

# Contents

<b>1</b>	<b>USING STATISTICS</b>	<b>1</b>
<b>2</b>	<b>DESCRIBING THE DATA.</b>	<b>3</b>
2.1	Types of Data	3
2.2	Bar Charts and Histograms	4
2.3	Averages.	6
2.3.1	The arithmetic mean.	6
★2.3.2	Alternatives to the arithmetic mean.	7
2.4	Measuring the Spread	8
2.4.1	The variance	8
2.4.2	The standard deviation	9
★2.4.3	Different definitions of the standard deviation	10
★2.4.4	Alternative measures of the spread	12
★2.5	Higher Powers of $x$ .	13
★2.5.1	Skew	13
★2.5.2	Higher powers	14
2.6	More Than One Variable	14
2.6.1	Covariance	15

2.6.2 Correlation . . . . .	15
★2.6.3 More than two variables . . . . .	17
2.7 Problems . . . . .	18
<b>3 THEORETICAL DISTRIBUTIONS . . . . .</b>	<b>20</b>
3.1 General Properties of Distributions . . . . .	21
3.1.1 A simple distribution . . . . .	21
3.1.2 The law of large numbers . . . . .	21
3.1.3 Expectation values . . . . .	22
3.1.4 Probability density distributions . . . . .	23
3.2 The Binomial Distribution . . . . .	24
3.2.1 The binomial probability distribution formula . . . . .	25
★3.2.2 Proof of properties of the binomial distribution . . . . .	27
3.3 The Poisson Distribution . . . . .	28
3.3.1 The Poisson probability formula . . . . .	29
★3.3.2 Proof of properties of the Poisson distribution . . . . .	32
★3.3.3 Two Poisson distributions . . . . .	33
3.4 The Gaussian Distribution . . . . .	34
3.4.1 The Gaussian probability distribution function . . . . .	34
★3.4.2 Proof of properties of the Gaussian . . . . .	36
3.4.3 Definite integrals . . . . .	36
3.4.4 Indefinite integrals . . . . .	37
3.4.5 Gaussian as limit of the Poisson and binomial . . . . .	40
★3.4.6 The many-dimensional Gaussian . . . . .	41
★3.4.7 The binormal distribution . . . . .	42
★3.5 Other Distributions . . . . .	44
★3.5.1 The uniform distribution . . . . .	45
★3.5.2 The Weibull distribution . . . . .	45
★3.5.3 The Breit-Wigner or Cauchy distribution . . . . .	46
3.6 Problems . . . . .	46
<b>4 ERRORS . . . . .</b>	<b>48</b>
4.1 Why Errors are Gaussian . . . . .	49
4.1.1 The central limit theorem . . . . .	50
4.2 Working with Errors . . . . .	51
4.2.1 Repeated measurements . . . . .	51
4.2.2 Averaging weighted measurements . . . . .	53
4.2.3 A note of caution . . . . .	54
4.3 Combination of Errors . . . . .	55
4.3.1 One variable . . . . .	55
4.3.2 A function of two or more variables . . . . .	56
4.3.3 Percentage errors . . . . .	58

★4.3.4 Several functions of several variables . . . . .	58
4.4 Systematic Errors . . . . .	61
4.4.1 Finding, eliminating, and evaluating them . . . . .	62
★4.4.2 Living with systematic errors . . . . .	65
4.5 Problems . . . . .	67
<b>5 ESTIMATION . . . . .</b>	<b>68</b>
5.1 Properties of Estimators . . . . .	68
5.1.1 Consistency, bias, and efficiency . . . . .	69
★5.1.2 The likelihood function . . . . .	71
★5.1.3 Proof of the minimum variance bound . . . . .	74
5.2 Some Basic Estimators . . . . .	75
5.2.1 Estimating the mean . . . . .	76
5.2.2 Estimating the variance . . . . .	76
5.2.3 Estimating $\sigma$ . . . . .	78
★5.2.4 The correlation coefficient . . . . .	80
★5.3 Maximum Likelihood . . . . .	81
★5.3.1 ML: consistency, bias, and invariance . . . . .	84
★5.3.2 Maximum likelihood at large $N$ . . . . .	85
★5.3.3 Errors on the ML estimators . . . . .	86
★5.3.4 Several variables . . . . .	88
★5.3.5 Notes on maximum likelihood . . . . .	89
★5.4 Extended Maximum Likelihood . . . . .	90
★5.5 The Method of Moments . . . . .	92
★5.6 Maximum Likelihood and Least Squares . . . . .	93
★5.7 Stratified Sampling—Beating $\sqrt{N}$ . . . . .	93
5.8 Problems . . . . .	95
<b>6 LEAST SQUARES . . . . .</b>	<b>97</b>
6.1 Outline of the Method . . . . .	97
6.1.1 Fitting $y = mx$ — simple proportion . . . . .	98
6.2 The Straight Line Fit . . . . .	99
6.2.1 The slope and intercept for a straight line . . . . .	100
★6.2.2 Derivation of the errors for a straight line . . . . .	101
★6.2.3 Weighted straight line fit . . . . .	102
★6.2.4 Extrapolation . . . . .	103
★6.2.5 Systematic errors and a straight line fit . . . . .	103
★6.2.6 Regression . . . . .	104
6.3 Fitting Binned Data . . . . .	105
6.4 The $\chi^2$ Distribution . . . . .	106
★6.4.1 Proof of the $\chi^2$ distribution . . . . .	108
★6.5 Errors on $x$ and $y$ . . . . .	109

★6.6	Linear Least Squares and Matrices . . . . .	111
★6.6.1	Straight line fit using matrices . . . . .	113
★6.6.2	Higher polynomials . . . . .	114
★6.7	Non-linear Least Squares . . . . .	115
6.8	Problems . . . . .	116
<b>★7</b>	<b>PROBABILITY AND CONFIDENCE . . . . .</b>	<b>118</b>
★7.1	What is Probability? . . . . .	119
★7.1.1	Mathematical probability . . . . .	119
★7.1.2	Empirical—the limit of a frequency. . . . .	119
★7.1.3	Objective—propensity . . . . .	120
★7.1.4	Subjective probability . . . . .	121
★7.1.5	Bayesian statistics . . . . .	123
★7.1.6	Conclusions on probability . . . . .	124
★7.2	Confidence Levels . . . . .	125
★7.2.1	Confidence levels in descriptive statistics . . . . .	125
★7.2.2	Confidence intervals in estimation . . . . .	127
★7.2.3	Confidence levels from Gaussians . . . . .	129
★7.2.4	Measurement of a constrained quantity . . . . .	130
★7.2.5	Binomial confidence intervals . . . . .	132
★7.2.6	Poisson confidence intervals. . . . .	133
★7.2.7	Several variables—confidence regions . . . . .	134
★7.3	Student's <i>t</i> Distribution. . . . .	134
★7.3.1	Proof of the formula for Student's <i>t</i> . . . . .	138
★7.4	Problems . . . . .	139
<b>8</b>	<b>TAKING DECISIONS . . . . .</b>	<b>141</b>
8.1	Hypothesis Testing . . . . .	142
8.1.1	Hypotheses . . . . .	142
8.1.2	Type I and type II errors . . . . .	142
8.1.3	Significance . . . . .	143
8.1.4	Power . . . . .	143
★8.1.5	The Neyman Pearson test . . . . .	144
8.2	Interpreting Experiments . . . . .	145
8.2.1	The null hypothesis . . . . .	146
★8.2.2	Binomial probabilities . . . . .	147
★8.2.3	Is there a signal??—Poisson statistics . . . . .	148
8.3	Goodness of Fit. . . . .	149
8.3.1	The $\chi^2$ test . . . . .	150
★8.3.2	The run test . . . . .	153
★8.3.3	The Kolmogorov test . . . . .	155
8.4	The Two-sample Problem . . . . .	156

8.4.1	Two Gaussian samples with known $\sigma$ . . . . .	156
★8.4.2	Two Gaussian samples with unknown $\sigma$ . . . . .	157
★8.4.3	Matched and correlated samples . . . . .	159
★8.4.4	The <i>F</i> distribution . . . . .	160
★8.4.5	The general case . . . . .	161
★8.5	Analysis Methods for Several Samples. . . . .	164
★8.5.1	The analysis of variance (basic method). . . . .	164
★8.5.2	Multway analysis of variance . . . . .	166
★8.5.3	Contingency tables . . . . .	169
8.6	Problems . . . . .	170
<b>★9</b>	<b>RANKING METHODS . . . . .</b>	<b>172</b>
★9.1	Non-parametric Methods . . . . .	173
★9.1.1	The sign test for the median . . . . .	173
★9.2	Two Ranked Samples . . . . .	174
★9.2.1	The Mann–Whitney test. . . . .	174
★9.2.2	Matched pairs . . . . .	175
★9.2.3	Wilcoxon's matched pairs signed rank test . . . . .	176
★9.3	Measures of Agreement. . . . .	177
★9.3.1	Spearman's correlation coefficient . . . . .	177
★9.3.2	Concordance. . . . .	178
★9.4	Problems . . . . .	179
<b>10</b>	<b>NOTES FOR NUMBER CRUNCHERS . . . . .</b>	<b>180</b>
10.1	Significance . . . . .	180
10.1.1	Subtraction . . . . .	181
10.1.2	Computing the standard deviation . . . . .	181
10.1.3	Addition . . . . .	182
10.1.4	Quadratic equations . . . . .	183
10.2	Other Do's and Don't's . . . . .	183
10.2.1	Matrix inversion . . . . .	183
10.2.2	Fitting curves . . . . .	184
10.3	Random Number Generation. . . . .	184
10.4	Style . . . . .	185
	Bibliography . . . . .	187
	Appendix 1 Answers to Problems . . . . .	190
	Appendix 2 Proof of the Central Limit Theorem . . . . .	196
	Index . . . . .	199