

The background of the book cover is a close-up, high-contrast image of a fingerprint. The ridges and valleys are clearly visible, creating a complex, swirling pattern. The color palette is primarily shades of blue and purple, with the ridges appearing in a lighter, almost white-blue hue against a darker, deep purple background. The fingerprint pattern is centered and occupies most of the cover area.

David Lucy

Introduction to
Statistics *for*
Forensic Scientists

 WILEY

Contents

Preface	ix
List of figures	xi
List of tables	xiii
1 A short history of statistics in the law	1
1.1 History	1
1.2 Some recent uses of statistics in forensic science	3
1.3 What is probability?	4
2 Data types, location and dispersion	7
2.1 Types of data	7
2.2 Populations and samples	9
2.3 Distributions	9
2.4 Location	11
2.5 Dispersion	13
2.6 Hierarchies of variation	14
3 Probability	17
3.1 Aleatory probability	17
One throw of a six-sided die	17
A single throw with more than one outcome of interest	18
Two six-sided dice	19
3.2 Binomial probability	21
3.3 Poisson probability	24
3.4 Empirical probability	25
Modelled empirical probabilities	25
Truly empirical probabilities	27

4	The normal distribution	29
4.1	The normal distribution	29
4.2	Standard deviation and standard error of the mean	30
4.3	Percentage points of the normal distribution	32
4.4	The t -distribution and the standard error of the mean	34
4.5	t -testing between two independent samples	36
4.6	Testing between paired observations	40
4.7	Confidence, significance and p -values	42
5	Measures of nominal and ordinal association	45
5.1	Association between discrete variables	45
5.2	χ^2 test for a 2×2 table	46
5.3	Yules Q	48
5.4	χ^2 tests for greater than 2×2 tables	49
5.5	ϕ^2 and Cramers V^2	50
5.6	The limitations of χ^2 testing	51
5.7	Interpretation and conclusions	52
6	Correlation	55
6.1	Significance tests for correlation coefficients	59
6.2	Correlation coefficients for non-linear data	60
6.3	The coefficient of determination	63
6.4	Partial correlation	63
6.5	Partial correlation controlling for two or more covariates	69
7	Regression and calibration	75
7.1	Linear models	75
7.2	Calculation of a linear regression model	78
7.3	Testing 'goodness of fit'	80
7.4	Testing coefficients a and b	81
7.5	Residuals	83
7.6	Calibration	85
	A linear calibration model	86
	Calculation of a confidence interval for a point	89
7.7	Points to remember	91
8	Evidence evaluation	95
8.1	Verbal statements of evidential value	95
8.2	Evidence types	96
8.3	The value of evidence	97
8.4	Significance testing and evidence evaluation	102
9	Conditional probability and Bayes' theorem	105
9.1	Conditional probability	105
9.2	Bayes' theorem	108
9.3	The value of evidence	112

10	Relevance and the formulation of propositions	117
	10.1 Relevance	117
	10.2 Hierarchy of propositions	118
	10.3 Likelihood ratios and relevance	120
	10.4 The logic of relevance	122
	10.5 The formulation of propositions	123
	10.6 What kind of propositions can we not evaluate?	124
11	Evaluation of evidence in practice	129
	11.1 Which database to use	129
	Type and geographic factors	129
	DNA and database selection	131
	11.2 Verbal equivalence of the likelihood ratio	133
	11.3 Some common criticisms of statistical approaches	136
12	Evidence evaluation examples	139
	12.1 Blood group frequencies	139
	12.2 Trouser fibres	141
	12.3 Shoe types	144
	12.4 Airweapon projectiles	148
	12.5 Height description from eyewitness	150
13	Errors in interpretation	155
	13.1 Statistically based errors of interpretation	155
	Transposed conditional	156
	Defender's fallacy	157
	Another match error	157
	Numerical conversion error	158
	13.2 Methodological errors of interpretation	159
	Different level error	158
	Defendant's database fallacy	159
	Independence assumption	159
14	DNA I	161
	14.1 Loci and alleles	161
	14.2 Simple case genotypic frequencies	162
	14.3 Hardy-Weinberg equilibrium	164
	14.4 Simple case allelic frequencies	166
	14.5 Accounting for sub-populations	168
15	DNA II	171
	15.1 Paternity – mother and father unrelated	171
	15.2 Database searches and value of evidence	174
	15.3 Discussion	176
16	Sampling and sample size estimation	179
	16.1 Estimation of a mean	179

16.2	Sample sizes for <i>t</i> -tests	181
	Two sample <i>t</i> -test	181
	One sample <i>t</i> -test	183
16.3	How many drugs to sample	184
16.4	Concluding comments	188
17	Epilogue	191
17.1	Graphical models and Bayesian Networks	192
	Graphical models	192
	Bayesian networks	194
17.2	Kernel density estimation	195
17.3	Multivariate continuous matching	196
	Appendices	
A	Worked solutions to questions	199
B	Percentage points of the standard normal distribution	225
C	Percentage points of <i>t</i>-distributions	227
D	Percentage points of χ^2-distributions	229
E	Percentage points of beta-beta distributions	231
F	Percentage points of F-distributions	233
G	Calculating partial correlations using Excel software	235
H	Further algebra using the “third law”	239
	References	243
	Index	249