

Estatística Aplicada **1**

Para quem se inicia na estatística.

Para quem, tendo já alguns conhecimentos, necessita esclarecer, aprofundar ou rever tópicos específicos.

Com casos e exemplos do dia a dia que garantem uma aprendizagem fácil e eficaz.



**Elizabeth Reis
Paulo Melo
Rosa Andrade
Teresa Calapez**

6ª Edição
Revista e Aumentada



EDIÇÕES SÍLABO

ESTADÍSTICA APLICADA

Volume 1

ESTATÍSTICA APLICADA

Volume 1

**Probabilidades, Variáveis aleatórias,
Distribuições Teóricas**

Volume 2

**Amostragem, Estimação pontual e por intervalos,
Ensaio de hipóteses paramétricos e não paramétricos**

EXERCÍCIOS DE ESTATÍSTICA APLICADA

Volume 1

**Probabilidades, Variáveis aleatórias,
Distribuições Teóricas**

Volume 2

**Amostragem, Estimação pontual e por intervalos,
Ensaio de hipóteses paramétricos e não paramétricos**

ESTATÍSTICA APLICADA

Volume 1

**Probabilidades, Variáveis aleatórias,
Distribuições Teóricas**

*Elizabeth Reis
Paulo Melo
Rosa Andrade
Teresa Calapez*

6ª Edição
Revista e Aumentada



É expressamente proibido reproduzir, no todo ou em parte, sob qualquer forma ou meio, **NOMEADAMENTE FOTOCÓPIA**, esta obra. As transgressões serão passíveis das penalizações previstas na legislação em vigor.

Visite a Sílabo na rede

www.silabo.pt

Editor: Manuel Robalo

FICHA TÉCNICA:

Título: Estatística Aplicada – Vol. 1

Autores: Elizabeth Reis, Paulo Melo, Rosa Andrade, Teresa Calapez

© Edições Sílabo, Lda.

1ª Edição – Lisboa, outubro de 1997.

6ª Edição – Lisboa, setembro de 2015.

Impressão e acabamentos: Europress, Lda.

Depósito Legal: 398128/15

ISBN: 978-972-618-819-3

EDIÇÕES SÍLABO, LDA.

R. Cidade de Manchester, 2

1170-100 Lisboa

Tel.: 218130345

Fax: 218166719

e-mail: silabo@silabo.pt

www.silabo.pt

Índice

ÍNDICE VOLUME 2	12
NOTA À SEXTA EDIÇÃO	15
PREFÁCIO	17

Capítulo I – Introdução

1. DUAS RAZÕES PARA SE ESTUDAR ESTATÍSTICA	21
2. A NECESSIDADE DA ESTATÍSTICA NAS CIÊNCIAS ECONÓMICAS E DE GESTÃO	21
3. MÉTODO ESTATÍSTICO DE RESOLUÇÃO DE UM PROBLEMA	23
4. ESTATÍSTICA DESCRITIVA E INFERÊNCIA ESTATÍSTICA	24
5. ESCALAS DE MEDIDA DOS DADOS ESTATÍSTICOS	26
5.1. Escala nominal	27
5.2. Escala ordinal	28
5.3. Escala por intervalos	28
5.4. Escala de rácios	29
6. ALGUMAS CONSIDERAÇÕES FINAIS	29
7. UTILIZAÇÃO DO COMPUTADOR	30

Capítulo II – Teoria das probabilidades

1. RESUMO HISTÓRICO	33
2. CONCEITOS DA TEORIA DAS PROBABILIDADES	36
2.1. Experiência aleatória	36
2.2. Espaço de resultados	37
2.3. Acontecimentos	39

3. ÁLGEBRA DOS ACONTECIMENTOS	43
3.1. União de acontecimentos	43
3.2. Interseção de acontecimentos	44
3.3. Diferença de acontecimentos	46
3.4. Propriedades das operações	48
4. CONCEITOS DE PROBABILIDADE	49
4.1. Conceito clássico de probabilidade (<i>a priori</i>)	50
4.2. Conceito frequencista de probabilidade (<i>a posteriori</i>)	52
4.3. Conceito subjectivo ou personalista de probabilidade	53
5. AXIOMAS DA TEORIA DAS PROBABILIDADES	55
6. PROBABILIDADES CONDICIONADAS	66
6.1. Axiomática e teoremas da teoria das probabilidades na probabilidade condicionada	69
7. PROBABILIDADE DE INTERSECÇÃO DE ACONTECIMENTOS. ACONTECIMENTOS INDEPENDENTES	72
7.1. Probabilidade de interseção de acontecimentos	72
7.2. Acontecimentos independentes	74
7.3. Acontecimentos independentes versus acontecimentos incompatíveis ou mutuamente exclusivos	80
8. TEOREMA DA PROBABILIDADE TOTAL E FÓRMULA DE BAYES	82
8.1. Teorema da probabilidade total	83
8.2. Fórmula de Bayes	84
EXERCÍCIOS PROPOSTOS	87

Capítulo III – Variáveis aleatórias

1. DEFINIÇÃO	97
1.1. Enquadramento e exemplos	97
1.2. Cálculo de probabilidades através de variáveis aleatórias	102
1.3. Variáveis aleatórias unidimensionais e bidimensionais	104
2. FUNÇÕES DE PROBABILIDADE E DE DISTRIBUIÇÃO DE VARIÁVEIS ALEATÓRIAS UNIDIMENSIONAIS	107
2.1. Variáveis aleatórias discretas	107
2.1.1. Função de probabilidade	107
2.1.2. Função de distribuição	112

2.2. Variáveis aleatórias contínuas	115
3. FUNÇÕES DE PROBABILIDADE E DE DISTRIBUIÇÃO DE VARIÁVEIS ALEATÓRIAS BIDIMENSIONAIS	123
3.1. Variáveis aleatórias discretas	123
3.1.1. Função de probabilidade conjunta	123
3.1.2. Função de distribuição conjunta	125
3.1.3. Função de probabilidade marginal	127
3.1.4. Função de probabilidade condicionada	128
3.1.5. Independência de variáveis aleatórias	130
3.2. Variáveis aleatórias contínuas	131
3.2.1. Definição	131
3.2.2. Cálculo de probabilidades	133
3.2.3. Funções de densidade de probabilidade marginais	135
3.2.4. Independência	136
4. PARÂMETROS DE VARIÁVEIS ALEATÓRIAS: VALOR ESPERADO E VARIÂNCIA	137
4.1. Média ou valor esperado	137
4.1.1. Definição	137
4.1.2. Propriedades do valor esperado	139
4.1.3. Valor esperado de função de variável aleatória	141
4.1.4. Valor esperado monetário (V.E.M.)	143
4.2. Variância e desvio-padrão	147
4.2.1. Propriedades da variância	149
4.3. Covariância e coeficiente de correlação linear	150
5. MOMENTOS	155
5.1. Função geradora de momentos	157
6. DESIGUALDADES DE MARKOV E CHEBYSHEV	158
EXERCÍCIOS PROPOSTOS	164

Capítulo IV – Distribuições teóricas mais importantes

1. DISTRIBUIÇÕES DISCRETAS	179
1.1. A distribuição uniforme	179
1.2. Prova de Bernoulli	184
1.3. A distribuição de Bernoulli	187

1.4. A distribuição binomial	189
1.4.1. A função de probabilidade da binomial	190
1.4.2. Aspecto gráfico da função de probabilidade da binomial	195
1.4.3. Parâmetros da distribuição binomial	199
1.4.4. A aditividade nas distribuições binomiais	202
1.4.5. Outras aplicações da distribuição binomial	203
1.5. A distribuição multinomial	207
1.5.1. Parâmetros mais importantes da multinomial	209
1.6. A distribuição binomial negativa	213
1.6.1. Relação entre a binomial e a binomial negativa	214
1.6.2. Parâmetros mais importantes da binomial negativa	215
1.7. A distribuição geométrica ou de Pascal	216
1.7.1. Parâmetros mais importantes da distribuição geométrica	217
1.8. A distribuição hipergeométrica	218
1.8.1. Parâmetros mais importantes da distribuição hipergeométrica	221
1.8.2. Generalização da distribuição hipergeométrica	222
1.9. A distribuição de Poisson	224
1.9.1. O processo de Poisson	224
1.9.2. Parâmetros mais importantes da distribuição de Poisson	227
1.9.3. A aditividade nas distribuições de Poisson	230
1.9.4. Aproximação da distribuição binomial à Poisson	232
2. DISTRIBUIÇÕES CONTÍNUAS	237
2.1. A distribuição uniforme	237
2.2. A distribuição normal	240
2.2.1. Características da distribuição normal	241
2.2.2. Cálculo de probabilidades na distribuição normal	243
2.2.3. A aditividade da distribuição normal	250
2.2.4. A distribuição normal como uma aproximação da distribuição binomial	252
2.2.5. A distribuição normal como uma aproximação da distribuição de Poisson	253
2.3. A distribuição exponencial	255
2.4. A distribuição Gama	258
2.5. A distribuição do Qui-quadrado	262
2.5.1. Principais características da distribuição do χ^2	262
2.5.2. Alguns teoremas	263

2.6. A distribuição <i>t</i> de <i>Student</i>	264
2.6.1. Principais características da distribuição <i>t</i> de <i>Student</i> . . .	265
2.6.2. Alguns teoremas	265
2.7. A distribuição do <i>F</i> de Snedecor	266
2.7.1. Principais características da distribuição <i>F</i> de Snedecor . .	267
2.7.2. Alguns teoremas	267
EXERCÍCIOS PROPOSTOS	269

Apêndice – Tabelas de distribuição

Distribuição Binomial	281
Distribuição de Poisson	286
Distribuição Normal Padrão	293
Distribuição χ^2	294
Distribuição <i>t</i> de <i>Student</i>	295
Distribuição <i>F</i> de Snedecor	296
BIBLIOGRAFIA	299

Índice Volume 2

NOTA INTRODUTÓRIA À SEGUNDA EDIÇÃO

PREFÁCIO

Capítulo V – O processo de amostragem

1. INTRODUÇÃO

2. ALGUNS CONCEITOS IMPORTANTES NA TEORIA
DA AMOSTRAGEM

3. QUESTÕES PRÉVIAS AO PROCESSO DE AMOSTRAGEM

4. AS FASES DO PROCESSO DE AMOSTRAGEM

EXERCÍCIOS PROPOSTOS

Capítulo VI – Distribuições amostrais

1. INTRODUÇÃO

2. DISTRIBUIÇÕES AMOSTRAIS TEÓRICAS

3. DISTRIBUIÇÕES AMOSTRAIS DAS ESTATÍSTICAS
MAIS IMPORTANTES

EXERCÍCIOS PROPOSTOS

Capítulo VII – Estimação de parâmetros

1. INTRODUÇÃO

2. ESTIMAÇÃO PONTUAL

3. ESTIMAÇÃO POR INTERVALOS

EXERCÍCIOS PROPOSTOS

Capítulo VIII – Ensaio de hipóteses

1. A NECESSIDADE DOS ENSAIOS DE HIPÓTESES
 2. HIPÓTESES E ERROS
 3. COMO FAZER UM ENSAIO DE HIPÓTESES
 4. ERROS NOS ENSAIOS DE HIPÓTESES
 5. ESCOLHA DA ESTATÍSTICA ADEQUADA AO ENSAIO
- EXERCÍCIOS PROPOSTOS

Capítulo IX – Testes não-paramétricos

1. INTRODUÇÃO
 2. TESTES DE AJUSTAMENTO
 3. TABELAS DE CONTINGÊNCIA
 4. TESTES À IGUALDADE DE DUAS OU MAIS DISTRIBUIÇÕES
 5. COMPARAÇÕES ENTRE DUAS AMOSTRAS EMPARELHADAS
- EXERCÍCIOS PROPOSTOS

APÊNDICE – TABELAS DE DISTRIBUIÇÃO

BIBLIOGRAFIA

Nota à sexta edição

Esta edição da obra *Estatística Aplicada*, constitui uma nova versão revista e aumentada. Está estruturada em dois volumes para responder às solicitações de muitos leitores, docentes e alunos, coincidindo cada um dos volumes com programas específicos de *Estatística* lecionados nas várias Universidades.

Nesta edição, o número de exercícios propostos no final de cada capítulo, com as respetivas soluções, foi significativamente aumentado. Muitos destes exercícios encontram-se resolvidos na obra dos mesmos autores: *Exercícios, Estatística Aplicada, Volume 1*, editada em 2012 pela Sílabo.

O primeiro volume, para além do capítulo introdutório, inclui um segundo capítulo sobre Teoria das Probabilidades, um terceiro sobre Variáveis Aleatórias, sendo o quarto e último sobre Distribuições Teóricas mais Importantes.

O segundo volume, sendo maioritariamente dedicado aos métodos de Inferência Estatística (capítulos VII, VIII e IX, Estimação de Parâmetros, Ensaio de Hipóteses e Testes não-Paramétricos), inclui também uma breve introdução aos Processos de Amostragem (capítulo V) e a apresentação das Distribuições Amostrais (capítulo VI).

Finalmente, agradecemos as sugestões e indicações de muitos leitores que contribuíram para a produção desta nova edição revista e aumentada que, acreditamos constituir uma mais valia para os leitores, docentes e alunos.

Os autores
Lisboa, Setembro de 2015

Prefácio

Este livro de Estatística Aplicada destina-se a profissionais licenciados ou não e a estudantes universitários que, na vida prática ou no processo de aprendizagem, têm necessidade de saber Estatística e de a aplicar aos problemas mais variados do dia-a-dia. Como objetivos finais, este livro pretende tornar compreensíveis a linguagem e notação estatísticas, bem como exemplificar as suas potenciais utilizações, sem descuidar os pressupostos subjacentes e o rigor teórico necessário.

Deverá referir-se que a escolha do título não foi pacífica. De entre os vários alternativos — Probabilidades e Estatística, Inferência Estatística, etc. — a preferência por Estatística Aplicada justifica-se pela abordagem diferenciada de outras obras já publicadas sobre Inferência Estatística, e que resumidamente pode ser assim descrita: mais do que «ensinar», pretende-se com este livro, a) despertar e estimular o interesse dos leitores pelo método estatístico de resolução dos problemas; b) utilizando uma linguagem simples e acessível, apresentar os conceitos e métodos de análise estatística de modo mais intuitivo e informal; c) acompanhar a apetência teórica com exemplos apropriados a cada situação.

O livro encontra-se dividido em nove capítulos. No capítulo I (Introdução) são explicitadas várias razões para que um profissional, técnico, estudante ou mero cidadão adquira um nível mínimo de conhecimentos em Estatística.

A Teoria das Probabilidades é objeto de estudo do capítulo II. Nele são apresentados os diferentes conceitos de probabilidade e a sua axiomática, dando especial relevo aos teoremas da probabilidade total e de Bayes.

Os terceiro e quarto capítulos, tal como o segundo, são essenciais para a compreensão dos seguintes, relativos à Inferência Estatística. O capítulo III respeita às Variáveis Aleatórias, sua definição, características e propriedades. No quarto capítulo estudam-se em pormenor as distribuições de algumas variáveis aleatórias de importância maior nas áreas de aplicação das ciências sócio-económicas como sejam as distribuições de Bernoulli, binomial, Poisson, binomial negativa, hipergeométrica, multinomial, uniforme e normal.

O capítulo V é dedicado ao estudo dos processos de amostragem, incluindo os diferentes métodos de recolha de uma amostra, enquanto que no capítulo VI se apresentam as distribuições amostrais mais importantes.

Os três últimos capítulos são dedicados à Inferência Estatística propriamente dita. No capítulo VII apresentam-se métodos de estimação de parâmetros, com ênfase especial para o método de máxima verosimilhança. Inclui-se ainda a estimação por intervalos. Os capítulos VIII e IX destinam-se à apresentação, respetivamente, dos ensaios de hipóteses paramétricos e não-paramétricos.

Com exceção do primeiro, todos os restantes capítulos são finalizados com um conjunto de exercícios não resolvidos, acompanhados geralmente das respetivas soluções.

No Apêndice estão incluídas as Tabelas (das distribuições) necessárias à compreensão do texto e à resolução dos exemplos e dos exercícios propostos.

Este livro é o resultado de alguns anos de experiência docente dos seus autores na equipa de Estatística do ISCTE e da tentativa de responder às necessidades sentidas por muitos — alunos e docentes de variadas licenciaturas, docentes do ensino secundário, profissionais e técnicos de diferentes áreas científicas (gestão, economia, sociologia, psicologia, medicina, enfermagem, engenharia, informática, etc.) — que, no decorrer destes anos, e na falta de uma obra que os ajudasse a encontrar as soluções estatísticas apropriadas aos seus problemas, procuraram ajuda junto dos autores.

Sem dúvida que a responsabilidade desta obra é assumida pelos seus autores, mas a sua concretização só se tornou possível com a ajuda, apoio e disponibilidade de muitos. Por isso, não deixando de agradecer a todos os que, directa ou indirectamente, contribuíram para a sua realização, gostaríamos de, nominalmente, dar uma palavra especial de agradecimento aos seguintes docentes de Estatística do ISCTE: Ana Cristina Ferreira, Ana Paula Marques, António Robalo, Fátima Ferrão, Fátima Salgueiro, Graça Trindade, Helena Carvalho, Helena Pestana, João Figueira, J.C. Castro Pinto, J.J. Dias Curto, Margarida Perestrelo e Paula Vicente.

Finalmente, uma palavra de apreço a todos os alunos, quer das licenciaturas do ISCTE, quer dos mestrados do INDEG/ISCTE, cujas sugestões, dúvidas e problemas certamente contribuíram para enriquecer este livro.

Os autores

Capítulo I

Introdução

1. Duas razões para se estudar estatística

Existem duas boas razões para se saber Estatística. Primeiro, qualquer cidadão está diariamente exposto a um enorme conjunto de informações resultantes de estudos sociológicos e de mercado ou económicos, de sondagens políticas ou mesmo de pesquisa científica. Muitos destes resultados baseiam-se em inquéritos por amostragem. Alguns deles utilizam, para o efeito, uma amostra representativa de dimensão adequada e recolhida por um processo aleatório. Outros não. Para estes, a validade dos resultados não ultrapassa a amostra que os originou. A afirmação de que é fácil mentir com Estatística é quase um lugar comum. Qualquer manual que se preze apresenta nas primeiras páginas a famosa citação atribuída a Benjamin Disraeli: «There is three kinds of lies: lies, damned lies and statistics». E o pior é que, de certa forma, esta citação é verdadeira: é fácil distorcer e manipular resultados e conclusões e enganar alguém não-(in)formado. Mas saber Estatística permite que se avaliem os métodos de recolha, os próprios resultados, se detetem e rejeitem falsas conclusões.

Se, para muitos, a necessidade de saber Estatística advém do facto de serem cidadãos do mundo, para alguns essa necessidade é acrescida por uma atividade profissional que requer a utilização de métodos estatísticos de recolha, análise e interpretação de dados. E esta é a segunda razão para se estudar Estatística. A utilização da Estatística nas ciências sociais, políticas, económicas, biológicas, físicas, médicas, de engenharia, etc, é por demais conhecida: os métodos de amostragem e de inferência estatística tornaram-se um dos principais instrumentos do método científico. Para todos os que trabalham nestas áreas, é vital um conhecimento básico dos conceitos, possibilidades e limitações desses métodos.

2. A necessidade da estatística nas áreas profissionais e científicas

Nas várias áreas profissionais e científicas, a Estatística pode ser utilizada com três objetivos: (1) descrever e compreender relações entre diferentes características de uma população, (2) tomar decisões mais corretas e (3) fazer face à mudança.

A quantidade de informação recolhida, processada e finalmente apresentada a um comum mortal cresce tão rapidamente que um processo de seleção e identificação das relações mais importantes se torna imprescindível. É aqui que a Estatística poderá dar o seu primeiro contributo, quer através de métodos meramente descritivos, quer utilizando métodos mais sofisticados de generalização dos resultados de uma amostra a toda a população.

Uma vez identificadas as relações, estas poderão constituir uma ajuda preciosa à tomada de decisões corretas em situações de incerteza. Veja-se o seguinte exemplo.

Através de métodos estatísticos adequados, determinada instituição bancária identificou as características sócio-económicas daqueles que considera serem bons clientes. Esta identificação permite-lhe, no futuro, rejeitar pedidos de crédito por parte de potenciais clientes, cujas características mais se afastam das anteriores.

Planear significa determinar antecipadamente as ações a empreender no futuro. Para fazer face à mudança, é necessário que as decisões e o planeamento se apoiem numa análise cuidada da situação presente e numa previsão realista do que acontecerá no futuro.

Os métodos estatísticos de previsão não permitem adivinhar com uma precisão absoluta os acontecimentos futuros, mas permitem medir as variações atuais e estabelecer os cenários futuros mais prováveis, diminuindo, de algum modo, a incerteza inerente a esses acontecimentos futuros.

Nas várias áreas profissionais e científicas, em particular na gestão das organizações e das empresas, a tomada de decisão é crucial e faz parte do dia-a-dia de qualquer decisor. As consequências dessas decisões são demasiado importantes para que possam basear-se apenas na intuição ou *feeling* momentâneos.

Os decisores das organizações e das empresas são responsáveis pelas decisões mesmo quando estas se baseiam em informações incompletas ou incertas. É precisamente porque à informação disponível está associado um elevado grau de incerteza que a Estatística se tornou tão importante no processo de tomada de decisões: a Estatística permite a extração de conclusões válidas a partir de informação incompleta.

O ambiente de formação de uma decisão varia de um extremo em que pouca, ou nenhuma informação está disponível, ao extremo oposto em que o decisor detém toda ou quase toda a informação sobre a situação. Este último extremo significa que o decisor conhece a situação de todos os elementos da população. A informação disponível a partir dos recenseamentos do INE,

realizados de 10 em 10 anos, é um exemplo. Mas a situação mais comum para os decisores é aquela em que quase nenhuma informação se encontra disponível. Veja-se o exemplo do lançamento de um novo produto utilizando tecnologia de ponta praticamente desconhecida dos consumidores. Como irão estes reagir ao lançamento do novo produto? Ou o caso do estudo e análise da eficácia de certo medicamento, para o qual se conhecem apenas os resultados e efeitos ocorridos no tratamento de certa patologia de um número restrito de doentes. Ou o caso de estudo e análise da fiabilidade de determinada estrutura de engenharia ou equipamento sujeito a condições externas não usuais. À partida, pouca ou nenhuma informação existe para que os decisores possam responder a estas questões.

A Estatística fornece aos decisores instrumentos para que possam responder a estes problemas e tomar decisões com alguma confiança, mesmo quando a quantidade de informação disponível é pequena e as situações futuras são de elevada incerteza.

3. Método estatístico de resolução de um problema

Para que se obtenham resultados válidos, o investigador deve seguir todos os passos que definem o método estatístico de resolução de problemas:

1. *Identificar corretamente o problema* em análise. Mesmo em estudos exploratórios cujo objetivo é identificar possíveis relações entre as características dos indivíduos sem que, à partida, se defina um modelo regulador dessas relações, é necessário identificar o problema para o qual se pretendem encontrar respostas.

2. *Recolher a informação necessária*, relevante para o problema em estudo, em tempo útil e tão completa quanto possível. Esta informação poderá consistir em dados primários, recolhidos através de um questionário, ou dados secundários, recolhidos e publicados através de outra fonte de informação.

3. *Classificar e organizar os dados*, por exemplo, através da codificação e criação de uma base de dados em suporte informático. Uma vez ultrapassada esta fase, é já possível reduzir a quantidade de informação, fazendo desaparecer os pormenores menos importantes através de medidas de estatística descritiva (medidas de tendência central, dispersão, concentração, etc.), quadros e gráficos.

4. *Análise dos dados e apresentação dos resultados*: identificar relações, testar hipóteses, definir modelos com a ajuda de métodos estatísticos apropriados.

5. *Tomar a decisão* mais adequada, ponderando as possíveis opções face aos objetivos inicialmente propostos. A qualidade da informação recolhida e as capacidades do investigador determinam, em grande parte, a adequabilidade das opções propostas.

4. Estatística descritiva e inferência estatística

Embora a classificação e organização dos dados a que se faz referência no terceiro passo seja ainda um capítulo importante da Estatística — a Estatística Descritiva — um segundo capítulo torna-se muito mais importante, quando os dados recolhidos respeitam apenas a um subconjunto da população em estudo e não a toda a população — a Inferência Estatística. Só quando o grupo sobre o qual se pretende obter informação é de dimensão reduzida, se torna viável recolher essa informação para todos os elementos desse grupo. O recenseamento de uma população envolve custos e tempos demasiado elevados para serem suportados por organizações não vocacionadas para o efeito. Por essa razão, se tornaram populares e se generalizaram a todos os domínios científicos as técnicas de amostragem.

Contrariamente a um recenseamento, onde se recolhe informação sobre as características de toda uma população, uma amostra fornece informação sobre um subconjunto dessa população.

Os métodos de Inferência Estatística permitem (1) estimar as características desconhecidas de uma população (por exemplo, a proporção de consumidores que preferem uma dada marca de detergentes) e (2) testar se determinadas hipóteses sobre essas características desconhecidas são plausíveis (por exemplo, se a afirmação de um vendedor de que os resultados de imagem da marca que vende são superiores aos de outras marcas concorrentes).

Nos exemplos anteriores, as características das populações (proporção de consumidores e resultados médios da imagem da marca) são os *parâmetros*. Quando respeitam a uma amostra, estes indicadores estatísticos passam a chamar-se *estatísticas*.

Os métodos de Inferência Estatística envolvem o cálculo de estatísticas, a partir das quais se infere sobre os parâmetros da população, isto é, permitem, com determinado grau de probabilidade, generalizar à população certas conclusões, por comparação com os resultados amostrais.

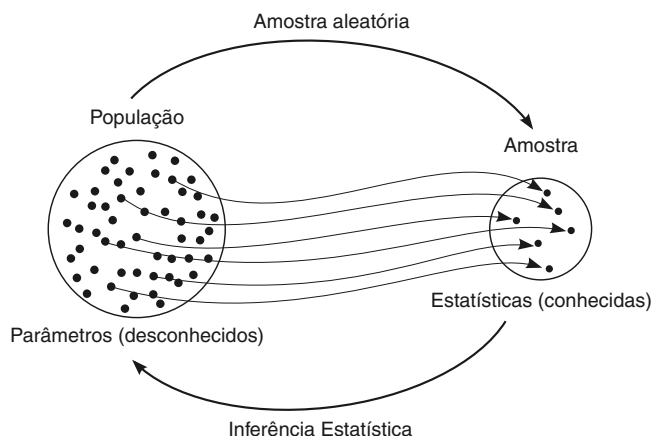
Exemplos de parâmetros são a média de uma população (μ), a variância (σ^2) ou o desvio-padrão (σ). Como exemplos de estatísticas: a média (\bar{X}), a variância (S^2) ou o desvio-padrão (S) amostrais.

A distinção entre parâmetro e estatística torna-se extremamente importante na Inferência Estatística. Muitas vezes pretende-se estimar o valor de um parâmetro ou fazer um teste de hipóteses sobre o seu valor. No entanto, o cálculo dos parâmetros é, geralmente, impossível ou impraticável, devido aos requisitos de tempo e dinheiro a que obriga. Nestes casos, a escolha de uma amostra aleatória permite que se obtenha uma estimativa para o parâmetro. A base da Inferência Estatística consiste, assim, na possibilidade de se tomarem decisões sobre os parâmetros de uma população, sem que seja necessário proceder a um recenseamento de toda a população.

Exemplo

Um industrial de máquinas de lavar quer determinar qual o número médio de lavagens de determinado tipo de máquina (lavar e secar), até que necessitem de reparação. O parâmetro que pretende conhecer é o número médio de lavagens das máquinas até serem reparadas. O técnico da sua fábrica seleciona aleatoriamente algumas máquinas da sua produção mensal, e verifica as lavagens efetuadas até ocorrer uma avaria, calculando, em seguida, para as máquinas da amostra, o número médio de lavagens, isto é, a média amostral.

A figura seguinte demonstra o processo seguido.



O processo de generalizar à população os resultados recolhidos na amostra é feito num ambiente de incerteza. A não ser que o valor dos parâmetros seja calculado a partir de todos os elementos da população, nunca se saberá com certeza se as estimativas ou inferências feitas são verdadeiras ou não. Num esforço para medir o grau de confiança ou de certeza associado aos resultados do processo de inferência, a Estatística utiliza a teoria das probabilidades. Por essa razão se dedica um capítulo deste livro ao estudo das probabilidades.

5. Escalas de medida dos dados estatísticos

Os exemplos de dados que diariamente se podem recolher são dos mais variados. Vejamos alguns: a temperatura máxima na cidade de Lisboa; a cotação do euro face ao dólar; as taxas de inflação dos países da União Europeia; as exportações de material eletrónico dos países da Ásia Oriental; a distribuição etária da população do concelho de Lisboa; a distribuição por sexo dessa mesma população; as profissões da população da Marinha Grande; a distribuição dos emigrantes portugueses por países de acolhimento; as preferências da população portuguesa no que respeita às suas viagens de férias; as preferências dos portugueses em relação aos quatro canais de televisão nacional; as quotas de mercado das diferentes marcas de automóveis utilitários.

Estes exemplos de dados estatísticos diferenciam-se, não só por se referirem a características de diferentes populações, mas também por estarem definidos em diferentes escalas de medida e, portanto, por necessitarem de diferentes métodos estatísticos para os descreverem e analisarem. São quatro os tipos de escalas de medida: nominal, ordinal, por intervalos e por rácios. Nem sempre é evidente a distinção entre estas escalas, sobretudo entre as duas últimas. A classificação que se descreverá em seguida é a adoptada pelos autores deste livro, embora se reconheça não existir unanimidade neste domínio.

5.1. Escala nominal

Os dados definidos numa escala nominal são dados qualitativos por excelência. Por exemplo, suponha-se que se pretendia conhecer a característica *profissão* da população constituída pelos pais dos alunos universitários. O estudo desta característica permitiria descrever o conjunto de profissões desta população, através de uma listagem que incluiria:

- trabalhador qualificado
- médico
- advogado
- militar
- professor
- bancário
- etc, etc, etc.

Suponha-se ainda que, para efeitos de processamento dos dados, se codificava cada um dos valores desta característica, dando o valor 1 ao trabalhador qualificado, 2 ao médico, 3 ao advogado, 4 ao militar, e assim por diante. Estes números são utilizados apenas como códigos e não como valores quantitativos, uma vez que, por exemplo, ao valor 4, não está associada uma maior quantidade do que aos valores 1, 2 e 3.

Os códigos numéricos são utilizados para diferenciar as categorias desta característica, não fazendo qualquer sentido calcular indicadores quantitativos (como a média ou desvio-padrão) a partir destes números.

Outros exemplos de características definidas em escalas nominais são a religião, a raça, a localização geográfica, o local de nascimento, o sexo, os setores de atividade económica.

Um caso particular deste tipo de escala de medida ocorre quando a característica em estudo tem apenas duas categorias: são as chamadas características binárias ou dicotómicas. São exemplos deste tipo de características o sexo (que pode ser masculino ou feminino), e a resposta à seguinte pergunta: «Reside em Lisboa?» (podendo ser *Sim* ou *Não*).

5.2. Escala ordinal

Quando numa característica nominal a ordem das categorias obedece a uma sequência com significado, está-se em presença de uma característica definida numa escala ordinal. Os códigos numéricos que identificam as categorias já não são dados de forma arbitrária mas sim de tal modo que as categorias às quais foram dados o primeiro e último códigos são as que mais distam e mais se diferenciam entre si.

As escalas ordinais tornam-se extremamente úteis para medir opiniões subjectivas sobre as qualidades de certos atributos, cuja medição objectiva é impossível. Por exemplo, poder-se-á perguntar a um consumidor qual a sua opinião sobre o sabor de determinado produto alimentar, de acordo com a seguinte escala:

- 1 – *detesta*
- 2 – *gosta pouco*
- 3 – *indiferente*
- 4 – *gosta*
- 5 – *adora*.

As respostas a esta questão podem ser resumidas numa escala ordinal, com cinco categorias, vulgarmente conhecida por escala de Likert.

Outro exemplo consistiria em solicitar aos consumidores que ordenassem por ordem decrescente de preferência, de 1 até 8, oito marcas de sabonetes.

Um outro modo de obtenção de uma escala ordinal consiste em dividir uma escala contínua em múltiplos intervalos. Por exemplo, os indivíduos de uma população podem ser classificados em três grandes grupos, resultantes da divisão de um intervalo contínuo de idades: jovens (até 18 anos), adultos (de 18 a 65 anos) e idosos (mais de 65 anos).

5.3. Escala por intervalos

Para além das propriedades da escala ordinal, a escala por intervalos tem ainda a propriedade de as distâncias iguais corresponderem quantidades iguais. As escalas por intervalos podem ser contínuas ou discretas. São contínuas se podem tomar um número infinito não numerável de valores e são discretas se



PAULO MELO licenciou-se em Economia pelo ISCTE em 1977. Desde então tem lecionado várias disciplinas na área de Métodos Quantitativos, no ISCTE e na Faculdade de Economia da Universidade Nova de Lisboa, bem como em cursos de pós-graduação promovidos pelo INDEG/ISCTE. Em 1985 concluiu o mestrado em Economia, na área de Economia das Telecomunicações, pela Faculdade de Economia da UNL. Colaborou com a Direção de Planeamento do Banco Fonseca & Burnay e com a Direção de Marketing da TMN. Nos CTT – Telecomunicações foi responsável pelas áreas de Política Tarifária e de Planeamento de Marketing.

ROSA ANDRADE licenciou-se em Economia pelo ISCTE, em 1977. Em 1982 completou a parte escolar do mestrado em Métodos Matemáticos Aplicados à Economia e Gestão de Empresas do ISEG. Lecionou em diversas unidades curriculares na área dos Métodos Quantitativos no ISCTE, tendo iniciado a partir de 1986, uma carreira profissional na área do Marketing e do Planeamento e Controlo. Atualmente é professora auxiliar convidada no ISCTE-IUL onde leciona Estatística I e Estatística II na licenciatura de Gestão e é consultora na Autoridade Nacional de Comunicações (ICP-ANACOM) tendo desempenhado, desde 1993, várias funções técnicas e de chefia nesta organização.

TERESA CALAPEZ licenciou-se em Matemáticas Aplicadas, em 1986, pela Faculdade de Ciências de Lisboa. Concluiu o mestrado em Estatística e Investigação Operacional em 1991 pela mesma Faculdade e o doutoramento em Métodos Quantitativos, na especialidade de Análise Multivariada de Dados, pelo ISCTE-IUL, em 2004. Atualmente é professora auxiliar no ISCTE-IUL onde leciona e coordena várias unidades curriculares na área da Estatística e da Análise de Dados, nos diversos ciclos de ensino. É membro da Sociedade Portuguesa de Estatística, investigadora da BRU/UNIDE-IUL, Business Research Unit do ISCTE-IUL e colabora regularmente com o Dinâmia/CET-IUL.

ELIZABETH REIS licenciou-se em Economia pela Faculdade de Economia da Universidade do Porto em 1979. Fez o primeiro curso de pós-graduação em Economia Europeia (1981) na Universidade Católica Portuguesa. Em 1984 concluiu o M.Sc. em Social Statistics na Universidade de Southampton e, em 1987, o Ph.D. na mesma área científica. É doutora em Métodos Quantitativos para Gestão pela Universidade Técnica de Lisboa, tendo obtido a Agregação em Métodos Quantitativos pelo ISCTE-IUL em 1999, instituição onde atualmente é Professora Catedrática e leciona e coordena unidades curriculares de Estatística, Análise de Dados, Metodologias de Recolha de Informação, Métodos de Investigação e Pesquisa de Mercados nos níveis de Licenciatura, Mestrado e Doutoramento, em diversos domínios da Gestão Empresarial e das Ciências Sociais.

Este livro de Estatística sintetiza e beneficia da experiência acumulada pelos autores em mais de três décadas de ensino em licenciaturas, mestrados e doutoramentos.

Depois de várias edições e reimpressões, que atestam a sua popularidade e excelente acolhimento pelo seu público-alvo, surge agora em sexta edição, revista e ampliada, com mais casos práticos, exercícios, exemplos e a introdução de melhorias na exposição das matérias teóricas apresentadas.

Numa linguagem simples e acessível, profusamente ilustrado, com exemplos e situações do dia a dia este livro desperta e estimula o interesse dos leitores que, quer por motivos académicos quer por motivos profissionais, tenham de aprender ou aplicar a estatística.

As matérias abordadas são as que fazem parte da generalidade das disciplinas de estatística dos diversos graus de ensino.

Profissionais e estudantes de gestão, economia, sociologia, psicologia, antropologia, medicina, engenharia, informática, comunicação social e muitos outros dispõem agora de um novo manual, revisto e aumentado, a que poderão sempre recorrer.

Estatística Aplicada 1

ISBN 978-972-618-819-3



9 789726 188193