

Contents

Part I Human Life, Medicine, and Genetics

1	Season of Birth and Human Longevity: A New Theory Why Children Born in November Live Longer	3
	Walter Krämer and Katharina Schüller	
1.1	The Date of Birth Matters	3
1.2	A New Look from Switzerland	5
1.3	In Search of the Reasons	7
1.4	Data Limitations	10
1.5	Further Reading	10
2	Where Do Drugs Work in the Body? A Systematic Statistical Data Analysis	11
	Claus Weihs	
2.1	Pharmacokinetics and the Pre-clinical Phase	11
2.2	Standard Procedure for Statistical Data Analysis	12
2.3	The Distribution of Drugs in the Body	13
2.3.1	Business Understanding	13
2.3.2	Data Understanding	13
2.3.3	Data Preparation	15
2.3.4	Modeling	16
2.3.5	Evaluation	18
2.3.6	Deployment	18
2.4	Further Reading	19
3	Drug Studies: Using Statistics to Achieve the Optimal Dose	21
	Holger Dette and Kirsten Schorning	
3.1	The Three Clinical Test Phases	21
3.2	The Optimization of Phase 2	22
3.3	It is All About the Experimental Design	23
3.4	Towards Practical Application	26
3.5	Further Reading	26

4	Statistical Alarm Systems in Intensive Care Medicine	27
	Roland Fried, Ursula Gather, and Michael Imhoff	
4.1	Alarms in Acute Medical Care	27
4.2	Smoothing as Part of Data Preprocessing	29
4.3	Joint Analysis of Characteristics	31
4.4	Validation of the Results	34
4.5	Further Reading	34
5	Personalized Medicine: How Statistics Helps Not to Drown in the Flood of Data	37
	Jörg Rahnenführer	
5.1	Genetic Decision Support in Medicine	37
5.2	Efficacy and Side Effects of Therapies	38
5.3	In Search for Genetic Patterns	39
5.4	Statistical Tricks	40
5.5	Medical Application	41
5.6	Models for Estimating Disease Progression	41
5.7	Summary	43
5.8	Further Reading	43
6	Modulating Genetic Effects on Bladder Cancer Risk in an Area of Coal, Iron, and Steel Industries	45
	Silvia Selinski, Katja Ickstadt, and Klaus Golka	
6.1	Environment, Genes, and Urinary Bladder Cancer	45
6.2	Epidemiology and Genetics	47
6.3	Gene-Environment Interactions	50
6.4	Conclusion	52
6.5	Further Reading	53
7	Statistics and the Maximum Human Lifespan	55
	Jan Feifel and Markus Pauly	
7.1	Background	55
7.2	From the Average to Extreme Value Theory	56
7.3	Challenges in Working with Demographic Data	59
7.4	Results	60
7.5	Conclusion	60
7.6	Further Reading	61
Part II Sports and Entertainment		
8	Statistics and Soccer	65
	Andreas Groll and Gunther Schauburger	
8.1	More Goals by Means of Statistics	65
8.2	A Statistical Model for Predicting Goals	66
8.3	Influential Variables	67
8.4	Conclusion	72
8.5	Further Reading	72

9	The Players' Anxiety at the Penalty Kick: Who Is the Best Penalty Taker, Who the Best Goalkeeper?	73
	Peter Gnändinger, Leo N. Geppert, and Katja Ickstadt	
9.1	Penalties in Soccer	73
9.2	The Penalty Data Set	74
9.3	Factors Associated with the Outcome of the Penalty	75
9.3.1	Modeling Penalty Probabilities	75
9.3.2	Fixed and Random Effects	77
9.3.3	Important Factors	78
9.4	Rankings of Goalkeepers and Penalty Takers	79
9.4.1	Penalty Takers' Influence as Random Effect	79
9.4.2	Leaderboards	80
9.5	Conclusion and Outlook	82
9.6	Further Reading	82
10	Music Data Analysis	83
	Claus Weihs	
10.1	What Is Music?	83
10.2	Music Data	85
10.3	The Studies	87
10.3.1	Classification	87
10.3.2	Pitch Identification	87
10.3.3	Instrument Recognition	89
10.3.4	Onset Detection	89
10.3.5	Automatic Transcription	90
10.3.6	Genres	91
10.4	Further Reading	92
11	Statistics and Horse Race Betting: Favorites vs. Longshots	93
	Martin Kukuk	
11.1	Horse Race Betting	93
11.2	Betting Payouts	94
11.3	Empirical Explanations for the Favorite-Longshot Bias	95
11.4	Favorite-Longshot-Bias Caused by Subjective Estimates	96
11.5	Conclusion	98
11.6	Further Reading	99
12	The Statistics of the German 6/49 Lotto	101
	Walter Krämer	
12.1	Lotto as an Investment	101
12.2	Optimizing the Payout	103
12.3	Further Reading	105

Part III Money and Business

13	Statistics at the Stock Exchange	109
	Walter Krämer and Tileman Conring	
13.1	Beware of Dependencies	109
13.2	Investing in Stocks	110
13.3	Time-Varying Dependencies	111
13.4	The Not So Normal Normal Distribution	113
13.5	Cointegration	115
13.6	Further Reading	116
14	Statistics in the Risk Assessment of Bank Portfolios	117
	Dominik Wied and Robert Löser	
14.1	The Problem	117
14.2	Expected Shortfall Compared to Value-at-Risk	118
14.3	Estimation of Risk Measures	119
14.4	Validation of Risk Models	120
14.5	Further Reading	123
15	On Rating the Raters: Statistics in the Rating Industry	125
	Walter Krämer and Simon Neumärker	
15.1	Obligations and Obligors	125
15.2	How to Judge the Quality of Default Forecasts?.....	127
15.3	A Numerical Example	129
15.4	Partial Orderings of Probability Forecasts	130
15.5	Scalar Valued Measures of Forecasting Quality	130
15.6	Further Reading	132
16	Gross Domestic Product, Greenhouse Gas Emissions, and Global Warming	133
	Martin Wagner, Fabian Knorre, and Christina Kopetzky	
16.1	Economic Activity and Emissions.....	133
16.2	Statistical Analysis	136
16.3	Parameter Estimation in the Presence of Nonlinear Cointegration	138
16.4	Interpretation	140
16.5	Further Reading	140

Part IV Nature and Technology

17	Flood Statistics: Still on the River Bank or Already in the Water? ...	143
	Svenja Fischer, Roland Fried, and Andreas Schumann	
17.1	Getting a Grip on Floods.....	143
17.2	What Is a Flood?	145
17.3	Flood Risk and Probabilities	146
17.4	Robust Estimation	148

17.5	Flood Types and Changes Over Time	149
17.6	Regionalization	151
17.7	Further Reading	152
18	How Statistics Helps to Reduce Rejects	153
	Claus Weihs and Nadja Bauer	
18.1	Defects in Deep Drilling	153
18.2	Quality Improvement: Six Sigma.....	154
18.2.1	Define	155
18.2.2	Measure	156
18.2.3	Analyze	158
18.2.4	Improve	160
18.2.5	Control	160
18.3	Further Reading	161
19	Statistics and Reliability of Technical Products	163
	Christine H. Müller	
19.1	Reliability and Randomness	163
19.2	Simple Service Lifetime Analysis	164
19.3	Lifetime Analysis Under Different Loads.....	165
19.4	Lifetime Analysis for Products with Several Components	166
19.5	Prediction Intervals.....	167
19.6	Outlook	169
19.7	Further Reading	169
20	Durable Machine Components: How Statistical Design of Experiments Optimizes Wear Protection	171
	Sonja Kuhnt, Wolfgang Tillmann, Alexander Brinkhoff, and Eva-Christina Becker-Emden	
20.1	Wear Protection Through Coating	171
20.2	Optimization Through Statistical Design of Experiments.....	172
20.3	Challenges in Real Coating Processes.....	175
20.4	Further Reading	177

Part V Intricacies of Measurement

21	Measuring the Immeasurable: Statistics, Intelligence, and Education.....	181
	Philipp Doeblér, Gesa Brunn, and Fritjof Freise	
21.1	Educational Tests and Education	181
21.2	Latent Variables and Their Indicators	182
21.3	A Statistical Model for Learning Progress Diagnostics	185
21.4	From Data to Latent Variables	186
21.5	Further Reading	189

22	Uncovering Embarrassing Truths Through Statistics	191
	Andreas Quatember	
22.1	The Method of Indirect Questioning	191
22.2	A Modification of the Original Idea	192
22.3	Tasks for Future Research	194
22.4	Further Reading	195
23	Samples and Missing Data	197
	Andreas Quatember	
23.1	Sampling in Theory and Practice	197
23.2	Statistical Methods to Compensate for Non-responses	198
23.3	Further Reading	200

Part VI Language Data

24	Who Is Supposed to Read All This? Automatic Analysis of Text Data	203
	Jörg Rahnenführer and Carsten Jentsch	
24.1	Large Text Collections	203
24.2	Text Analysis in the Social Sciences	204
24.3	Preprocessing of Text Data	205
24.4	Topic-Based Classification of Large Text Collections	206
24.5	Finding Differences	207
24.6	Text Analysis of Election Programs	208
24.7	Summary and Outlook	211
24.8	Further Reading	211
25	Statistical Modeling of Current Linguistic Realities Around the World: The Case of Singapore	213
	Sarah Buschfeld and Claus Weihs	
25.1	Singapore and the World Englishes Paradigm	213
25.2	Data Collection and Preparation	214
25.3	Prediction of Linguistic Characteristics	216
	25.3.1 Design of Experiments	216
	25.3.2 Variation in British and Singaporean Englishes	217
25.4	Evaluation and Interpretation	221
25.5	Summary and Outlook	222
25.6	Further Reading	222
26	Linguistic Manifestations of Cultural Differences Across National Varieties of English: A Methodological Survey	225
	Edgar W. Schneider	
26.1	Introduction	225
26.2	Electronic Corpora: Representative Text Collections?	227
26.3	Linguistic Forms Representing Cultural Orientations	230
	26.3.1 Example of a Cultural Dimension: Collectivism	230

26.3.2	Example of a Structural Scheme: Recipientless Constructions	232
26.4	Statistical Testing.....	233
26.5	Sample Results	234
26.6	Summary and Conclusion.....	235
26.7	Further Reading	235
 Part VII From Here to Where?		
27	Is Data Science More Than Statistics? The Bigger Picture	239
	Claus Weihs and Katja Ickstadt	
27.1	Data Science: What Is It Anyway?	239
27.2	Data Science: Steps	240
27.2.1	General Structure	240
27.2.2	Data Acquisition and Enrichment.....	241
27.2.3	Data Exploration.....	243
27.2.4	Modeling: Statistical Data Analysis	244
27.2.5	Evaluation: Model Validation and Selection	245
27.2.6	Deployment of Results	246
27.3	Conclusion	246
27.4	Further Reading	246
 Index		249