

brief contents

- 1 ■ Laying the groundwork 1
- 2 ■ Exploring probability and counting 16
- 3 ■ Exploring probability distributions and conditional probabilities 41
- 4 ■ Fitting a linear regression 79
- 5 ■ Fitting a logistic regression 111
- 6 ■ Fitting a decision tree and a random forest 140
- 7 ■ Fitting time series models 184
- 8 ■ Transforming data into decisions with linear programming 222
- 9 ■ Running Monte Carlo simulations 242
- 10 ■ Building and plotting a decision tree 271
- 11 ■ Predicting future states with Markov analysis 294
- 12 ■ Examining and testing naturally occurring number sequences 319
- 13 ■ Managing projects 349
- 14 ■ Visualizing quality control 378

contents

<i>preface</i>	<i>xiii</i>
<i>acknowledgments</i>	<i>xv</i>
<i>about this book</i>	<i>xvii</i>
<i>about the author</i>	<i>xxiv</i>
<i>about the cover illustration</i>	<i>xxv</i>

1 Laying the groundwork 1

1.1 Stats and quant 2

Understanding the basics 2 ■ *Why they matter* 2 ■ *The broader effect* 3 ■ *Diving deeper: Core concepts* 3

1.2 Why Python? 4

Rich ecosystem 4 ■ *Ease of learning* 5 ■ *Online support and community* 5 ■ *Industry adoption* 5 ■ *Versatility* 6

1.3 Python IDEs 6

IDLE: A starting point 6 ■ *PyCharm: A professional tool* 7
Other popular IDEs 7

1.4 Benefits and learning approach 8

From statistical measures to real-world application 9
Expanding beyond traditional techniques 9 ■ *A balanced approach to theory and practice* 10

- 1.5 How this book works 11
 - Foundational learning with exploration and practice* 12
 - Using Python for precision and efficiency* 13 ■ *Adaptable learning for diverse skill levels* 14
- 1.6 What this book does not cover 14

2 *Exploring probability and counting* 16

- 2.1 Basic probabilities 17
 - Probability types* 19 ■ *Converting and measuring probabilities* 19
- 2.2 Counting rules 22
 - Multiplication rule* 22 ■ *Addition rule* 22 ■ *Combinations and permutations* 23
- 2.3 Continuous random variables 30
 - Examples* 31 ■ *Probability density function* 31 ■ *Cumulative distribution function* 33
- 2.4 Discrete random variables 34
 - Examples* 35 ■ *Probability mass function* 36 ■ *Cumulative distribution function* 37

3 *Exploring probability distributions and conditional probabilities* 41

- 3.1 Probability distributions 42
 - Normal distribution* 42 ■ *Binomial distribution* 50
 - Discrete uniform distribution* 56 ■ *Poisson distribution* 60
- 3.2 Probability problems 63
 - Complement rule for probability* 65 ■ *Quick reference guide* 66 ■ *Applied probability: Examples and solutions* 68
- 3.3 Conditional probabilities 72
 - Examples* 72 ■ *Conditional probabilities and independence* 73
 - Intuitive approach to conditional probability* 74 ■ *Formulaic approach to conditional probability* 76

4 *Fitting a linear regression* 79

- 4.1 Primer on linear regression 81
 - Linear equation* 81 ■ *Goodness of fit* 85 ■ *Conditions for best fit* 85

4.2 Simple linear regression 87

Importing and exploring the data 88 ▪ *Fitting the model* 93
Interpreting and evaluating the results 95 ▪ *Testing model assumptions* 105

5 *Fitting a logistic regression* 111

5.1 Logistic regression vs. linear regression 113

5.2 Multiple logistic regression 114

Importing and exploring the data 115 ▪ *Fitting the model* 125
Interpreting and evaluating the results 128 ▪ *Calculating and evaluating classification metrics* 131

6 *Fitting a decision tree and a random forest* 140

6.1 Understanding decision trees and random forests 141

6.2 Importing, wrangling, and exploring the data 142

Understanding the data 143 ▪ *Wrangling the data* 144
Exploring the data 148

6.3 Fitting a decision tree 157

Splitting the data 158 ▪ *Fitting the model* 160 ▪ *Predicting responses* 161 ▪ *Evaluating the model* 161 ▪ *Plotting the decision tree* 163 ▪ *Interpreting and understanding decision trees* 164 ▪ *Advantages and disadvantages of decision trees* 173

6.4 Fitting a random forest 174

Fitting the model 175 ▪ *Predicting responses* 177 ▪ *Evaluating the model* 177 ▪ *Feature importance* 179 ▪ *Extracting random trees* 181

7 *Fitting time series models* 184

7.1 Distinguishing forecasts from predictions 185

7.2 Importing and plotting the data 186

Fetching financial data 186 ▪ *Understanding the data* 189
Plotting the data 190

7.3 Fitting an ARIMA model 191

Autoregression (AR) component 192 ▪ *Integration (I) component* 192 ▪ *Moving average (MA) component* 192
Combining ARIMA components 192 ▪ *Stationarity* 193
Differencing 195 ▪ *Stationarity and differencing applied* 197
AR and MA components 205 ▪ *Fitting the model* 207
Evaluating model fit 209 ▪ *Forecasting* 213

- 7.4 Fitting exponential smoothing models 215
Model structure 216 ▪ *Applicability* 216 ▪ *Mathematical properties* 216 ▪ *Types of exponential smoothing models* 216
Choosing between ARIMA and exponential smoothing 217
SES and DES models 217 ▪ *Holt–Winters model* 218

8 *Transforming data into decisions with linear programming* 222

- 8.1 Problem formulation 223
The scenario 224 ▪ *The challenge* 224 ▪ *The approach* 225
Feature summaries 227
- 8.2 Developing the linear optimization framework 229
Explanation of linear equations and inequalities 230
Data definition 230 ▪ *Objective function* 232
Constraints 233 ▪ *Decision variable bounds* 236 ▪ *Solving the linear programming problem* 236 ▪ *Result evaluation* 239

9 *Running Monte Carlo simulations* 242

- 9.1 Applications and benefits of Monte Carlo simulations 243
- 9.2 Step-by-step process 244
- 9.3 Hands-on approach 246
Establishing a probability distribution (step 1) 246 ▪ *Computing a cumulative probability distribution (step 2)* 248 ▪ *Establishing an interval of random numbers for each variable (step 3)* 250
Generating random numbers (step 4) 252 ▪ *Simulating a series of trials (step 5)* 253 ▪ *Analyzing the results (step 6)* 254
- 9.4 Automating simulations on discrete data 255
Plotting and analyzing the results 257
- 9.5 Automating simulations on continuous data 259
Predicting stock prices with Monte Carlo simulations 259
Analyzing historical data (step 1) 261 ▪ *Calculating log returns (step 2)* 262 ▪ *Computing statistical parameters (step 3)* 264
Generating random daily returns (step 4) 265 ▪ *Simulating prices (step 5)* 266 ▪ *Simulating multiple trials (step 6)* 267
Analyzing the results (step 7) 268

10 *Building and plotting a decision tree* 271

- 10.1 Decision-making without probabilities 272
Maximax method 273 ▪ *Maximin method* 276 ▪ *Minimax Regret method* 277 ▪ *Expected Value method* 279

- 10.2 Decision trees 282
 - Creating the schema 283* ■ *Plotting the tree 289*

11 *Predicting future states with Markov analysis 294*

- 11.1 Understanding the mechanics of Markov analysis 295
- 11.2 States and state probabilities 296
 - Understanding the vector of state probabilities for multistate systems 297* ■ *Matrix of transition probabilities 300*
- 11.3 Equilibrium conditions 307
 - Predicting equilibrium conditions programmatically 308*
- 11.4 Absorbing states 311
 - Obtaining the fundamental matrix 313* ■ *Predicting absorbing states 315* ■ *Predicting absorbing states programmatically 316*

12 *Examining and testing naturally occurring number sequences 319*

- 12.1 Benford's law explained 320
- 12.2 Naturally occurring number sequences 324
- 12.3 Uniform and random distributions 325
 - Uniform distribution 325* ■ *Random distribution 327*
 - Plotted distributions 328*
- 12.4 Examples 330
 - Street addresses 330* ■ *World population figures 333*
 - Payment amounts 336*
- 12.5 Validating Benford's law 337
 - Chi-square test 338* ■ *Mean absolute deviation 341*
 - Distortion factor and z-statistic 343* ■ *Mantissa statistics 344*

13 *Managing projects 349*

- 13.1 Creating a work breakdown structure 350
- 13.2 Estimating activity times with PERT 354
- 13.3 Finding the critical path 357
 - Earliest times 357* ■ *Latest times 359* ■ *Slack 360*
 - Finding the critical path programmatically 362*
- 13.4 Estimating the probability of project completion 369
- 13.5 Crashing the project 374

14 Visualizing quality control 378

14.1 Quality control measures 380

Upper control limit and lower control limit 380 ▪ *Mean and center line* 381 ▪ *Standard deviation* 381 ▪ *Range* 382 ▪ *Sample size* 383 ▪ *Proportion defective* 383 ▪ *Number of defective items* 384 ▪ *Number of defects* 385 ▪ *Defects per unit* 385 ▪ *Moving range* 386 ▪ *z-score* 386 ▪ *Process capability indices* 387

14.2 Control charts for attributes 388

p-charts 388 ▪ *np-charts* 392 ▪ *c-charts* 394 ▪ *g-charts* 396

14.3 Control charts for variables 398

x-bar charts 400 ▪ *r-charts* 401 ▪ *s-charts* 403 ▪ *I-MR charts* 405 ▪ *EWMA charts* 407

index 411