCÍCIO 1.5**

Dê um exemplo de uma boa amostra (representativa da população em estudo) e dê um exemplo de uma amostra enviesada.

Resolução

Um exemplo de uma boa amostra é: pretende-se fazer um estudo de mercado para determinada marca nacional sobre um produto conhecido e de uso corrente. Com este objetivo foram realizados telefonemas, em que os números de telefone eram selecionados aleatoriamente, mas a amostra foi estratificada ou seja foram questionados todos os distritos em quantidade proporcional à população residente num total de 2 000 questionários.

Um exemplo de uma má amostra é: pretende-se conhecer a opinião dos turistas que visitam Portugal sobre três questões fundamentais; acolhimento, qualidade dos serviços prestados e gastronomia. Para isso foram recolhidos ao acaso 1 500 questionários na zona do Algarve, nos três meses de verão.

EXERCÍCIO 1.6

Quando deve recorrer a uma amostra simples ou estratificada? Exemplifique.

Resolução

Para certo tipo de estudos estatísticos há que considerar subconjuntos da população (amostras). Para alguns deles importa evidenciar vários tipos de características existentes em grupos diferenciados da população (estratos), então, o estudo só é válido se todos os estratos estiverem representados na amostra; desta forma, surge a amostra estratificada. De realçar que, dentro de cada subconjunto cada elemento tem igual probabilidade de ser selecionado.

Uma empresa, com 1 500 funcionários tem, certamente, elementos na administração, no secretariado, em diferentes setores. Caso se pretenda fazer um estudo que abranja todos os funcionários deverá ser criada uma amostra estratificada.

Um outro exemplo pode ser um estudo sobre o território nacional que obriga que numa amostra estejam representados não só todos os distritos como também habitantes da cidade, de vilas e do campo.

Sendo a amostra um subconjunto de elementos retirados da população representativo da mesma, chama-se **dimensão ou tamanho** da amostra ao número de elementos que foram selecionados. Uma vez analisada a amostra de forma a criar informação válida, isto para que possa ser generalizada à população, o nível de confiança dos resultados inferidos é tanto maior quanto maior a dimensão dessa amostra.

Atendendo a que as dificuldades para a realização de um censo podem ser suficientes para que muitas vezes ele não se possa realizar; a utilização de amostras pode ser uma boa opção, muitas vezes a única, para a viabilização do estudo pretendido.

Os resultados da observação efetuada sobre as unidades estatísticas que constituem a amostra têm uma característica importante que é a sua variabilidade.

Atendendo a que são muitas as dificuldades para a realização de um censo (custos envolvidos, meios humanos necessários, longa duração, etc.) elas podem impedir a sua própria realização pelo que a utilização de amostras pode ser uma boa opção, muitas vezes a única para a viabilização do estudo pretendido.

Por exemplo, numa empresa só o facto de trabalhar com amostras se poder tirar conclusões rapidamente pode permitir, em tempo útil, implementar atempadamente decisões que melhorem a performance.

Uma classificação das amostras quanto às características que se pretende:

- A amostra diz-se **univariada**, quando se pretende observar e tratar estatisticamente apenas uma característica para cada objeto ou para cada indivíduo da amostra. Os dados resultantes dizem-se **univariados**.
- Por outro lado, a amostra é **bivariada** quando se pretende observar e tratar simultaneamente duas características para cada objeto ou para cada indivíduo da amostra. Os dados resultantes são **bivariados**.
- A amostra é **multivariada** se pretendemos, para cada objeto ou indivíduo, observar e tratar simultaneamente valores de várias características. Então, os dados resultantes são **multivariados**.

Nos exemplos seguintes ilustram-se situações onde os dados são univariados, bivariados ou multivariados.

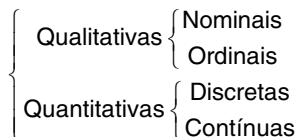
- A nota final dos alunos de determinada turma na unidade curricular Estatística é um dado univariado.
- A nota final e o número de faltas injustificadas de cada aluno de uma turma na unidade curricular Estatística são dados bivariados.

Neste texto estudam-se apenas dados univariados.

3. Variáveis estatísticas – classificação

Define-se **variável estatística**, ou, simplesmente, **variável**, como sendo uma propriedade ou uma característica dos elementos da população à qual se possa atribuir um número ou uma categoria.

Figura 3. Classificação de Variáveis



As variáveis quanto ao conteúdo podem ser **qualitativas** ou **quantitativas**.

Variável **qualitativa** (ou categórica) é aquela em que a característica associada permite classificar as unidades estatísticas em categorias ou modalidades e os dados estatísticos correspondentes traduzem essas mesmas categorias ou modalidades. As variáveis qualitativas podem ainda ser classificadas em **nominais** e **ordinais**. Classifica-se como **variável qualitativa nominal** aquela cujas categorias associadas são identificadas por um nome, não existindo qualquer critério relevante que permita estabelecer uma ordenação de uma em relação às restantes. São exaustivas e exclusivas, isto é, permitem a classificação de todas as unidades estatísticas numa e só numa categoria. Mesmo que a cada categoria seja associado um número, não é possível estabelecer qualquer relação de ordem entre esses números enquanto representativos dessas categorias.



Elsa Rosário Negas, licenciou-se em 1992 em Gestão de Empresas e é Mestre em Matemáticas Aplicadas pela Faculdade de Ciências da Universidade de Lisboa. Atualmente é docente na Universidade Lusíada e no Instituto Superior de Contabilidade e Administração de Lisboa, nas áreas de Matemática, Investigação Operacional, Estatística e Análise de Dados. Exerce ainda tutoria em Matemática Aplicada à Gestão, Probabilidade e Estatística e Estatística para as Ciências Sociais na Universidade Aberta. Como investigadora, colabora com o CITAD (Centro de Investigação em Território, Arquitectura e Design) e organizadora e sido membro da Comissão Científica dos seminários internacionais de Arquitectura e Matemática – As Fórmulas na Arquitectura, que se realizam na Universidade Lusíada de Lisboa.

Este manual apresenta uma exposição teórica dos conceitos básicos da Análise de Dados, apoiada em exemplos ilustrativos e exercícios resolvidos. A abordagem dos conceitos é rigorosa sem ser exaustiva nos seus aspectos matemáticos, de modo a que os conteúdos possam ser facilmente entendidos, mesmo por leitores que não tenham uma preparação matemática avançada.

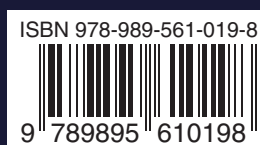
Partindo de exemplos práticos, que na sua maior parte se baseiam e descrevem circunstâncias da vida real, introduzem-se os conceitos teóricos e chama-se a atenção para alguns pontos em particular que, pela sua dificuldade ou pela sua importância, necessitam de um maior esclarecimento.

Em cada capítulo, são propostos dois tipos de exercícios de aplicação: exercícios resolvidos e exercícios de autoavaliação. Nos exercícios resolvidos é apresentada a resolução detalhada, apresentando-se, quando necessário, notas de esclarecimento para a compreensão plena da sua resolução. Nos exercícios de autoavaliação, o leitor é convidado a desenvolver autonomamente a sua resolução e a aferir os conhecimentos adquiridos, podendo confrontar os resultados com os da solução.

Este livro beneficia da vasta experiência da autora no apoio e lecionação das matérias expostas ao nível do ensino superior nas disciplinas de Estatística nas áreas das Ciências Humanas e das Ciências Exactas e, em particular, do seu trabalho desenvolvido com alunos do ensino a distância. É, por isso, um instrumento conciso, com a clareza necessária ao estudo autónomo, de forma a responder às necessidades de alunos do ensino superior, profissionais e professores do ensino secundário.

ESTATÍSTICA DESCRITIVA

Explicação Teórica, Casos de Aplicações
e Exercícios Resolvidos



623