

# Contents

---

<b>1. Introduction</b>	1
1.1 The scope of statistics	1
1.2 Statistics and medical research	1
1.3 Statistics and mathematics	2
1.4 Statistics and computing	3
1.5 The scope of this book	4
<b>2. The design of experiments</b>	6
2.1 Comparing treatments	6
2.2 Random allocation	8
2.3 Methods of allocation without random numbers	13
2.4 Volunteer bias	15
2.5 Cross-over designs	17
2.6 Selection of experimental subjects	18
2.7 Response bias and placebos	20
2.8 Assessment bias and double-blind studies	21
2.9 Laboratory experiments	22
2.10 Experimental units	24
2.11 Further points about trial design	24
<b>3. Sampling and observational studies</b>	30
3.1 Observational studies	30
3.2 Censuses	30
3.3 Sampling	31
3.4 Random sampling	33
3.5 Sampling in clinical studies	36
3.6 Sampling in epidemiological studies	38
3.7 Case control studies	40
3.8 Cohort studies	43
3.9 Questionnaire bias in observational studies	44
<b>4. Summarizing data</b>	51
4.1 Types of data	51
4.2 Frequency distributions	51
4.3 Histograms and other frequency graphs	57

4.4	Shapes of frequency distribution	61
4.5	Medians and quantiles	63
4.6	The mean	64
4.7	Variance and standard deviation	65
4A.1	The divisor for the variance	69
4A.2	Formulae for the sum of squares about the mean	72
<b>5.</b>	<b>Presenting data</b>	<b>75</b>
5.1	Rates and proportions	75
5.2	Significant figures	76
5.3	Presenting tables	78
5.4	Pie charts	80
5.5	Bar charts	80
5.6	Misleading graphs	82
5.7	Scatter diagrams	84
5.8	Line graphs and time series	85
5.9	Logarithmic scales	88
5A.	Logarithms	89
<b>6.</b>	<b>Probability</b>	<b>95</b>
6.1	Probability	95
6.2	Properties of probability	96
6.3	Probability distributions and random variables	97
6.4	The Binomial Distribution	98
6.5	Mean and variance	101
6.6	Properties and means and variances	102
6.7	The Poisson Distribution	104
6A.1	Combinations	105
6A.2	Expected value of a sum of squares	106
<b>7.</b>	<b>The Normal Distribution</b>	<b>112</b>
7.1	Probability distributions for continuous variables	112
7.2	The Normal Distribution	116
7.3	Properties of the Normal Distribution	120
7.4	Variables which are themselves Normally distributed	125
7.5	Assessing the fit of the Normal Distribution	126
7A	The Chi-squared and Student's $t$ Distributions	129
<b>8.</b>	<b>Estimation, standard error, and confidence intervals</b>	<b>134</b>
8.1	Sampling distributions	134
8.2	Standard error of a sample mean	136
8.3	Confidence intervals	138
8.4	Standard error of a proportion	140

8.5	Standard error of the difference between two means	140
8.6	Standard error of the difference between two proportions	142
8.7	Standard error of a sample standard deviation	143
8.8	Sample size for an estimate	144
8A.	The derivation of the standard error of the mean	145
<b>9.</b>	<b>Significance tests</b>	<b>148</b>
9.1	Testing a hypothesis	148
9.2	An example: the sign test	149
9.3	General principles of significance tests	151
9.4	Significance levels	152
9.5	One- and two-sided tests of significance	152
9.6	Significant, real and important	154
9.7	Comparing the means of large samples using the Normal Distribution	154
9.8	Comparison of two proportions	157
9.9	The power of a test	159
9.10	Sample size for a comparison	160
9.11	Multiple significance tests	161
<b>10.</b>	<b>Analysis of the means of small samples using the <math>t</math> Distribution</b>	<b>165</b>
10.1	The $t$ Distribution	165
10.2	The one sample $t$ method	169
10.3	The means of two independent samples	172
10.4	The use of transformations	175
10.5	Deviations from the assumptions of Normal Distribution and uniform variance	178
10.6	What is a large sample?	182
10A	Why the mean/standard error follows the $t$ Distribution	184
<b>11.</b>	<b>Regression and correlation</b>	<b>188</b>
11.1	Scatter diagrams	188
11.2	Regression	189
11.3	The method of least squares	191
11.4	The regression of X on Y	194
11.5	The standard error of the regression coefficient	195
11.6	Using the regression line for prediction	197
11.7	Analysis of residuals	200
11.8	Deviations from assumptions in regression	201
11.9	Extensions of the regression method	203
11.10	Correlation	203
11.11	Confidence interval and significance test for the correlation coefficient	207

11.12	Use of the correlation coefficient	208
11A.1	Derivation of the least-squares regression equation	209
11A.2	The standard error of $b$ and the variance about the line	210
<b>12.</b>	<b>Methods based on rank order</b>	<b>216</b>
12.1	Non-parametric methods	216
12.2	The Mann–Whitney $U$ test	217
12.3	The Wilcoxon matched pairs test	224
12.4	Spearman’s rank correlation coefficient	227
12.5	Kendall’s rank correlation coefficient	230
12.6	Continuity corrections	236
12.7	Parametric or non-parametric methods?	238
<b>13.</b>	<b>The analysis of cross-tabulations using the Chi-squared Distribution</b>	<b>241</b>
13.1	The chi-squared test for association	241
13.2	Validity of the chi-squared test for small samples	244
13.3	Tests for 2 by 2 tables	246
13.4	Chi-squared test for trend	247
13.5	Fisher’s exact test	250
13.6	Yates’ continuity correction for a 2 by 2 table	254
13.7	The validity of Fisher’s exact test and Yates’ correction	255
13.8	McNemar’s test for matched samples	255
13A.1	Why the chi-squared test works	258
13A.2	Derivation of the formula for Fisher’s exact test	260
<b>14.</b>	<b>Choosing the statistical method</b>	<b>264</b>
14.1	Method-orientated and problem-orientated teaching	264
14.2	Types of data	266
14.3	Comparing two groups	266
14.4	One sample and paired samples	267
14.5	Relationship between two variables	268
<b>15.</b>	<b>Clinical measurement</b>	<b>276</b>
15.1	Repeatability and precision in measurement	276
15.2	Digit preference	278
15.3	Comparing two methods of measurement	280
15.4	Sensitivity and specificity	283
15.5	Normal or reference ranges	285
15.6	Survival data	288
15.7	Computer aided diagnosis	291
15A.	Standard deviation for measurement error	293

<b>16. Mortality statistics and the structure of human populations</b>	297
16.1 Mortality rates	297
16.2 Age standardized mortality rates using the direct method	299
16.3 Standardized mortality ratios by the indirect method	300
16.4 Demographic life tables	302
16.5 Vital statistics	305
16.6 The population pyramid	305
<b>17. Solutions to exercises</b>	312
<b>References</b>	353
<b>Index</b>	357