

# Contents

<i>Preface</i>	xi
<i>Preface to the second edition</i>	xii
<i>Preface to the first edition</i>	xiii
<i>Acknowledgements</i>	xv
<i>1 What is statistics?</i>	1
1.1 Statistics as a science	1
1.2 Types of statistical data	2
Worksheet 1	3
<i>2 Some statistical notation</i>	5
2.1 $\Sigma$	5
2.2 Factorials	6
2.3 $x^y$	7
2.4 $e^x$	7
2.5 Decimal places and significant figures	7
Worksheet 2	8
<i>3 Summarizing data by tables and by graphical methods</i>	10
3.1 Tables and graphs for one continuous variable	10
3.2 Tables and graphs for one discrete variable	14
3.3 Tables and graphs for one categorical variable	17
3.4 Tables and graphs for two-variable data	17
3.5 Summary	20
Worksheet 3	20
<i>4 Summarizing data by numerical measures</i>	24
4.1 Averages	24
4.2 Sample mean ( $\bar{x}$ )	25
4.3 Sample median	25

4.4	Sample mode	26
4.5	When to use the mean, median and mode	28
4.6	Measures of variation	30
4.7	Sample standard deviation ( $s$ )	30
4.8	Sample inter-quartile range	32
4.9	When to use standard deviation and inter-quartile range	33
4.10	Box and whisker plots	33
4.11	Other measures of variation	34
4.12	A measure of skewness	34
4.13	Summary	35
	Worksheet 4	36
<b>5</b>	<b><i>Probability</i></b>	<b>38</b>
5.1	Introduction	38
5.2	Basic ideas of probability	39
5.3	The <i>a priori</i> definition of probability for equally likely outcomes	40
5.4	The relative frequency definition of probability, based on experimental data	41
5.5	The range of possible values for a probability value	42
5.6	Probability, percentage, proportion and odds	42
5.7	Subjective probability	42
5.8	Probabilities involving more than one event	43
5.9	Multiplication law (the 'and' law)	43
5.10	Addition law (the 'or' law)	45
5.11	Mutually exclusive and exhaustive events	47
5.12	Complementary events and the calculation of $P$ (at least 1 . . .)	48
5.13	Probability trees	48
5.14	Venn diagrams and Rees diagrams	50
5.15	Summary	51
	Worksheet 5	52
<b>6</b>	<b><i>Discrete probability distributions</i></b>	<b>55</b>
6.1	Introduction	55
6.2	Binomial distribution, an example	55
6.3	The general binomial distribution	56
6.4	Calculating binomial probabilities, an example	57
6.5	The mean and standard deviation of the binomial distribution	58
6.6	Binomial probabilities using tables	59
6.7	Binomial probabilities using Minitab	60
6.8	Simulation of binomial distributions using Minitab	61

6.9	Poisson distribution, an introduction	62
6.10	Some examples of Poisson variables	62
6.11	The general Poisson distribution	63
6.12	Calculating Poisson probabilities, an example	64
6.13	The mean and standard deviation of the Poisson distribution	64
6.14	Poisson probabilities using tables	65
6.15	Poisson probabilities using Minitab	66
6.16	Simulation of Poisson distributions using Minitab	66
6.17	Poisson approximation to the binomial distribution	66
6.18	Summary	67
	Worksheet 6	68
<b>7</b>	<b><i>Continuous probability distributions</i></b>	<b>71</b>
7.1	Introduction	71
7.2	The normal distribution	73
7.3	An example of a normal distribution	74
7.4	Normal probabilities using Minitab	76
7.5	Simulation of the normal distribution using Minitab	77
7.6	Rectangular distribution	78
7.7	The normal approximation to the binomial distribution	79
7.8	Summary	80
	Worksheet 7	80
<b>8</b>	<b><i>Samples and populations</i></b>	<b>84</b>
8.1	Introduction	84
8.2	Reasons for sampling	84
8.3	Sampling methods	85
8.4	Sample size	87
8.5	Sampling distribution of the sample mean	87
8.6	Simulation of the sampling distribution of the sample mean using Minitab	89
8.7	Summary	90
	Worksheet 8	91
<b>9</b>	<b><i>Confidence interval estimation</i></b>	<b>93</b>
9.1	Introduction	93
9.2	95% confidence intervals	94
9.3	Calculating a 95% confidence interval for the mean, $\mu$ , of a population: large sample size, $n$	94
9.4	Calculating a 95% confidence interval for the mean, $\mu$ , of a population: small sample size, $n$	96
9.5	The $t$ distribution	99

9.6	The choice of sample size when estimating the mean of a population	99
9.7	Degrees of freedom	100
9.8	95% confidence interval for a binomial probability	101
9.9	The choice of sample size when estimating a binomial probability	102
9.10	95% confidence interval for the mean of a population of differences, 'paired' samples data	102
9.11	95% confidence interval for the difference in the means of two populations, 'unpaired' samples data	104
9.12	Summary	107
	Worksheet 9	107

<b>10</b>	<b><i>Hypothesis testing</i></b>	<b>112</b>
10.1	Introduction	112
10.2	What is a hypothesis?	113
10.3	Which is the null hypothesis and which is the alternative hypothesis?	113
10.4	What is a significance level?	114
10.5	What is a test statistic, and how do we calculate it?	114
10.6	How do we find the tabulated test statistic?	115
10.7	How do we compare the calculated and tabulated test statistics?	115
10.8	What is our conclusion, and what assumptions have we made?	116
10.9	Hypothesis test for the mean, $\mu$ , of a population	116
10.10	Two examples of hypothesis tests with one-sided alternative hypotheses	117
10.11	Hypothesis test for a binomial probability	118
10.12	Hypothesis test for the mean of a population of differences, 'paired' samples data	120
10.13	Hypothesis test for the difference in the means of two populations, 'unpaired' samples data	121
10.14	Hypothesis test for the equality of the variances of two normally distributed populations	123
10.15	The effect of choosing significance levels other than 5%	123
10.16	What if the assumptions of a hypothesis test are not valid?	124
10.17	The connection between confidence interval estimation and hypothesis testing	124
10.18	Summary	125
	Worksheet 10	125

<i>11 Non-parametric hypothesis tests</i>	130
11.1 Introduction	130
11.2 Sign test for the median of a population	130
11.3 Sign test for the median of a population of differences, ‘paired’ samples data	132
11.4 Sign test for large samples ( $n > 10$ )	133
11.5 Sign test using Minitab	135
11.6 Wilcoxon signed rank test for the median of a population of differences, ‘paired’ samples data	135
11.7 Wilcoxon signed rank test for large samples ( $n > 25$ )	137
11.8 Wilcoxon signed rank test using Minitab	138
11.9 Mann–Whitney $U$ test for the difference in the medians of two populations, ‘unpaired’ samples data	139
11.10 Mann–Whitney $U$ test for large samples	141
11.11 Mann–Whitney $U$ test using Minitab	142
11.12 Summary	142
Worksheet 11	142
<i>12 Association of categorical variables</i>	146
12.1 Introduction	146
12.2 Contingency tables	146
12.3 $\chi^2$ test for independence, $2 \times 2$ contingency table data	147
12.4 $\chi^2$ test for independence, $3 \times 3$ table	149
12.5 $\chi^2$ test for independence using Minitab	151
12.6 Fisher exact test	152
12.7 $\chi^2$ trend test	154
12.8 Summary	155
Worksheet 12	155
<i>13 Correlation of quantitative variables</i>	159
13.1 Introduction	159
13.2 Pearson’s correlation coefficient	159
13.3 Hypothesis test for Pearson’s population correlation coefficient, $\rho$	162
13.4 The interpretation of significant and non-significant correlation coefficients	163
13.5 Spearman’s rank correlation coefficient	165
13.6 Hypothesis test for Spearman’s rank correlation coefficient	167
13.7 Spearman’s rank correlation coefficient in the case of ties	167
13.8 Correlation coefficients using Minitab	169
13.9 Summary	170
Worksheet 13	170

<i>14 Regression analysis</i>	174
14.1 Introduction	174
14.2 Determining the regression equation from sample data	175
14.3 Plotting the regression line on the scatter diagram	176
14.4 Predicting values of $y$	177
14.5 Confidence intervals for predicted values of $y$	178
14.6 Hypothesis test for the slope of the regression line	180
14.7 The connection between regression and correlation	181
14.8 Regression analysis using Minitab	181
14.9 Transformations to produce linear relationships	182
14.10 Summary	184
Worksheet 14	185
<i>15 Goodness-of-fit tests</i>	189
15.1 Introduction	189
15.2 Goodness-of-fit for a 'simple proportion' distribution	189
15.3 Goodness-of-fit for a binomial distribution	191
15.4 Goodness-of-fit for a Poisson distribution	193
15.5 The Shapiro–Wilk test for normality	195
15.6 Summary	197
Worksheet 15	197
<i>Appendix A Data set for a random sample of 40 students</i>	201
<i>Appendix B Multiple-choice test</i>	203
<i>Appendix C Solutions to worksheets and multiple-choice test</i>	209
<i>Appendix D Statistical tables</i>	229
<i>Appendix E Glossary of symbols</i>	252
<i>Appendix F Introduction to Minitab data entry and list of Minitab commands</i>	255
<i>Index</i>	260