

---

# Contents

## Part I Describing Datasets

<b>1</b>	<b>First Tools for Looking at Data</b>	3
1.1	Datasets	3
1.2	What's Happening? Plotting Data	4
1.2.1	Bar Charts	5
1.2.2	Histograms	6
1.2.3	How to Make Histograms	6
1.2.4	Conditional Histograms	7
1.3	Summarizing 1D Data	7
1.3.1	The Mean	7
1.3.2	Standard Deviation	9
1.3.3	Computing Mean and Standard Deviation Online	12
1.3.4	Variance	12
1.3.5	The Median	13
1.3.6	Interquartile Range	14
1.3.7	Using Summaries Sensibly	15
1.4	Plots and Summaries	16
1.4.1	Some Properties of Histograms	16
1.4.2	Standard Coordinates and Normal Data	18
1.4.3	Box Plots	20
1.5	Whose is Bigger? Investigating Australian Pizzas	20
1.6	You Should	24
1.6.1	Remember These Definitions	24
1.6.2	Remember These Terms	25
1.6.3	Remember These Facts	25
1.6.4	Be Able to	25
<b>2</b>	<b>Looking at Relationships</b>	29
2.1	Plotting 2D Data	29
2.1.1	Categorical Data, Counts, and Charts	29
2.1.2	Series	31
2.1.3	Scatter Plots for Spatial Data	33
2.1.4	Exposing Relationships with Scatter Plots	34
2.2	Correlation	36
2.2.1	The Correlation Coefficient	39
2.2.2	Using Correlation to Predict	42
2.2.3	Confusion Caused by Correlation	44
2.3	Sterile Males in Wild Horse Herds	45
2.4	You Should	47
2.4.1	Remember These Definitions	47
2.4.2	Remember These Terms	47

2.4.3	Remember These Facts	47
2.4.4	Use These Procedures	47
2.4.5	Be Able to	47

## Part II Probability

<b>3</b>	<b>Basic Ideas in Probability</b>	53
3.1	Experiments, Outcomes and Probability	53
3.1.1	Outcomes and Probability	53
3.2	Events	55
3.2.1	Computing Event Probabilities by Counting Outcomes	56
3.2.2	The Probability of Events	58
3.2.3	Computing Probabilities by Reasoning About Sets	60
3.3	Independence	61
3.3.1	Example: Airline Overbooking	64
3.4	Conditional Probability	66
3.4.1	Evaluating Conditional Probabilities	67
3.4.2	Detecting Rare Events Is Hard	70
3.4.3	Conditional Probability and Various Forms of Independence	71
3.4.4	Warning Example: The Prosecutor's Fallacy	72
3.4.5	Warning Example: The Monty Hall Problem	73
3.5	Extra Worked Examples	75
3.5.1	Outcomes and Probability	75
3.5.2	Events	76
3.5.3	Independence	77
3.5.4	Conditional Probability	78
3.6	You Should	80
3.6.1	Remember These Definitions	80
3.6.2	Remember These Terms	80
3.6.3	Remember and Use These Facts	80
3.6.4	Remember These Points	80
3.6.5	Be Able to	81
<b>4</b>	<b>Random Variables and Expectations</b>	87
4.1	Random Variables	87
4.1.1	Joint and Conditional Probability for Random Variables	89
4.1.2	Just a Little Continuous Probability	91
4.2	Expectations and Expected Values	93
4.2.1	Expected Values	93
4.2.2	Mean, Variance and Covariance	95
4.2.3	Expectations and Statistics	98
4.3	The Weak Law of Large Numbers	99
4.3.1	IID Samples	99
4.3.2	Two Inequalities	100
4.3.3	Proving the Inequalities	100
4.3.4	The Weak Law of Large Numbers	102
4.4	Using the Weak Law of Large Numbers	103
4.4.1	Should You Accept a Bet?	103
4.4.2	Odds, Expectations and Bookmaking: A Cultural Diversion	104
4.4.3	Ending a Game Early	105
4.4.4	Making a Decision with Decision Trees and Expectations	105
4.4.5	Utility	106

4.5	You Should	108
4.5.1	Remember These Definitions	108
4.5.2	Remember These Terms	108
4.5.3	Use and Remember These Facts	109
4.5.4	Remember These Points	109
4.5.5	Be Able to	109
<b>5</b>	<b>Useful Probability Distributions</b>	<b>115</b>
5.1	Discrete Distributions	115
5.1.1	The Discrete Uniform Distribution	115
5.1.2	Bernoulli Random Variables	116
5.1.3	The Geometric Distribution	116
5.1.4	The Binomial Probability Distribution	116
5.1.5	Multinomial Probabilities	118
5.1.6	The Poisson Distribution	118
5.2	Continuous Distributions	120
5.2.1	The Continuous Uniform Distribution	120
5.2.2	The Beta Distribution	120
5.2.3	The Gamma Distribution	121
5.2.4	The Exponential Distribution	122
5.3	The Normal Distribution	123
5.3.1	The Standard Normal Distribution	123
5.3.2	The Normal Distribution	124
5.3.3	Properties of the Normal Distribution	124
5.4	Approximating Binomials with Large $N$	126
5.4.1	Large $N$	127
5.4.2	Getting Normal	128
5.4.3	Using a Normal Approximation to the Binomial Distribution	129
5.5	You Should	130
5.5.1	Remember These Definitions	130
5.5.2	Remember These Terms	130
5.5.3	Remember These Facts	131
5.5.4	Remember These Points	131

### Part III Inference

<b>6</b>	<b>Samples and Populations</b>	<b>141</b>
6.1	The Sample Mean	141
6.1.1	The Sample Mean Is an Estimate of the Population Mean	141
6.1.2	The Variance of the Sample Mean	142
6.1.3	When The Urn Model Works	144
6.1.4	Distributions Are Like Populations	145
6.2	Confidence Intervals	146
6.2.1	Constructing Confidence Intervals	146
6.2.2	Estimating the Variance of the Sample Mean	146
6.2.3	The Probability Distribution of the Sample Mean	148
6.2.4	Confidence Intervals for Population Means	149
6.2.5	Standard Error Estimates from Simulation	152
6.3	You Should	154
6.3.1	Remember These Definitions	154
6.3.2	Remember These Terms	154
6.3.3	Remember These Facts	154

6.3.4	Use These Procedures	154
6.3.5	Be Able to	154
<b>7</b>	<b>The Significance of Evidence</b>	159
7.1	Significance	160
7.1.1	Evaluating Significance	160
7.1.2	P-Values	161
7.2	Comparing the Mean of Two Populations	165
7.2.1	Assuming Known Population Standard Deviations	165
7.2.2	Assuming Same, Unknown Population Standard Deviation	167
7.2.3	Assuming Different, Unknown Population Standard Deviation	168
7.3	Other Useful Tests of Significance	169
7.3.1	F-Tests and Standard Deviations	169
7.3.2	$\chi^2$ Tests of Model Fit	171
7.4	P-Value Hacking and Other Dangerous Behavior	174
7.5	You Should	174
7.5.1	Remember These Definitions	174
7.5.2	Remember These Terms	175
7.5.3	Remember These Facts	175
7.5.4	Use These Procedures	175
7.5.5	Be Able to	175
<b>8</b>	<b>Experiments</b>	179
8.1	A Simple Experiment: The Effect of a Treatment	179
8.1.1	Randomized Balanced Experiments	180
8.1.2	Decomposing Error in Predictions	180
8.1.3	Estimating the Noise Variance	181
8.1.4	The ANOVA Table	182
8.1.5	Unbalanced Experiments	183
8.1.6	Significant Differences	185
8.2	Two Factor Experiments	186
8.2.1	Decomposing the Error	188
8.2.2	Interaction Between Effects	189
8.2.3	The Effects of a Treatment	190
8.2.4	Setting Up An ANOVA Table	191
8.3	You Should	194
8.3.1	Remember These Definitions	194
8.3.2	Remember These Terms	194
8.3.3	Remember These Facts	194
8.3.4	Use These Procedures	194
8.3.5	Be Able to	194
<b>9</b>	<b>Inferring Probability Models from Data</b>	197
9.1	Estimating Model Parameters with Maximum Likelihood	197
9.1.1	The Maximum Likelihood Principle	198
9.1.2	Binomial, Geometric and Multinomial Distributions	199
9.1.3	Poisson and Normal Distributions	201
9.1.4	Confidence Intervals for Model Parameters	204
9.1.5	Cautions About Maximum Likelihood	206
9.2	Incorporating Priors with Bayesian Inference	206
9.2.1	Conjugacy	209
9.2.2	MAP Inference	210
9.2.3	Cautions About Bayesian Inference	211

9.3	Bayesian Inference for Normal Distributions . . . . .	211
9.3.1	Example: Measuring Depth of a Borehole . . . . .	212
9.3.2	Normal Prior and Normal Likelihood Yield Normal Posterior . . . . .	212
9.3.3	Filtering . . . . .	214
9.4	You Should . . . . .	215
9.4.1	Remember These Definitions . . . . .	215
9.4.2	Remember These Terms . . . . .	216
9.4.3	Remember These Facts . . . . .	216
9.4.4	Use These Procedures . . . . .	217
9.4.5	Be Able to . . . . .	217

## Part IV Tools

<b>10</b>	<b>Extracting Important Relationships in High Dimensions . . . . .</b>	<b>225</b>
10.1	Summaries and Simple Plots . . . . .	225
10.1.1	The Mean . . . . .	226
10.1.2	Stem Plots and Scatterplot Matrices . . . . .	226
10.1.3	Covariance . . . . .	227
10.1.4	The Covariance Matrix . . . . .	228
10.2	Using Mean and Covariance to Understand High Dimensional Data . . . . .	231
10.2.1	Mean and Covariance Under Affine Transformations . . . . .	231
10.2.2	Eigenvectors and Diagonalization . . . . .	232
10.2.3	Diagonalizing Covariance by Rotating Blobs . . . . .	233
10.2.4	Approximating Blobs . . . . .	235
10.2.5	Example: Transforming the Height-Weight Blob . . . . .	235
10.3	Principal Components Analysis . . . . .	236
10.3.1	The Low Dimensional Representation . . . . .	236
10.3.2	The Error Caused by Reducing Dimension . . . . .	238
10.3.3	Example: Representing Colors with Principal Components . . . . .	241
10.3.4	Example: Representing Faces with Principal Components . . . . .	242
10.4	Multi-Dimensional Scaling . . . . .	242
10.4.1	Choosing Low D Points Using High D Distances . . . . .	243
10.4.2	Factoring a Dot-Product Matrix . . . . .	245
10.4.3	Example: Mapping with Multidimensional Scaling . . . . .	246
10.5	Example: Understanding Height and Weight . . . . .	247
10.6	You Should . . . . .	250
10.6.1	Remember These Definitions . . . . .	250
10.6.2	Remember These Terms . . . . .	250
10.6.3	Remember These Facts . . . . .	250
10.6.4	Use These Procedures . . . . .	250
10.6.5	Be Able to . . . . .	250
<b>11</b>	<b>Learning to Classify . . . . .</b>	<b>253</b>
11.1	Classification: The Big Ideas . . . . .	253
11.1.1	The Error Rate, and Other Summaries of Performance . . . . .	254
11.1.2	More Detailed Evaluation . . . . .	254
11.1.3	Overfitting and Cross-Validation . . . . .	255
11.2	Classifying with Nearest Neighbors . . . . .	256
11.2.1	Practical Considerations for Nearest Neighbors . . . . .	256
11.3	Classifying with Naive Bayes . . . . .	257
11.3.1	Cross-Validation to Choose a Model . . . . .	259

11.4	The Support Vector Machine	260
11.4.1	The Hinge Loss	261
11.4.2	Regularization	262
11.4.3	Finding a Classifier with Stochastic Gradient Descent	262
11.4.4	Searching for $\lambda$	264
11.4.5	Example: Training an SVM with Stochastic Gradient Descent	266
11.4.6	Multi-Class Classification with SVMs	268
11.5	Classifying with Random Forests	268
11.5.1	Building a Decision Tree: General Algorithm	270
11.5.2	Building a Decision Tree: Choosing a Split	270
11.5.3	Forests	272
11.6	You Should	274
11.6.1	Remember These Definitions	274
11.6.2	Remember These Terms	274
11.6.3	Remember These Facts	275
11.6.4	Use These Procedures	275
11.6.5	Be Able to	276
<b>12</b>	<b>Clustering: Models of High Dimensional Data</b>	<b>281</b>
12.1	The Curse of Dimension	281
12.1.1	Minor Banes of Dimension	281
12.1.2	The Curse: Data Isn't Where You Think It Is	282
12.2	Clustering Data	283
12.2.1	Agglomerative and Divisive Clustering	283
12.2.2	Clustering and Distance	285
12.3	The K-Means Algorithm and Variants	287
12.3.1	How to Choose K	288
12.3.2	Soft Assignment	290
12.3.3	Efficient Clustering and Hierarchical K Means	291
12.3.4	K-Medoids	292
12.3.5	Example: Groceries in Portugal	292
12.3.6	General Comments on K-Means	293
12.4	Describing Repetition with Vector Quantization	294
12.4.1	Vector Quantization	296
12.4.2	Example: Activity from Accelerometer Data	298
12.5	The Multivariate Normal Distribution	300
12.5.1	Affine Transformations and Gaussians	301
12.5.2	Plotting a 2D Gaussian: Covariance Ellipses	301
12.6	You Should	302
12.6.1	Remember These Definitions	302
12.6.2	Remember These Terms	302
12.6.3	Remember These Facts	303
12.6.4	Use These Procedures	303
<b>13</b>	<b>Regression</b>	<b>305</b>
13.1	Regression to Make Predictions	305
13.2	Regression to Spot Trends	306
13.3	Linear Regression and Least Squares	308
13.3.1	Linear Regression	308
13.3.2	Choosing $\beta$	309
13.3.3	Solving the Least Squares Problem	309
13.3.4	Residuals	310
13.3.5	R-Squared	310

13.4	Producing Good Linear Regressions	313
13.4.1	Transforming Variables	313
13.4.2	Problem Data Points Have Significant Impact	314
13.4.3	Functions of One Explanatory Variable	317
13.4.4	Regularizing Linear Regressions	318
13.5	Exploiting Your Neighbors for Regression	321
13.5.1	Using Your Neighbors to Predict More than a Number	323
13.6	You Should	323
13.6.1	Remember These Definitions	323
13.6.2	Remember These Terms	324
13.6.3	Remember These Facts	324
13.6.4	Remember These Procedures	324
<b>14</b>	<b>Markov Chains and Hidden Markov Models</b>	<b>331</b>
14.1	Markov Chains	331
14.1.1	Transition Probability Matrices	333
14.1.2	Stationary Distributions	335
14.1.3	Example: Markov Chain Models of Text	336
14.2	Estimating Properties of Markov Chains	338
14.2.1	Simulation	338
14.2.2	Simulation Results as Random Variables	339
14.2.3	Simulating Markov Chains	341
14.3	Example: Ranking the Web by Simulating a Markov Chain	342
14.4	Hidden Markov Models and Dynamic Programming	344
14.4.1	Hidden Markov Models	344
14.4.2	Picturing Inference with a Trellis	344
14.4.3	Dynamic Programming for HMM's: Formalities	346
14.4.4	Example: Simple Communication Errors	348
14.5	You Should	349
14.5.1	Remember These Definitions	349
14.5.2	Remember These Terms	349
14.5.3	Remember These Facts	350
14.5.4	Be Able to	350
 <b>Part V Mathematical Bits and Pieces</b>		
<b>15</b>	<b>Resources and Extras</b>	<b>355</b>
15.1	Useful Material About Matrices	355
15.1.1	The Singular Value Decomposition	356
15.1.2	Approximating A Symmetric Matrix	356
15.2	Some Special Functions	358
15.3	Splitting a Node in a Decision Tree	359
15.3.1	Accounting for Information with Entropy	359
15.3.2	Choosing a Split with Information Gain	360
<b>Index</b>		<b>363</b>

<http://www.springer.com/978-3-319-64409-7>

Probability and Statistics for Computer Science

Forsyth, D.

2018, XXIV, 367 p. 124 illus., 84 illus. in color.,

Hardcover

ISBN: 978-3-319-64409-7