

# **A Basic Course in Statistics**

**Fourth Edition**

**G. M. Clarke and D. Cooke**



**A member of the Hodder Headline Group**  
**LONDON • NEW YORK • SYDNEY • AUCKLAND**

# Contents

<i>Preface to the fourth edition</i>	xv
<i>Introduction</i>	xvii
<i>List of projects</i>	xxi
<i>Notation</i>	xxiii
1. Populations and variates	1
1.1 A statistical investigation	1
1.2 Constructing tables and graphs	5
1.3 Tables and diagrams	12
1.4 Exercises on Chapter 1	13
1.5 Computing exercises	16
1.6 Projects	17
Summary	17
2. Measures of the centre of a set of observations	19
2.1 Introduction	19
2.2 The median	19
2.3 The arithmetic mean	21
2.4 $\Sigma$ -notation	24
2.5 The mean of grouped data	27
2.6 Coding	28
2.7 The mode	31
2.8 The geometric mean	31
2.9 Weighted averages and index numbers	31
2.10 Exercises on Chapter 2	34
2.11 Computing exercises	37
2.12 Projects	38
Summary	39
3. Samples and populations	41
3.1 A sampling investigation	41
3.2 Some methods of choosing a sample	42
3.3 Completing the investigation	45

3.4	Random numbers	46
3.5	Pseudo-random numbers	47
3.6	Systematic samples	48
3.7	Some useful terms	48
3.8	Sampling with and without replacement	49
3.9	Exercises on Chapter 3	50
3.10	Computing exercises	52
3.11	Projects	52
	Summary	53
4.	The measurement of variability	54
4.1	Introduction	54
4.2	Range	54
4.3	Mean deviation	55
4.4	Variance and standard deviation	56
4.5	Variance and standard deviation of a grouped frequency distribution	59
4.6	The semi-inter-quartile range (or quartile deviation)	63
4.7	Quantiles	65
4.8	Exercises on Chapter 4	69
4.9	Computing exercises	70
4.10	Projects	71
	Summary	72
5.	Looking at data	74
5.1	Introduction	74
5.2	Data preparation and scrutiny	75
5.3	Data description	78
5.4	Conclusion	83
5.5	Exercises on Chapter 5	83
5.6	Project	85
	Summary	86
	Revision Exercises A	86
6.	Probability	93
6.1	Introduction	93
6.2	Winning the toss	95
6.3	Terminology and notation	97
6.4	Probability when outcomes are not equally likely	102
6.5	Definition of probability	104
6.6	Setting up probability models	105
6.7	Counting methods	105
6.8	Use of counting methods in probability	110
6.9	Use of computers	111
6.10	Exercises on Chapter 6	112
6.11	Computing exercises	113
6.12	Projects	114
	Summary	114

7. Probabilities of compound events	116
7.1 Introduction	116
7.2 The intersection of two events	116
7.3 The union of two events	117
7.4 Theorems on the probability of union of two events	119
7.5 Complementary events	122
7.6 The intersection and union of three or more events	124
7.7 Conditional probability	126
7.8 Independence	130
7.9 Inverse probability	135
7.10 Simulation	141
7.11 Exercises on Chapter 7	143
7.12 Computing exercises	144
7.13 Projects	145
Summary	146
8. Discrete random variables	148
8.1 Predicting the number of left-handed children in a class	148
8.2 Random variables	151
8.3 Families of random variables	153
8.4 The discrete uniform distribution	153
8.5 The Bernoulli distribution	154
8.6 The binomial distribution	155
8.7 The geometric distribution	159
8.8 Acceptance sampling in quality control	160
8.9 Cumulative distribution functions and medians	167
8.10 Distribution of a function $g(R)$ given the distribution of $R$	169
8.11 Choice of a random sample from a distribution	170
8.12 Exercises on Chapter 8	172
8.13 Computing exercises	174
8.14 Projects	175
Summary	176
9. Expectation of random variables	177
9.1 Expected values	177
9.2 The means of the binomial and geometric distributions	179
9.3 Expectation of a linear function of a random variable	181
9.4 The expected value of a function of a random variable	183
9.5 Variance as an expected value	187
9.6 The variance of the binomial distribution	188
9.7 Variance of a linear function of a random variable	189
9.8 Average sample number in acceptance sampling	190
9.9 Probability generating functions	194
9.10 Exercises on Chapter 9	197
9.11 Computing exercises	199
9.12 Projects	199

Summary	200
Revision Exercises B	201
10. Joint distributions	207
10.1 Bivariate distributions	207
10.2 Mean of the sum of two random variables	211
10.3 Mean of the product of two independent random variables	214
10.4 The variance of the sum of two independent random variables	217
10.5 Exercises on Chapter 10	219
10.6 Computing exercises	221
10.7 Projects	222
Summary	223
11. Estimation	224
11.1 The distribution of estimates of a proportion	224
11.2 Distribution of estimates of a mean	227
11.3 The mean and variance of $\bar{X}$	228
11.4 Unbiased estimators	229
11.5 Unbiased estimation of variance	230
11.6 Exercises on Chapter 11	232
11.7 Computing exercises	235
11.8 Projects	235
Summary	236
12. Collecting data	238
12.1 Introduction	238
12.2 Surveys	238
12.3 Stratified random sampling	240
12.4 Experiments	241
12.5 Basic principles of experimental design	242
12.6 Exercises on Chapter 12	245
12.7 Projects	246
Summary	247
13. Significance testing	248
13.1 Testing a hypothesis	248
13.2 Another alternative hypothesis	253
13.3 Two types of error	257
13.4 Significance	258
13.5 The Wilcoxon rank sum (or Mann–Whitney $U$ ) test	259
13.6 The Wilcoxon signed-rank test	265
13.7 Type II errors and the power of a test	267
13.8 Summary and definitions	271
13.9 Exercises on Chapter 13	273
13.10 Computing exercises	275
13.11 Projects	275

Summary	277
Revision Exercises C	277
14. Continuous random variables	284
14.1 Modelling continuous variates	284
14.2 The specification of continuous random variables	290
14.3 The mean and variance of continuous random variables	293
14.4 Use of the expectation operator with continuous random variables	294
14.5 The continuous uniform distribution	297
14.6 The exponential distribution	298
14.7 Uses of the exponential distribution	300
14.8 The moment generating function	303
14.9 Sums of independent random variables	306
14.10 Distribution of a function of a random variable	309
14.11 Exercises on Chapter 14	311
14.12 Computing exercises	314
14.13 Projects	314
Summary	315
15. The normal distribution	316
15.1 Introduction	316
15.2 The standard normal distribution	317
15.3 The normal distribution as a model for data	319
15.4 Derivation of mean and variance	321
15.5 The calculation and use of standard variates when data are normally distributed	323
15.6 Exploring data	328
15.7 Residuals	329
15.8 Graphical study of symmetry and normality	331
15.9 Exercises on Chapter 15	334
15.10 Computing exercises	336
15.11 Projects	336
Summary	337
16. Sampling distributions of means and related quantities	338
16.1 The central limit theorem	338
16.2 Distribution of sample total	341
16.3 Normal approximation to the binomial distribution	343
16.4 Proportions	347
16.5 Distribution of linear combinations of normal random variables	348
16.6 Control charts	350
16.7 Exercises on Chapter 16	359
16.8 Computing exercises	362
16.9 Projects	363
Summary	363
Revision Exercises D	364

17. Significance tests using the normal distribution	371
17.1 Introduction	371
17.2 The standard normal distribution	372
17.3 A single observation from $\mathcal{N}(\mu, \sigma^2)$	374
17.4 The mean of $n$ observations from $\mathcal{N}(\mu, \sigma^2)$	376
17.5 Difference between two means from normal distributions	378
17.6 Large-sample tests	380
17.7 One-tail and two-tail tests	381
17.8 Paired comparison data	383
17.9 Tests for proportions	385
17.10 Normal approximations to nonparametric test statistics	387
17.11 Test on sample mean when variance estimated from sample	389
17.12 Paired comparisons using estimated variance	391
17.13 Tests on means of two unpaired samples using estimated variances	393
17.14 Inference when the null hypothesis is the hypothesis of interest	396
17.15 Parametric versus nonparametric tests	397
17.16 Exercises on Chapter 17	398
17.17 Computing exercises	402
17.18 Projects	403
Summary	403
18. Estimation of intervals and parameters	405
18.1 Introduction	405
18.2 Confidence interval for mean of a normal distribution	406
18.3 Confidence interval for population mean based on a large sample	408
18.4 Confidence interval for a population proportion	409
18.5 Confidence intervals for means of normal distributions when the variance $\sigma^2$ is unknown and the sample size is small	411
18.6 Investigation of a null hypothesis	414
18.7 Confidence interval for variance of a normal distribution	414
18.8 General methods of estimating parameters	416
18.9 Exercises on Chapter 18	424
18.10 Computing exercises	427
18.11 Projects	428
Summary	428
19. Hypothesis tests using the $\chi^2$ distribution	430
19.1 Random digits	430
19.2 The $\chi^2$ family of distributions	432
19.3 The rule for finding degrees of freedom in $\chi^2$ tests	434
19.4 Why the $X^2$ statistic has an approximate $\chi^2$ distribution	435
19.5 Testing the fit of a theoretical distribution to observed data	436
19.6 Testing the fit of data to a binomial distribution	438
19.7 Contingency tables	441
19.8 The $2 \times 2$ table	443
19.9 Testing the fit of a continuous distribution	446
19.10 Exercises on Chapter 19	450

19.11	Computing exercises	453
19.12	Projects	453
	Summary	453
	Revision Exercises E	454
20.	The Poisson distribution	464
20.1	Introduction	464
20.2	Two mathematical results	464
20.3	The binomial distribution as $n \rightarrow \infty, \pi \rightarrow 0$	464
20.4	The Poisson distribution as the outcome of a random process	468
20.5	The conditions for a Poisson distribution	470
20.6	Conditions in which a Poisson model breaks down	471
20.7	The mean and variance of the Poisson distribution	472
20.8	The sum of two Poisson random variables	474
20.9	Tests of goodness-of-fit to a Poisson distribution	476
20.10	The normal approximation to the Poisson distribution	478
20.11	Testing a hypothesis about the mean in a Poisson distribution	479
20.12	Comparing two means	480
20.13	Giving limits to an estimated mean	481
20.14	The 'index of dispersion'	482
20.15	Significance tests when $\chi^2$ has a very large number of degrees of freedom	484
20.16	Exercises on Chapter 20	485
20.17	Computing exercises	488
20.18	Projects	488
	Summary	489
21.	Correlation	491
21.1	Introduction	491
21.2	Scatter diagrams	491
21.3	The correlation coefficient	495
21.4	Correlation between two random variables	499
21.5	Testing values of $r$ for significance	500
21.6	Fallacies in interpreting calculated correlation coefficients	502
21.7	Comparing two correlation coefficients	507
21.8	Confidence interval for $\rho$	509
21.9	Rank correlation	510
21.10	Testing $r_s$ for significance	512
21.11	Exercises on Chapter 21	514
21.12	Computing exercises	517
21.13	Projects	518
	Summary	519
22.	The analysis of variance	520
22.1	Introduction	520
22.2	Comparison of the means of populations: one-way analysis of variance	522



22.3	Confidence intervals and tests on means	526
22.4	The $F$ -distribution and the $F$ -test	527
22.5	A mathematical model for comparing population means	529
22.6	Two-way analysis of variance	531
22.7	Design of experiments	537
22.8	Exercises on Chapter 22	539
22.9	Computing exercises	543
22.10	Projects	544
	Summary	545
23.	Simple linear regression	547
23.1	Introduction	547
23.2	Fitting a straight line when there is error in only one variable	547
23.3	Fitting a straight line that need not pass through the origin	552
23.4	The distributions of $a$ and $b$	555
23.5	Regression	559
23.6	Prediction of the total score on two dice	559
23.7	Prediction of son's height from father's height	561
23.8	Two possible regression lines	564
23.9	An analysis of variance for regression	567
23.10	Transformations to make a model 'linear'	571
23.11	Exercises on Chapter 23	574
23.12	Computing exercises	579
23.13	Projects	580
	Summary	581
	Revision Exercises F	582
24.	Multiple regression	591
24.1	Introduction	591
24.2	Estimation of unknown parameters	593
24.3	Interpretation by analysis of variance	595
24.4	Mathematical derivation of analysis of variance identity	600
24.5	Confidence intervals and $t$ -tests for the regression coefficients	601
24.6	Regression with more than two explanatory variates	602
24.7	Regression by computer	603
24.8	Checking the fit of the regression	603
24.9	General strategies	604
24.10	Exercises on Chapter 24	605
24.11	Computing exercises	608
24.12	Project	610
	Summary	610
Appendix		613
I	The binomial series expansion	613
II	The exponential function	614
III	Derivatives and integrals of the exponential function	615
IV	Integrals related to the normal distribution	615

V	The limit of $(1 + x/n)^n$ as $n \rightarrow \infty$	617
VI	A derivation of the Poisson distribution	617
VII	Partial differentiation	618
Bibliography		620
I	Particular topics	620
II	Reference texts	621
Answers		622
Hints on computing exercises		638
Tables		651
Index		662