

POTENCIAR OS NEGÓCIOS? A ESTATÍSTICA DÁ UMA AJUDA!

MUITAS APLICAÇÕES EM EXCEL
e poucas fórmulas...

JOSÉ DIAS CURTO

<https://diascurto.wixsite.com/sitedc/estatistica>
dias.curto@iscte-iul.pt

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 página do livro:

<https://diascurto.wixsite.com/sitedc/estatistica>

FICHA TÉCNICA:

Título: Potenciar os negócios? A Estatística dá uma ajuda!

Autor: José Joaquim Dias Curto

© José Joaquim Dias Curto

Design: Starline - Freepik.com; Freepik

Direção de arte: Filipa Pestana

Impressão e acabamentos: GUIDE – Artes Gráficas, Lda

3ª edição, setembro de 2019

Depósito Legal: 414740/16

ISBN: 978-989-20-6961-6

Para o Tomás e para a Ticas

Índice

| | |
|--|------------|
| Prefácio | v |
| Prefácio à segunda edição | vii |
| Prefácio à primeira edição | ix |
| A estatística sempre “por perto” | xi |
| 1 “Coisas” simples mas necessárias | 1 |
| 1.1 Género e Salário: são a mesma coisa? | 2 |
| 1.1.1 Dados qualitativos | 3 |
| 1.1.2 Dados quantitativos | 4 |
| 1.2 Dados de séries cronológicas, seccionais e em painel | 4 |
| 1.2.1 Variáveis discretas e contínuas | 6 |
| 1.3 Apresentação dos dados: em tabelas, por classes... | 6 |
| 1.4 Preparar o Excel | 8 |
| 1.4.1 Suplemento “Analysis ToolPack” | 8 |
| 1.4.2 Outros suplementos para o Excel | 10 |
| 1.4.3 Funções de estatística | 10 |
| 1.5 Ficheiros de dados | 12 |
| 1.6 Breve caracterização dos trabalhadores da PIXIES | 13 |
| 1.7 À procura de dados para analisar | 21 |
| 2 O que “dizem” os dados? (Não é jogo nem tarot...) | 22 |
| 2.1 Poupar para comprar um automóvel | 25 |
| 2.2 Quanto vale uma empresa de SI em Portugal? | 30 |
| 2.2.1 Indicação sobre o VNme e RLme por empresa | 34 |
| 2.2.2 Quantis e percentis | 36 |
| 2.3 Para destacar as diferenças entre empresas | 37 |
| 2.4 O peso das “anormalidades” | 42 |
| 2.5 Rendibilidade e risco nos mercados financeiros | 48 |

| | |
|---|------------|
| Apêndices | 61 |
| Apêndice 2.A Cálculo dos percentis em Excel | 61 |
| 3 É possível “dormir” mais descansado? | 63 |
| 3.1 Não é certo? Mas tem uma probabilidade... | 65 |
| 3.2 Variáveis aleatórias | 68 |
| 3.3 Caraterísticas das distribuições de probabilidade | 73 |
| 3.3.1 Média ou valor esperado | 74 |
| 3.3.2 Variância | 75 |
| 3.3.3 Covariância e coeficiente de correlação simples | 76 |
| 3.4 Distribuições de probabilidade teóricas | 78 |
| 3.4.1 Distribuições: o que são e para que servem? | 79 |
| 3.4.2 Todos têm a mesma probabilidade | 80 |
| 3.4.3 Sucessos/Insucessos: quantos são, quantos são? | 81 |
| 3.4.4 Com reposição: distribuição hipergeométrica | 90 |
| 3.4.5 Limitar no espaço ou no tempo | 92 |
| 3.4.6 Distribuição Normal | 93 |
| 3.4.7 Distribuição do qui-quadrado | 103 |
| 3.4.8 Distribuição t de Student | 105 |
| 3.4.9 Distribuição F de Snedecor | 107 |
| 3.5 Simulação de vendas | 110 |
| 3.5.1 Distribuição Uniforme: $U(a, b)$ | 111 |
| 3.5.2 Distribuição Triangular: $T(a, b, c)$ | 117 |
| 3.5.3 Distribuição Exponencial | 120 |
| 3.5.4 Simulação de vendas da empresa WILDNOTHING | 122 |
| Apêndices | 126 |
| Apêndice 3.A Momentos e parâmetros de ordem de uma distribuição | 126 |
| 4 Uma parte para “revelar” o todo | 128 |
| 4.1 População e amostra | 129 |
| 4.2 Métodos de seleção das amostras | 131 |
| 4.3 Parâmetros e estatísticas | 137 |
| 4.4 Distribuições por amostragem (DPA) | 138 |
| 4.4.1 Teorema do Limite Central | 142 |
| 4.4.2 DPA da média amostral | 143 |
| 4.4.3 DPA da variância amostral | 144 |
| 4.4.4 DPA da proporção amostral | 145 |
| 4.5 Estimação paramétrica | 145 |
| 4.6 Propriedades em amostras finitas | 147 |
| 4.7 Propriedades em grandes amostras | 150 |
| 4.8 Estimação por intervalos | 153 |

| | | |
|----------|---|------------|
| 4.9 | Dimensão de uma amostra aleatória simples | 157 |
| 4.9.1 | Populações de Bernoulli | 158 |
| 4.9.2 | Populações normais | 160 |
| 5 | Testar para decidir | 162 |
| 5.1 | Na Bolsa, é para ganhar ou para perder? | 164 |
| 5.1.1 | Abordagem do intervalo de confiança | 166 |
| 5.1.2 | A abordagem dos testes (ou ensaios) de hipóteses | 167 |
| 5.1.3 | “Aceitar” ou “Não Rejeitar”, que terminologia utilizar? | 173 |
| 5.1.4 | Tipo de erros | 175 |
| 5.2 | O impacte da formação nas vendas dos lojistas | 177 |
| 5.2.1 | Teste à diferença de duas médias – amostras independentes | 178 |
| 5.2.2 | Teste à igualdade de variâncias | 179 |
| 5.2.3 | Teste à diferença de duas médias – amostras emparelhadas | 185 |
| 5.3 | Testes à normalidade da distribuição | 188 |
| 5.3.1 | Testes aos coeficientes de assimetria e de curtose | 188 |
| 5.3.2 | Teste de Jarque-Bera (JB) | 189 |
| 5.3.3 | Teste de Kolmogorov-Smirnov (KS) | 190 |
| 6 | Variação com variação dá... relação | 193 |
| 6.1 | Análise de correlação paramétrica | 194 |
| 6.1.1 | Diagrama de dispersão | 195 |
| 6.1.2 | Covariância e Coeficiente de correlação simples | 198 |
| 6.2 | Análise de correlação não paramétrica | 204 |
| 6.3 | Adivinhar? Pode-se tentar... com a regressão (simples) | 208 |
| 6.3.1 | O modelo de regressão linear simples | 211 |
| 6.3.2 | Função de regressão | 213 |
| 6.3.3 | O método dos mínimos quadrados ordinários | 221 |
| 6.3.4 | Coeficiente de determinação (r^2) | 224 |
| 6.3.5 | Hipóteses do modelo de regressão linear | 225 |
| 6.3.6 | Distribuição por amostragem dos estimadores OLS | 230 |
| 6.3.7 | Inferência estatística no modelo | 234 |
| 6.3.8 | Predição no modelo | 237 |
| 6.3.9 | Consumo de eletricidade | 239 |
| 7 | Mais “coisas” na relação entre duas variáveis | 246 |
| 7.1 | Escala e unidades de medida | 247 |
| 7.2 | Modelo logarítmico ou modelo linear-log | 249 |
| 7.3 | Modelo log-linear (ou modelo semilogarítmico) | 254 |
| 7.4 | Modelo duplo-log (ou modelo log-log) | 258 |
| 7.5 | Relações inversas | 261 |
| 7.6 | Polinómios | 261 |

| | | |
|---------------------|---|------------|
| 7.7 | Comparação dos valores de R^2 entre modelos | 263 |
| Apêndices | | 271 |
| | Apêndice 7.A Efeitos marginais e Elasticidades | 271 |
| | Apêndice 7.B Funções exponencial e logarítmica | 273 |
| | 7.B.1 Propriedades das funções | 274 |
| | 7.B.2 Propriedades das derivadas das funções | 275 |
| | 7.B.3 O conceito de elasticidade | 276 |
| 8 | Explicar melhor com a regressão | 278 |
| 8.1 | Os estimadores OLS | 280 |
| | 8.1.1 Coeficientes estandardizados | 281 |
| | 8.1.2 Coeficientes de correlação part e parcial | 282 |
| 8.2 | Coeficiente de determinação corrigido (ou ajustado) | 283 |
| 8.3 | Mais sobre o teste F | 285 |
| | 8.3.1 Poder explicativo e significância estatística | 287 |
| | 8.3.2 Igualdade de coeficientes | 292 |
| 8.4 | Erros de especificação | 294 |
| | 8.4.1 O problema da multicolinearidade | 295 |
| | 8.4.2 Forma funcional da relação | 302 |
| | 8.4.3 Variáveis explicativas endógenas | 304 |
| | 8.4.4 Teste de Chow | 304 |
| | 8.4.5 Heteroscedasticidade | 306 |
| | 8.4.6 Autocorrelação | 311 |
| Bibliografia | | 321 |

Lista de Tabelas

| | | |
|-------|---|-----|
| 1.1 | Dados de Trabalhadores | 2 |
| 1.2 | Tabela de Frequências | 15 |
| 1.3 | Tabela de Frequências: Resultados | 16 |
| 1.4 | Tabela de Frequências: Idade e Salário | 17 |
| 2.1 | Despesas mensais da família Pereira | 25 |
| 2.2 | Poupança dos Pereira para comprar o automóvel | 29 |
| 2.3 | Volume de Negócios (Vendas: VN), Resultados Líquidos (RL) e Número de Trabalhadores (FT) das 30 Maiores Empresas Portuguesas de Sistemas de Informação (Ano 2011) | 31 |
| 2.4 | Medidas de estatística descritiva: resultados | 33 |
| 2.5 | Quartis e Percentis: cálculo | 38 |
| 2.6 | Quartis e Percentis: resultados | 38 |
| 2.7 | Cotações e Rendibilidades diárias do PSI20, SONAE e REN | 50 |
| 2.8 | Medidas de estatística descritiva: Resultados | 53 |
| 2.9 | Taxas médias de rendibilidade diária | 56 |
| 2.10 | Análise de <i>outliers</i> : Resultados | 58 |
| 3.1.1 | Conceito frequencista de probabilidade | 67 |
| 3.4.1 | Função de distribuição | 107 |
| 3.4.2 | Quantis | 108 |
| 3.5.1 | Medidas de estatística descritiva | 114 |
| 3.5.2 | Medidas de estatística descritiva: $N(0,1)$ | 116 |
| 3.5.3 | Medidas de estatística descritiva: $T(0, 50, 10)$ | 119 |
| 3.5.4 | Medidas de estatística descritiva: $\text{Exp}(\lambda = 0.5)$ | 122 |
| 4.9.1 | Dimensão da amostra: cálculos | 161 |
| 4.9.2 | Dimensão da amostra: resultados | 161 |
| 5.2.1 | Resultados do teste F (igualdade de variâncias) | 180 |
| 5.2.2 | Resultados do teste F (igualdade de variâncias) | 181 |
| 5.2.3 | Teste T para as variáveis $L1CFDepois$ e $L1SFDepois$ | 183 |

| | | |
|-------|--|-----|
| 5.2.4 | Teste T para a diferença de médias | 185 |
| 5.2.5 | Resultados do teste T para amostras emparelhadas (L1CF) | 187 |
| 6.1.1 | Variâncias e Covariâncias | 201 |
| 6.1.2 | Coefficiente de correlação simples | 202 |
| 6.2.1 | Coefficiente de Spearman: ordenação dos valores | 207 |
| 6.3.1 | Número e receita dos turistas | 208 |
| 6.3.2 | Salários por anos de escolaridade | 214 |
| 6.3.3 | Média condicional, valor estimado, erros e resíduos | 218 |
| 6.3.4 | Cálculos preliminares para estimar β_1 e β_2 | 223 |
| 6.3.5 | Modelo de regressão linear simples: resultados | 241 |
| 6.3.6 | Resíduos OLS | 244 |
| 7.7.1 | Efeitos marginais decrescentes | 267 |
| 7.A.1 | Formas funcionais, efeitos marginais e elasticidades | 272 |
| 8.3.1 | Modelo de regressão linear múltipla: resultados | 290 |
| 8.4.1 | Coefficientes de correlação <i>part</i> : Resultados | 299 |
| 8.4.2 | Tolerância e VIF | 299 |
| 8.4.3 | Correlação e Multicolinearidade | 300 |
| 8.4.4 | Modelo de regressão linear múltipla: sem variável rendimento | 301 |
| 8.4.5 | RL <i>vs</i> VN e TRAB | 309 |
| 8.4.6 | Valores estimados de Y e Resíduos OLS | 309 |
| 8.4.7 | Teste de White | 310 |

Lista de Figuras

| | | |
|-------|--|------|
| 1 | Rendibilidade e Risco de um Fundo de Ações | xiii |
| 1.1 | Comando <code>Data Analysis</code> (Análise de Dados) | 8 |
| 1.2 | Instalação do suplemento <code>ANALYSIS TOOLPACK</code> | 9 |
| 1.3 | Janela de comandos estatísticos | 10 |
| 1.4 | Instalação de suplemento extra | 11 |
| 1.5 | Suplemento <code>REALSTATS</code> | 11 |
| 1.6 | Introdução de função | 12 |
| 1.7 | Funções estatísticas | 12 |
| 1.8 | Função <code>FREQUENCY/FREQUÊNCIA</code> | 15 |
| 1.9 | Frequências | 16 |
| 1.10 | Histograma: Introdução de dados | 19 |
| 1.11 | Histograma: Desempenho | 19 |
| 1.12 | Histograma: Salário | 19 |
| 1.13 | Histograma: Salário | 20 |
| 2.1 | Medidas de estatística descritiva: cálculo | 33 |
| 2.2 | Outliers severos e moderados | 41 |
| 2.3 | Tipos de assimetria | 43 |
| 2.4 | Efeito do “peso” das caudas na assimetria | 45 |
| 2.5 | Tipos de curtose | 46 |
| 2.6 | Histograma: Introdução de dados | 51 |
| 2.7 | Histograma: PSI20 | 51 |
| 2.8 | Medidas de estatística descritiva: Cálculo | 53 |
| 2.9 | Análise de outliers: Cálculo | 57 |
| 2.10 | Medidas de estatística descritiva: tabela dinâmica | 59 |
| 2.11 | Medidas de estatística descritiva: tabela dinâmica 1 | 60 |
| 2.12 | Medidas de estatística descritiva: tabela dinâmica 2 | 60 |
| 3.1.1 | Frequências relativas do PSI20, SONAE e REN | 68 |
| 3.2.1 | Função densidade de probabilidade | 71 |
| 3.2.2 | Funções densidade de probabilidade: $f(x)$, e de distribuição: $F(x)$, de uma variável aleatória com distribuição normal estandardizada | 73 |

| | |
|--|-----|
| 3.4.1 Distribuição das vendas (milhares de euros) | 80 |
| 3.4.2 Função BINOM.DIST | 86 |
| 3.4.3 Função NEGBINOM.DIST | 89 |
| 3.4.4 Função densidade de probabilidade de uma variável aleatória com distribuição normal | 94 |
| 3.4.5 Distribuição normal: cálculo de probabilidades em Excel | 97 |
| 3.4.6 RETPSI20 – Introdução da linha | 102 |
| 3.4.7 RETPSI20 – Histograma com sobreposição da curva Normal | 103 |
| 3.4.8 Distribuição qui-quadrado: função densidade de probabilidade | 105 |
| 3.4.9 Distribuição t de Student: função densidade de probabilidade para diferentes graus de liberdade | 107 |
| 3.4.10 Distribuição F : função densidade de probabilidade para diferentes graus de liberdade | 108 |
| 3.5.1 Função densidade de probabilidade $U(2,4)$ | 112 |
| 3.5.2 Dados, Análise de Dados, Histograma | 115 |
| 3.5.3 Histograma: distribuição $U(0,1)$ | 115 |
| 3.5.4 Histograma: distribuição $N(0,1)$ | 116 |
| 3.5.5 Função densidade de probabilidade $T(0, 12, 4)$ | 118 |
| 3.5.6 Histograma: distribuição $T(0, 50, 10)$ | 120 |
| 3.5.7 Função densidade de probabilidade $\text{Exp}(\lambda = 0.5)$ | 120 |
| 3.5.8 Histograma: distribuição $\text{Exp}(\lambda = 0.5)$ | 122 |
| 4.2.1 Amostra aleatória | 133 |
| 4.2.2 Geração aleatória | 134 |
| 4.4.1 Amostras aleatórias | 140 |
| 4.4.2 Amostras aleatórias | 140 |
| 4.4.3 Histograma: introdução de dados | 141 |
| 4.4.4 Histograma: introdução de dados da distribuição normal | 141 |
| 4.4.5 Histograma com sobreposição da curva normal | 142 |
| 5.1.1 Regiões de Não Rejeição (RNR) e de Rejeição (RR) | 167 |
| 5.1.2 Testes Unilateral Esquerdo e Unilateral Direito | 168 |
| 5.1.3 Regiões de Não Rejeição (RNR) e de Rejeição (RR) | 171 |
| 5.1.4 Cálculo da probabilidade associada ao valor de Z | 172 |
| 5.1.5 Região de Não Rejeição (RNR) e de Rejeição (RR) | 175 |
| 5.2.1 Índices de cumprimento antes, durante e depois da formação | 177 |
| 5.2.2 Teste F à igualdade de variâncias | 180 |
| 5.2.3 Teste t à igualdade das médias | 182 |
| 5.2.4 Teste T para amostras emparelhadas | 186 |
| 6.1.1 Diagramas de dispersão | 196 |
| 6.1.2 Diagrama de dispersão | 197 |

| | | |
|--------|--|-----|
| 6.1.3 | Diagrama de dispersão SONAE-PSI20: procedimento de cálculo | 199 |
| 6.1.4 | Diagrama de dispersão SONAE-PSI20: gráfico | 200 |
| 6.1.5 | Covariância: Introdução de dados | 200 |
| 6.1.6 | Coefficiente de correlação: Introdução de dados | 201 |
| 6.1.7 | Diagrama de dispersão | 204 |
| 6.3.1 | Evolução da Receita e do N° de Turistas | 209 |
| 6.3.2 | Diagrama de dispersão | 210 |
| 6.3.3 | Retas “alternativas” | 211 |
| 6.3.4 | Distribuição condicional dos salários por anos de escolaridade | 215 |
| 6.3.5 | Funções de regressão da população e da amostra | 219 |
| 6.3.6 | Média condicional dos erros | 227 |
| 6.3.7 | Variância condicional dos erros (homoscedasticidade) | 229 |
| 6.3.8 | Variância condicional dos erros (heteroscedasticidade) | 230 |
| 6.3.9 | Ausência de autocorrelação | 231 |
| 6.3.10 | Modelo de regressão linear simples: Estimação | 240 |
| 6.3.11 | Normalidade dos erros | 245 |
| 6.3.12 | Homoscedasticidade e Autocorrelação | 245 |
| 7.2.1 | Modelo lin-log | 252 |
| 7.2.2 | yields - maturidade | 253 |
| 7.2.3 | yields - logaritmo da maturidade | 253 |
| 7.3.1 | PIB Norte-Americano | 257 |
| 7.4.1 | Modelo log-log | 260 |
| 7.5.1 | Risco de uma carteira de títulos | 262 |
| 7.7.1 | Funções Linear e Quadrática | 266 |
| 7.7.2 | Valores estimados da variável dependente | 269 |
| 7.B.1 | Função Exponencial | 273 |
| 7.B.2 | Função Logarítmica | 274 |
| 8.3.1 | Modelo de regressão linear múltipla: Estimação | 289 |
| 8.4.1 | Correlação: Introdução de dados | 298 |
| 8.4.2 | Resíduos ao quadrado <i>vs</i> valores estimados de RL | 309 |
| 8.4.3 | Autocorrelação dos erros | 314 |
| 8.4.4 | Regiões da estatística de Durbin-Watson | 315 |

Prefácio

E um livro de “Estatística” continua a ter interesse para alguns leitores, pois já vamos na terceira edição. Muito obrigado a todos os que contribuem para o sucesso deste projeto.

Os comentários que tenho recebido de alunos, colegas do ensino e de profissionais, que encontram no livro uma ferramenta de trabalho capaz de responder às suas necessidades, têm sido muito importantes para todas as edições do livro (as que já “eram” e as que estão para vir). Transcrevo a seguir dois desses comentários (que muito agradeço): *Relativamente à obra, com a qual muito aprendi, apreciei e me delicieei, felicito-o pelo encadeamento singular, pela distinta elegância e simplicidade da linguagem usada, o que permite que a assimilação, compreensão e absorção da matéria abordada se torne mais inteligível e assim acessível a um público mais alargado. Para esta acessibilidade, foram cruciais as ‘Aplicações’ que da forma simples e rigorosa como foram expostas, o público, à partida, menos preparado, familiarizado, apático e renitente, dão conta da ‘GIGANTE’ importância da Estatística, que, lamentavelmente, vê várias vezes o seu valor subestimado em grande medida.* (Flávio Nunes, FEUC).

Livro muito honesto que trata de assuntos importantes de uma maneira simples e intuitiva. Muito bem organizado e com um forte apoio prático. Muito útil para cursos e escolas de índole muito prática. Parabéns. (Diogo Júdice, ISCAC).

Nesta edição (a terceira) do livro as alterações não são muito significativas (uma palavra muito “cara” à estatística, não acha?). “Aprimoramos” a explicação sobre algumas propriedades dos estimadores e corrigimos gralhas e problemas no ficheiro de dados detetados pela Ângela Lopes (que está muito longe, no Brasil). A Sisi Marques também deu uma ajuda importante ao ler a versão (quase) final desta nova edição.

O livro continua a ser uma edição de autor, e isto tem vantagens e desvantagens. Não posso aproveitar a “máquina promocional” de uma Editora e o livro não está à venda em livrarias. Ainda fiz uma tentativa (não vou dizer o nome da livraria para não “se sentirem atingidos”) mas queriam cobrar 50% sobre o preço de capa. Um abuso, não acham? E a venda era à consignação... Triste sina a de quem escreve e publica livros em Portugal... Portanto, as vendas continuam a depender muito do “passa-palavra” e uma vez mais apelo à vossa generosidade para divulgação do livro.

Quem quiser adquirir o livro pode dirigir-se diretamente a mim (pessoalmente ou através do e-mail: dias.curto@iscte-iul.pt) ou fazendo o pedido através da página do livro no meu site:

<https://diascurto.wixsite.com/sitedc/estatistica>

E qual é a grande vantagem de ser eu a publicar? O preço, que é substancialmente mais baixo do que aquele que seria praticado por uma Editora.

E “prontos” resta-me agradecer novamente a todos os que têm comprado o livro e contribuído para a sua divulgação e... está na hora de “embarcar” nesta viagem de conhecimento. Espero que sejam bem sucedidos na aprendizagem e que, de uma vez por todas, passem a “adorar” a Estatística. Qualquer dúvida é só enviar para o meu e-mail: dias.curto@iscte-iul.pt.

Até breve e que a nova edição não tarde a aparecer.

Lisboa, setembro de 2019.

Prefácio à segunda edição

A primeira edição do livro foi bem aceite e muitos contribuíram para o sucesso deste projeto: os meus alunos (mais os que frequentam os mestrados executivos do que os alunos das licenciaturas e dos mestrados de continuidade), formandos de “outras andanças”, com destaque para os Revisores Oficiais de Contas que têm tido uma grande consideração pelo meu trabalho, colegas e alunos de outras escolas e muitos profissionais de diversas áreas: Imobiliário, Banca, Contabilidade, Auditoria, Finanças, Gestão portuária, Indústria do papel, etc. Ao longo deste último ano recebi também *feedback* e respondi sempre às dúvidas e questões que me foram colocadas. Por tudo isto, e a todos aqueles que me apoiaram, só me resta dizer MUITO OBRIGADO!

E porque a vida não para, é tempo de avançar com a segunda edição. As alterações em relação à primeira edição não são muito significativas mas ainda assim convém referir:

- O título do livro. “Estatística”, o título anterior, parece descrever um manual para alunos de licenciatura que têm de perceber umas coisitas de estatística e fazer uns exercícios para passar à cadeira. Mas o livro é isso e muito muito mais! São resolvidas dezenas de aplicações com dados reais que permitem ao leitor perceber como utilizar a estatística no processo de tomada de decisão e esta é a razão do novo título “Potenciar os negócios? A Estatística dá uma ajuda!”.
- Na secção 2.4 esclarece-se melhor a diferença entre “curtose” e “excesso de curtose” de uma distribuição, introduzindo-se um gráfico com as funções densidade de probabilidade relativas aos três tipos de distribuição: leptocúrtica, mesocúrtica e platicúrtica.
- Alguns capítulos (ou secções) começam agora com citações da vida real publicadas em jornais e identificadas com a temática estatística respetiva.
- A capa também foi renovada com a preciosa ajuda da Filipa a quem agradeço mais uma vez.
- Por fim, corrigimos as gralhas (com a ajuda de alguns leitores) que foram sendo detetadas ao longo deste último ano.

Importa ainda renovar um pedido de desculpas. Na escrita do livro utilizo o Latex, e por não ter um corretor automático de português, é muito provável que, mesmo depois de toda a minha atenção e revisão, ainda apareçam gralhas.

Por fim, a todos os que tornaram possível esta nova edição, o meu muito obrigado e conto com todos para a divulgação deste trabalho.

Lisboa, novembro de 2017

Prefácio à primeira edição

Há já vários anos que leciono e coordeno disciplinas (agora designadas unidades curriculares, UCs) em cursos de pós-graduação e em mestrados executivos. As designações são variadas (Estatística Aplicada, Técnicas Quantitativas para Finanças, Pesquisa e Análise de Dados, Análise de Dados e Informação, Métodos Quantitativos Aplicados às Vendas, Métodos Quantitativos Aplicados, etc.) mas todas “andam à volta” da matemática, da estatística e da econometria, as matérias mal-amadas da maior parte dos estudantes. Não é que eles não vejam interesse nos temas lecionados, mas o problema “são as bases”. Assustam-se com a equação da curva de Gauss, estranham a palavra “ln”, perguntam “como é que passo para o outro lado da equação” e pedem, quase sempre, “para explicar outra vez e/ou mais devagar”.

Muitos já não têm matemática desde o 9º ano, outros tiveram até ao 12º ano mas já não se lembram de quase nada e muitos já “deram tudo” mas o tempo encarregou-se de lhes limpar a memória.

Perante este cenário, e quando as aulas começam, muitos dos alunos têm uma expectativa quase “catastrófica” sobre o resultado final: “O que é que vai acontecer?”, “Se calhar vou chumbar...” E fazem muitas perguntas “O professor chumba muitos alunos?”, “As notas são muito baixas?”, “É preciso saber muita matemática? Eu já não me lembro de quase nada...”, “E estatística? Já foi há tantos anos...”

Mas depois da primeira aula, e pelo menos até bem próximo do exame, a calma regressa e quase todos admitem que “sou capaz!”. Deve ser porque o método de leção lhes transmite coragem e confiança. A exposição das matérias, mais baseada em “números” do que “em letras”, relativiza o carácter abstrato das letras e as fórmulas complexas, “assustadoras” e “desmobilizadoras” reduzem-se ao que é essencial. Para além disto, e antes de qualquer fórmula se revelar intimidatória, há sempre uma explicação prévia que a contextualiza na matéria e que permite, através de palavras, compreender a sua necessidade e perceber todos (ou quase todos) os elementos (abstratos) que a compõem. Como já devem ter reparado, a utilização da palavra “fórmulas” é uma forma suave de me referir à matemática...

Mas para a acalmia ser definitiva é ainda necessário dar muitos exemplos de situações reais onde as matérias lecionadas podem ser aplicadas (e assim realçar

a utilidade respetiva) e libertá-los da necessidade de “fazer contas” sempre que os cálculos envolvem fórmulas mais complexas. Essa tarefa fica a “carga” do computador e dos programas utilizados: nas aulas os alunos aprendem a trabalhar com o suplemento **Analysis ToolPack** do Excel e com o SPSS. Outros ainda, dependendo do curso e da UC lecionada, utilizam também o EViews. Sempre que é possível as aulas decorrem num laboratório de informática com computadores à disposição, mas há também cursos em que os alunos são convidados a trazer o computador pessoal e a instalar o SPSS. Em qualquer dos casos os alunos “servem-se” do computador para produzir resultados a partir dos programas utilizados.

“**É ou não é**” este o método que mais serve a quem frequenta uma pós-graduação ou um mestrado executivo? Tenho a certeza que sim. Para estes alunos e para as empresas onde trabalham, e “**espero que estejam a perceber**”, os fundamentos matemáticos não são assim tão importantes. O que importa é que saibam utilizar os programas para produzir informação a partir dos dados disponíveis e que sejam capazes de interpretar os resultados de forma a tirar o máximo proveito dessa informação. E “**está claro**” que este é o objetivo principal do método que utilizo.

E agora o porquê deste livro. Ao longo destes anos tenho preparado e disponibilizado materiais à luz deste método: fórmulas quanto baste, cópias de janelas do Excel (para ilustrar os comandos) e texto muito objetivo no sentido de apresentar conceitos, descrever métodos e comentar resultados. Penso que chegou o momento de lhes “dar uma arrumação” e apresentá-los sob a forma de livro esperando que, pela sua simplicidade e objetividade, possa constituir um instrumento de trabalho para todos aqueles que utilizam a estatística como método de apoio à tomada de decisão.

Termino este prefácio agradecendo, primeiro, aos grandes mestres e professores (portugueses) por me terem ajudado a perceber “alguma coisa” de estatística: Bento de Jesus Caraça, Tiago de Oliveira, Bento Murteira, Daniel Müller, Silva Ribeiro, Filomena Pimenta, Dinis Pestana, Elizabeth Reis e Ivette Gomes. E em segundo, a todos os meus alunos pela estima, dedicação e carinho que me merecem. Espero poder continuar a contar com eles (os que já foram e os que hão-de vir a ser) para comentar, encontrar “coisas” menos explícitas, identificar erros (se os houver) e contribuir assim para uma nova edição ainda mais de acordo com as necessidades dos leitores: profissionais que encaram os métodos estatísticos não como um “bicho de sete cabeças” mas como ferramentas importantes na análise e tratamento de dados. Ainda um obrigado muito especial à Filipa pela capa bonita que “inventou”.

Lisboa, setembro de 2016

A estatística sempre “por perto”

“SEM ESTATÍSTICAS ANDAMOS PERDIDOS NA SOCIEDADE EM QUE VIVEMOS.” (Maria João Valente, diretora da PORDATA, (ECO, 08/06/2017).

Todos os dias recorremos à matemática e à estatística, mesmo que não nos demos conta. No supermercado utilizamos a *soma* para saber quanto gastámos e a *diferença* para calcular o troco. Na bomba de gasolina *multiplicamos* o preço pelos litros de combustível para saber quanto é que vão custar os próximos quilómetros a andar de carro. Em casa *dividem-se* as 8 pastilhas pelos quatro filhos do casal (apesar do mais velho reclamar uma quantidade *maior...*).

Mas a matemática não se esgota nestas coisas simples da vida. Muito do “bem bom” que temos hoje, e que aumentou de forma considerável nos últimos anos, deve-se à matemática. Sim, não tenham dúvidas, mas como não estamos aqui para destacar as “benfeitorias” da matemática, sugiro que leiam os livros de Guillen (1995) e Stewart (2003) para me darem razão¹.

Apesar de não ser fácil separar a estatística da matemática, o nosso enfoque é a estatística e o grande objetivo é evidenciar como ela “faz parte das nossas vidas”.

Começamos pelo Totoloto (apesar do Euromilhões estar na moda). “Joguei outra vez esta semana e... nada! Não tenho sorte nenhuma.” Pois... a *probabilidade* de sair o totoloto, por cada aposta jogada, é de 1 em 13983816, ou seja, 0.00000007151, e “é quase impossível” ser “bafejado” pela sorte (a Santa Casa, e os que dela dependem, que não nos ouçam). Aliás, há mesmo quem diga que é maior a *probabilidade* de ser atropelado por um carro do que sair o totoloto. Mas há também quem diga “que ele sai, sai!” Portanto, não é um *acontecimento impossível*.

Quando vamos ao futebol queremos que o nosso clube ganhe. Mas nunca sabemos antes do jogo se tal vai acontecer. Para a vitória ser *certa*, sempre se pode “comprar” o árbitro... Mas, independentemente do árbitro e da capacidade dos nossos jogadores, há sempre fatores *aleatórios* (imprevisíveis, mas que ainda

¹Já Galileu (1564-1642) no distante século XVII acreditava que a matemática seria a chave para a compreensão do universo. Newton, Bernoulli, Faraday, Clausius e Einstein, entre tantos outros, viriam a dar-lhe razão. Para Galileu, o grande livro do universo estava escrito em linguagem matemática.

assim têm uma probabilidade de acontecer) que influenciam o resultado (por exemplo, o nosso melhor avançado lesionou-se e a equipa não marcou golos). Portanto, quando se entra no estádio a *esperança* é grande mas a *incerteza* no resultado ainda é maior.

Probabilidade, *aleatório* e *acontecimento* são termos comuns que têm fundamento na estatística, como se explica mais adiante. *Certo*, *incerto* e *esperança* também não lhes são estranhos. Mas não se fica por aqui...

Quando pergunta ao filho (ou ao amigo) a nota que teve no exame de física e ele lhe responde: 12, não há razões para grande entusiasmo. Mas se ele disser que a *média* da turma foi 6, aí a “coisa” muda de figura e até lhe pode dar um abraço de parabéns!

A *média* é talvez o melhor exemplo da nossa “convivência” com a estatística pela frequência com que é utilizada. “Quantos quilos tem este saco de laranjas?” “O fornecedor disse que o peso médio de cada saco são dois quilos.” “Em média quanto tempo demoras a chegar à escola?” “Cerca de 30 minutos.” Até o banco regista o saldo médio da sua conta bancária...

E já que falamos de bancos, o que vai fazer com os 5000 euros que poupou no último ano? Se não os gastar, sempre pode fazer uma aplicação que lhe proporcione algum rendimento. Na hora de decidir revela-se o quanto é afoito e destemido (aquilo que em finanças se designa por *aversão ao risco*). Se optar por um depósito a prazo, o banco dá-lhe 2% ao fim de um ano (que representam 100 euros). Se comprar obrigações da empresa BAUHAUS, a taxa do cupão é 4% (o dobro em euros: 200) e pode também comprar, à cotação de 5 euros, 1000 unidades de participação (UP) do fundo de ações CURE de um banco a operar em Portugal (a informação que consta do prospeto do fundo, e que nos interessa, é apresentada na figura 1).

Neste caso não é possível antecipar o valor do rendimento pois depende da cotação das UP no momento da venda: pode ganhar e ter uma *mais-valia*, perder e incorrer numa *menos-valia* ou “ficar em casa” se a cotação das UP estiver acima, abaixo ou nos 5 euros, respetivamente. Se a cotação estiver nos 6 euros, por exemplo, tem uma rendibilidade de 20%. Fantástico, vai ganhar 1000 euros!... Mas também pode acontecer o contrário e o preço baixar para os 4 euros. Neste caso perde 1000 euros e, ao final de um ano, dos 5000 euros iniciais restam apenas 4000. Ou seja, do depósito a prazo para as ações o *risco* do seu investimento vai aumentando.

O termo *risco* pode até nem lhe ser familiar, mas percebe, de certeza, que a possibilidade de ter um rendimento elevado é maior nas ações. Aliás, se atentar na informação do prospeto, a rendibilidade do fundo foi de 27.34% e 8.47% nos últimos 12 e 24 meses, respetivamente, bem acima dos 2% e 4% oferecidos pelo depósito bancário e pelas obrigações. Mas leia também a nota que se apresenta: *as rendibilidades divulgadas representam dados passados, não constituindo ga-*

Figura 1: Rendibilidade e Risco de um Fundo de Ações

| | | | |
|---------------|-----------------------|------------------------------------|-----------------------------------|
| Data Cotação: | Cotação UP: | Rentab. 12M ² : 27,34 % | Rentab. 24M ² : 8,47 % |
| | Data Rentab. e Risco: | Risco 12M: 13,55 % | Risco 24M: 17,26 % |

¹ Classe de Risco referente a 12 meses:

| Desvio-padrão anualizado (%) | Classe de risco | Escalão de risco |
|------------------------------|-----------------|-------------------|
| [0 ; 1,5[| 1 | risco baixo |
| [1,5 ; 5[| 2 | risco médio baixo |
| [5 ; 10[| 3 | risco médio |
| [10 ; 15[| 4 | risco médio alto |
| [15 ; 20[| 5 | risco alto |
| >=20 | 6 | risco muito alto |

² Estas rentabilidades não consideram a retenção de IRS que existe no resgate

NOTAS:

^{1*} - As Rendibilidades divulgadas representam dados passados, não constituindo garantia de rentabilidade futura, porque o valor das unidades de participação pode aumentar ou diminuir em função do nível de risco que varia entre 1 (risco mínimo) e 6 (risco máximo).

rentia de rentabilidade futura. Aliás, também está escrito que o valor das UP pode cair e... neste caso incorre numa perda. Portanto, é mais “arriscado” aplicar os 5000 euros no fundo de ações, pois com o depósito a prazo ou as obrigações estão garantidos o capital inicial mais 100 ou 200 euros de juros, respetivamente (admitindo que o banco e a empresa BAUHAUS não vão à falência). Os financeiros chamam *risco* à possibilidade “da coisa correr mal” e de perder parte (ou até mesmo a totalidade) do capital investido. Nos casos do BES e do BANIF, por exemplo, os acionistas perderam tudo...

Ora bem, o risco “é das finanças”¹ mas “são da estatística” as formas de o quantificar. A medida mais utilizada é o *desvio-padrão*, como se apresenta mais adiante, e antes de ser uma medida de risco é uma medida de *dispersão*, designação dada pela estatística a um conjunto de medidas para aferir sobre a variação do fenómeno em causa. Se “olhar” novamente para a informação do prospeto, e mesmo que não saiba o que é o desvio-padrão (mas que ficará a saber mais adiante), as colunas *Classe de risco* e *Escalão de risco* são bem claras sobre o “perigo” que é investir no fundo de ações. Nos últimos 2 anos o risco foi superior a 17% (é como se pudéssemos dizer, apesar de não ser totalmente correto, que o valor das UP tanto pode subir como descer 17%) e está na penúltima classe de risco (*risco elevado*). Para melhor contextualizar os 17%, e admitindo que o banco do depósito a prazo e a empresa BAUHAUS não vão à falência, ao fim de um ano receberá os 5000 euros mais os juros vencidos. Portanto, o risco é 0%, que compara com os 17% do fundo de ações.

Lá estão o risco e a estatística por detrás de uma decisão importante que

¹Da meteorologia, da medicina, da engenharia...

pode tomar: comprar ou não as unidades de participação daquele fundo.

E para não me tornar chato ao “defender a minha dama” deixo apenas mais três exemplos da nossa familiaridade com a estatística. “No dia 9 de agosto a temperatura em Vilamoura atingiu os 40°, coisa muito rara nos últimos 20 anos”. Em estatística estes valores pouco frequentes merecem uma atenção especial e designam-se por *outliers*¹, se quisermos utilizar o termo em inglês, ou valores extremos. “A amplitude térmica (AT) diária na Beira Baixa é muito elevada”. A AT (não confunda com Autoridade Tributária...) é a diferença entre as temperaturas máxima e mínima diárias e na estatística esta diferença constitui também uma medida de dispersão designada por *intervalo de variação*. “Na minha empresa cerca de 95% dos trabalhadores ganham entre 1200 e 3800 euros”. O que quer dizer que o salário dos restantes 5% é inferior a 1200 ou superior a 3800 euros. Não sei se o fez, mas para chegar à conclusão este “patrão” pode ter recorrido à *distribuição normal*, mais um “instrumento” muito importante da estatística.

E podíamos continuar... mas penso que já os convenci da “convivência” e da “importância” da estatística no nosso dia a dia. E prometo que vai descobrir muitas outras “coisas” onde a estatística já é, ou virá a ser, importante para a sua tomada de decisão. A aplicação dos 5000 euros já foi um princípio, mas para já “vamos” à estrutura deste livro.

No próximo capítulo começa-se (de forma ligeira...) a falar dos dados, o objeto da estatística: classificação e formas de apresentação. Define-se variável e prepara-se o Excel para utilização futura (sempre na perspectiva da estatística...). Calculam-se e interpretam-se as frequências absolutas e relativas, o valor da moda e constrói-se o primeiro histograma.

No capítulo 2 introduz-se e aplica-se a generalidade das medidas de estatística descritiva para ajudar a família Pereira na compra de um automóvel, para analisar as vendas e os resultados líquidos das 30 maiores empresas de sistemas de informação em Portugal no ano 2011 e para avaliar o risco e a rentabilidade do índice PSI20 e das ações das empresas SONAE e REN.

No capítulo 3 discutem-se e aplicam-se os instrumentos estatísticos para lidar com os fenómenos de natureza aleatória, introduzindo os conceitos de probabilidade, variável aleatória e distribuição de probabilidade. A seguir descrevem-se com algum detalhe e aplicam-se as distribuições teóricas mais importantes: Binomial, Poisson, Normal, t-Student, F-Snedecor, etc. O capítulo termina com a simulação de vendas de uma empresa considerando-se as distribuições Triangular, Normal e Exponencial para modelo probabilístico teórico das quantidades vendidas de três produtos.

A distinção entre amostra e população constitui o mote do capítulo 4. Se não se consegue chegar a toda a população, pelo menos pode recolher-se uma

¹Apesar de ser um termo em inglês, e por aparecer de forma recorrente no texto, a partir de agora a palavra outlier vai deixar de ser apresentada em itálico.

amostra, o mais representativa possível daquela população. Explica-se também a diferença entre parâmetro e estatística e fala-se de distribuições por amostragem das estatísticas mais importantes: média, variância e proporção amostrais, não esquecendo o teorema do limite central (que ajuda que ele nos dá!...). A partir das distribuições por amostragem é possível deduzir intervalos de confiança para os parâmetros e calcular também a dimensão de uma amostra aleatória simples tendo em conta a dimensão da população (se conhecida), o nível de confiança e o erro amostral em que se pretende incorrer.

O capítulo 5 trata dos testes de hipóteses, começando por apresentar as abordagens do intervalo de confiança e dos testes de hipóteses na realização de inferências sobre determinada população. Distinguem-se ensaios de significância de testes de hipóteses e explica-se o significado dos erros tipo I e II na tomada de decisão. Por último, procede-se à descrição e aplicação de alguns testes, nomeadamente o teste t para a diferença de médias com o propósito de avaliar o efeito das ações de formação nas vendas dos lojistas de uma empresa portuguesa, e os testes mais populares (Jarque-Bera e Kolmogorov-Smirnov) para aferir sobre a normalidade da distribuição de uma variável aleatória.

Nos três capítulos seguintes “andamos às voltas” com a relação entre grandezas económicas e financeiras. Se o preço de um bem aumentar, qual é o efeito expectável sobre a quantidade vendida? Bem, se não for um daqueles bens mesmo essenciais (pão, por exemplo), as vendas devem baixar (que se cuidem as empresas fornecedoras). No capítulo 6 apresentam-se e aplicam-se as análises de correlação e regressão linear simples. Com a primeira “consegue-se” avaliar o tipo (linear ou não linear), o sentido (direta ou inversa) e a intensidade (forte ou fraca) da relação. Para isso costuma recorrer-se a uma representação gráfica (diagrama de dispersão) e a uma medida (coeficiente de correlação simples). A análise de regressão linear, por ser mais completa, permite avaliar como é que a variação de uma variável impacta na variação da outra (se o preço do pão aumentar 10 cêntimos, qual é a redução expectável no valor da vendas). Neste capítulo discute-se ainda o método dos mínimos quadrados ordinários, as propriedades dos estimadores, os pressupostos do modelo de regressão linear simples e a inferência estatística através dos testes t e F .

Uma vez que as relações não lineares são também bastante comuns entre variáveis de natureza económica e financeira, no capítulo 7 são analisadas outras funções de regressão que, apesar de constituírem relações não lineares na sua forma original, podem ser convertidas em funções lineares nos parâmetros através de transformações adequadas (a transformação logarítmica é a mais utilizada). Para isso são propostas formas funcionais alternativas para a função de regressão conhecidas vulgarmente por lin-lin, log-lin, lin-log e log-log. Também se analisam relações polinómicas e inversas bem como a alteração na escala dos dados e o seu impacte nas estimativas dos mínimos quadrados ordinários.

E o modelo de regressão linear múltipla (capítulo 8) põe fim a esta pequena “odisseia”... É um modelo mais completo (com mais variáveis explicativas) para “responder” melhor à complexidade da vida real. Fala-se outra vez do método dos mínimos quadrados ordinários, do R^2 , dos testes t e F e introduzem-se novas medidas: coeficiente de determinação ajustado, coeficientes estandardizados, coeficientes de correlação parcial, etc. Por fim, avaliam-se de forma mais formal os pressupostos do modelo de regressão linear através dos procedimentos estatísticos mais comuns (quase sempre baseados em testes de hipóteses).

E mesmo mesmo a acabar... faz-se um apanhado das referências bibliográficas. Espero não me ter esquecido de nenhuma, mas se tal acontecer, as minhas sinceras desculpas a quem ficou injustamente de fora.