A Concise Course in A-Level Statistics

With Worked Examples

J. CRAWSHAW
J. CHAMBERS

SECOND EDITION

ELBS

FUNDED BY THE BRITISH GOVERNMENT

LOW-PRICED EDITION

A Concise Course in A-Level Statistics

With Worked Examples

Second Edition

J. CRAWSHAW BSc
Head of Mathematics Department, Clifton High School, Bristol

J. CHAMBERS MA
Head of Mathematics Department, Sutton High School, Surrey

ELBS with Stanley Thornes
Educational Low-Priced Books Scheme
funded by the British Government

Stanley Thornes (Publishers) Ltd
Ellenborough House, Wellington Street,
Cheltenham GL50 1YD

First published 1984
Reprinted 1985, 1986, 1987, 1988, 1989
Second edition 1990
Reprinted 1991, 1992, 1993

ELBS edition first published 1987
Reprinted 1988, 1989
ELBS edition of second edition 1991
Reprinted 1992, 1993

ISBN 0 7487 1207 0

Printed and bound in Great Britain at
The Bath Press, Avon

A Concise Course in A-Level Statistics

The Educational Low-Priced Books Scheme is funded by the Overseas Development Administration as part of the British Government overseas aid programme. It makes available low-priced, unabridged editions of British publishers' textbooks to students in developing countries. Below is a list of some other books on mathematics and statistics published under the ELBS imprint.

Bostock and Chandler
Further Mechanics and Probability
Stanley Thornes

Bostock and Chandler
Mathematics — The Core Course for A-Level
Stanley Thornes

Bostock and Chandler
Mathematics — Mechanics and Probability
Stanley Thornes

Bostock, Chandler and Rourke
Further Pure Mathematics
Stanley Thornes

Bowers
Statistics for Economics and Business
Macmillan

Clarke and Cooke
A Basic Course in Statistics
Edward Arnold

Francis
Business Mathematics and Statistics
DP Publications

Jeffrey
Mathematics for Engineers and Scientists
Chapman & Hall

Stephenson
Mathematical Methods for Science Students
Longman

Whitehouse
A-Level Pure Mathematics
HLT Publications

PREFACE

This text is intended primarily for use by students and teachers of the statistics section of A-level Pure Mathematics with Statistics, an increasingly popular course.

Points of theory are presented concisely and illustrated by suitable worked examples, many taken from previous A-level papers. These are then supported by very carefully graded exercises which serve to consolidate the theory, link it with previous work and build up the confidence of the reader. There are frequent summaries of main points and miscellaneous exercises containing mainly A-level questions.

Throughout the text we have aimed to provide the reader with a mathematical structure and a logical framework within which to work. We have given special attention to topics which, in our experience, cause great difficulty. These include probability theory, the theory of continuous random variables and significance testing.

The text covers the main theory required by all the major examining boards. We are very grateful to the following for permission to reproduce questions:

University of Cambridge Local Examinations Syndicate (C)
The Southern Universities' Joint Board (SUJB)
Joint Matriculation Board (JMB)
University of London (L)
Oxford and Cambridge School Examinations Board (O & C)
incorporating School Mathematics Project (SMP)
Mathematics in Education and Industry (MEI)
The Associated Examining Board (AEB)
Oxford Delegacy of Local Examinations (O)

A-level questions are followed by the name of the board. Questions from Additional Mathematics papers are indicated by the word Additional, and (P) indicates a part-question.

We are particularly indebted to The Associated Examining Board and The Southern Universities' Joint Board for allowing us to use some of their questions as worked examples, and would stress that they are in no way involved in, or responsible for, this working.

We extend our thanks to our families, colleagues and students for all their encouragement and support, in particular to Audrey Shepherd and Jane Ziesler.

J Crawshaw
J Chambers

PREFACE TO THE SECOND EDITION

In order to give a fully comprehensive coverage of the present A-level syllabuses the following material has been added:

Chapter 4 — The use of binomial and Poisson cumulative probability tables. The geometric distribution
Chapter 5 — The negative exponential distribution
Chapter 6 — The use of the standard normal cumulative tables $\Phi(z)$ (with the use of tables giving $Q(z)$ retained in the Appendix)
Chapter 7 — Random sampling and the use of random number tables
Chapter 9 — Significance testing relating to the binomial and Poisson distributions
Chapter 11 — A fuller treatment of correlation and linear regression, including significance testing relating to Spearman's and Kendall's coefficients of correlation.

Numerous recent A-level questions taken from all the major examining boards have been added, together with worked examples from the University of London Schools Examination Board which we would stress is in no way responsible for these solutions.

J Crawshaw
J Chambers
1990

CONTENTS

PREFACE ix

1 DESCRIPTIVE STATISTICS 1
Discrete and Continuous Data 1
Frequency Distribution 2
Histograms 5
Circular Diagrams or Pie Diagrams 11
Frequency Polygons 15
Frequency Curves 16
Cumulative Frequency 17
Cumulative Frequency Curve 18
The Median 22
Quartiles, Percentiles, Interquartile Range 29
The Mode 34
The Arithmetic Mean 37
The Mean Deviation from the Mean 46
The Standard Deviation 48
The Variance 48
The Use of Calculators 51
Scaling Similar Sets of Data 54
Combining Sets of Numbers 56
Miscellaneous Worked Examples 66

2 PROBABILITY 78
Classical Definition of Probability 78
Important Results 79
Mutually Exclusive Events 86
Exhaustive Events 89
Conditional Probability 90
Independent Events 94
Summary — Probability Laws 100
Extension of Results to more than Two Events 102
Probability Trees 108
Bayes' Theorem 115
Some Useful Methods 120
Arrangements 124
Permutations of r Objects from n Objects 130
Combinations of r Objects from n Objects 131
Summary — Arrangements, Permutations and Combinations 138
Miscellaneous Worked Examples 140

3 PROBABILITY DISTRIBUTIONS I — DISCRETE RANDOM VARIABLES 167
Discrete Random Variable 167
Probability Density Function 168
Expectation, $E(X)$ 171
Expectation of any Function of X, $E[g(X)]$ 179
Variance, $\mathrm{Var}(X)$ 183
The Cumulative Distribution Function 189
Two Independent Random Variables 192
The Distribution of $X_1 + X_2$ 197
Comparing the Distributions of $2X$ and $X_1 + X_2$ 200
Summary — Discrete Random Variables 204

4 SPECIAL DISCRETE PROBABILITY DISTRIBUTIONS 209
The Binomial Distribution 209
Expectation and Variance 214
Diagrammatic Representation of the Binomial Distribution 217
Cumulative Binomial Probability Tables 219
The Recurrence Formula 221
Fitting a Theoretical Distribution 225
Worked Example 229
Summary — Binomial Distribution 233
The Geometric Distribution 235
Expectation and Variance 236
The Poisson Distribution 242
Expectation and Variance 243
Uses 245
Unit Interval 246
Cumulative Poisson Probability Tables 251
Diagrammatic Representation of the Poisson Distribution 252
The Recurrence Formula 254
Fitting a Theroetical Distribution 255
The Distribution of Two Independent Poisson Variables 258
Miscellaneous Worked Examples 262
Summary — Poisson Distribution 268

5 PROBABILITY DISTRIBUTIONS II — CONTINUOUS RANDOM VARIABLES 272
Probability Density Function 272
Expectation 275
Variance 281
The Mode 286
Cumulative Distribution Function $F(x)$ 287
Obtaining the p.d.f. from the Cumulative Distribution 295
The Rectangular Distribution 302
Expectation and Variance 304
The Exponential Distribution 307
Expectation and Variance 308
The Link Between the Exponential Distribution and the Poisson Distribution 314

The Normal Distribution 317
Expectation and Variance 317
Miscellaneous Worked Examples 320
Summary — Continuous Random Variables 326

6 THE NORMAL DISTRIBUTION 331
Probability Density Function of Normal Variable 331
The Standard Normal Distribution 332
The Probability Density Function for Z, $\phi(z)$ 333
The Cumulative Distribution Function for Z, $\Phi(z)$ 333
Use of the Standard Normal Tables using $\Phi(z)$ 334
Use of the Standard Normal Tables for any Normal Distribution 338
Problems which Involve Finding the Value of μ or σ or both 346
Miscellaneous Worked Examples 350
The Normal Approximation to the Binomial Distribution 355
The Normal Approximation to the Poisson Distribution 362
When to Use the Different Approximations 364

7 RANDOM VARIABLES AND RANDOM SAMPLING 374
Sum and Difference of Two Independent Normal Variables 374
Extension to More than Two Independent Normal Variables 377
Multiples of Normal Variables 383
Distinguishing Between Multiples and Sums of Random Variables 385
Miscellaneous Worked Examples 388
Summary — Sums Differences and Multiples of Independent Normal Variables 391
The Sample Mean 395
Sampling Without Replacement 397
The Distribution of the Sample Mean
(a) from a normal population 399
(b) from any population 403
The Distribution of the Sample Proportion 406
Summary — The Sample Mean and the Sample Proportion 410
Random Sampling 410
Random Number Tables 411
Sampling from Given Distributions
(a) frequency distributions 412
(b) probability distributions 413

8 ESTIMATION OF POPULATION PARAMETERS 420
Point Estimation — Unbiased Estimator 420
Consistent Estimator 423
The Most Efficient Estimator of the Population Mean 425
The Most Efficient Estimator of the Population Variance 425
Estimator of Population Proportion 428
Pooled Estimators from Two Samples
mean and variance 429
proportion 431
Summary — Point Estimators 433
Interval Estimation — Confidence Intervals 433
Confidence Interval for the Population Mean μ
(a) σ^2 known 434
(b) (i) σ^2 unknown, sample size large 437
The t-distribution 441
Use of t-distribution Tables 442
Confidence Interval for the Population Mean μ
(b) (ii) σ^2 unknown, sample size small 444
Confidence Interval for the Proportion of Successes in a Population 448
Summary — Confidence Intervals 455

9 SIGNIFICANCE TESTING 458
Null and Alternative Hypotheses 458
Critical Regions and Critical Values 459
One-tailed and Two-tailed Tests 459
Testing a Single Sample Value 462
Testing a Mean
(1) σ^2 known 465
(2) σ^2 unknown, sample size large 473
(3) σ^2 unknown, sample size small 477
Testing the Difference between Means 481
Testing a Proportion 492
Testing the Difference between Proportions 497
Summary — Significance Testing 503
Tests Involving the Binomial Distribution 506
Summary — Testing a Binomial Proportion n not large 513
Tests Involving the Poisson Distribution 514
Summary — Testing a Poisson Mean 517
Type I and Type II Errors 518

10 THE χ^2 TEST 533
The Chi-squared Distribution 533
The χ^2 Test 534
Uniform Distribution 536
Distribution in a Given Ratio 539
Goodness of Fit Tests 540
Binomial Distribution, p known 540
Binomial Distribution, p unknown 542
Poisson Distribution 543
Normal Distribution, μ and σ^2 known 545
Normal Distribution, μ and σ^2 unknown 546
Use of χ^2 Test in Contingency Tables 548
2×2 Contingency Tables 548
$h \times k$ Contingency Tables 551
Summary — χ^2 Test and Degrees of Freedom 554

11 REGRESSION AND CORRELATION 559
Scatter Diagram 559
Regression Function 559
Linear Correlation and Regression Lines 559
Drawing a Regression Line 'by Eye' 560
Calculating the Equations of the Least Squares Regression Lines 564
Covariance 567
Alternative Method for Least Squares Regression Lines 568
Minimum Sum of Squares of Residuals 569
The Product-moment Correlation Coefficient 576
Alternative Method for the Minimum Sum of Squares of Residuals 579
Relationship Between Regression Coefficients and r 581
Using a Method of Coding 588
Coefficients of Rank Correlation 591
Spearman's Coefficient of Rank Correlation r_S 591
Significance of Spearman's Rank Correlation Coefficient 596
Kendall's Coefficient of Rank Correlation r_k 605
Significance of Kendall's Rank Correlation Coefficient 607
Miscellaneous Worked Examples 613
Summary — Regression and Correlation 620

APPENDIX 1 629
Random Numbers 629
Cumulative Binomial Probabilities 630
Cumulative Poisson Probabilities 632
The Distribution Function $\Phi(z)$ of the Normal Distribution N(0, 1) 634
Upper Quantiles $z_{[p]}$ of the Normal Distribution N(0, 1) 635
Upper Quantiles $t_{[p]}$ of the t-Distribution $t(v)$ 636
Chi-squared Tables 637
Tables A, B, C 638
The Upper Tail Probabilities $Q(z)$ of the Normal Distribution N(0, 1) 640

APPENDIX 2 641
Use of the Standard Normal Tables using $Q(z)$ 641
Use of the Standard Normal Tables for any Normal Distribution 644
Problems which Involve Finding the Value of σ or μ or both 650

ANSWERS 654

INDEX 675

1

DESCRIPTIVE STATISTICS

DISCRETE DATA

These are the marks obtained by 30 pupils in a test:

6	3	5	9	0	1	8	5	6	7	4	4	3	1	0
2	2	7	10	9	7	5	4	6	6	2	1	0	8	8

This is an example of **discrete raw data.**

Discrete data can assume only exact values, for example

the number of cars passing a checkpoint in a certain time,

the shoe sizes of children in a class,

the number of tomatoes on each of the plants in a greenhouse.

The data is 'raw' because it has not been ordered in any way.

To illustrate the data more concisely, a **frequency distribution** can be formed. We count the number of 0's, 1's, 2's, . . . , and form a table:

Mark	0	1	2	3	4	5	6	7	8	9	10	
Frequency	3	3	3	2	3	3	4	3	3	2	1	Total 30

Discrete data can be grouped into 'classes', but once this has been done some of the original information is lost:

Mark	0–1	2–3	4–5	6–7	8 and over	
Frequency	6	5	6	7	6	Total 30

CONTINUOUS DATA

These are the heights of 20 children in a school. The heights have been measured correct to the nearest cm.

133	136	120	138	133
131	127	141	127	143
130	131	125	144	128
134	135	137	133	129

This is an example of **continuous raw data.**

Continuous data cannot assume exact values, but can be given only within a certain range or measured to a certain degree of accuracy, for example

144 cm (correct to the nearest cm) could have arisen from any value in the range $143.5 \text{ cm} \leqslant h < 144.5 \text{ cm}$.

Other examples of continuous data are

the speeds of vehicles passing a particular point,

the masses of cooking apples from a tree,

the time taken by each of a class of children to perform a task.

FREQUENCY DISTRIBUTIONS

To form a frequency distribution for the heights of the 20 children we group the information into 'classes' or 'intervals':

	(Alternative ways of writing the interval)	
$119.5 \leqslant h < 124.5$	119.5–124.5	120–124
$124.5 \leqslant h < 129.5$	124.5–129.5	125–129
$129.5 \leqslant h < 134.5$	129.5–134.5	130–134
$134.5 \leqslant h < 139.5$	134.5–139.5	135–139
$139.5 \leqslant h < 144.5$	139.5–144.5	140–144

The values 119.5, 124.5, 129.5, . . . , are called the **class boundaries.**

NOTE: the upper class boundary (u.c.b.) of one interval is the lower class boundary (l.c.b.) of the next interval.

The width of an interval = u.c.b. − l.c.b.

$$\text{Therefore the width of the first interval} = 124.5 - 119.5$$
$$= 5$$

In fact, in this example, each of the classes has been chosen so that the width is 5.

To group the heights into the following classes it helps to use a 'tally' column, entering the numbers in the first row, then the second row, and so on.

~~133~~	~~136~~	~~120~~	~~138~~
131	127	141	127
130	131	125	144
134	135	137	133
133	143	128	129

Height (cm)	Tally
$119.5 \leqslant h < 124.5$	I
$124.5 \leqslant h < 129.5$	
$129.5 \leqslant h < 134.5$	I
$134.5 \leqslant h < 139.5$	II
$139.5 \leqslant h < 144.5$	

The final frequency distribution should read:

Height (cm)	Tally	Frequency
$119.5 \leqslant h < 124.5$	I	1
$124.5 \leqslant h < 129.5$	~~IIII~~	5
$129.5 \leqslant h < 134.5$	~~IIII~~ II	7
$134.5 \leqslant h < 139.5$	IIII	4
$139.5 \leqslant h < 144.5$	III	3
		Total 20

Example 1.1 The following table gives the diameters of 40 ball-bearings, each measured in cm correct to 2 decimal places (d.p.). Form a frequency distribution by taking classes of width 0.02 cm.

3.98	3.94	3.96	3.97	4.02	3.96	3.97	3.98
3.94	3.97	3.96	3.97	4.00	4.00	3.98	3.97
3.99	3.99	4.00	3.95	4.03	3.95	4.00	4.01
3.99	3.99	3.98	4.01	3.98	4.00	4.04	4.03
3.99	4.02	4.03	4.00	3.93	4.01	4.00	3.93

Solution 1.1 The smallest value in the table is 3.93 and the largest value is 4.04. As measurements have been taken in cm correct to 2 d.p., the lowest class boundary is 3.925 cm. As the class width is 0.02 cm, the first interval must have an upper class boundary of 3.945 cm.

So we take as class boundaries 3.925, 3.945, 3.965, . . . , 4.045.

The frequency distribution is as follows:

Diameter (cm)	Tally	Frequency
$3.925 \leqslant d < 3.945$	IIII	4
$3.945 \leqslant d < 3.965$	~~IIII~~	5
$3.965 \leqslant d < 3.985$	~~IIII~~ ~~IIII~~	10
$3.985 \leqslant d < 4.005$	~~IIII~~ ~~IIII~~ II	12
$4.005 \leqslant d < 4.025$	~~IIII~~	5
$4.025 \leqslant d < 4.045$	IIII	4
		Total 40

NOTE: The intervals are often written

Diameter (cm)
3.93–3.94
3.95–3.96
3.97–3.98
and so on

Remember to work out the class boundaries.

The following frequency distributions show some of the ways in which data may be grouped.

(i) Frequency distribution to show the lengths of 30 rods. Lengths have been measured to the nearest mm.

Length (mm)	27–31	32–36	37–46	47–51
Frequency	4	11	12	3

The interval '27–31' means 26.5 mm $\leqslant$ length $<$ 31.5 mm.

The class boundaries are 26.5, 31.5, 36.5, 46.5, 51.5
The class widths are 5, 5, 10, 5

(ii) Frequency distribution to show the marks in a test of 100 students

Mark	30–39	40–49	50–59	60–69	70–79	80–89
Frequency	10	14	26	20	18	12

The class boundaries are 29.5, 39.5, 49.5, 59.5, 69.5, 79.5, 89.5
The class widths are 10, 10, 10, 10, 10, 10,

(iii) Frequency distribution to show the lengths of 50 telephone calls

Length of call (min)	0–	3–	6–	9–	12–	18–
Frequency	9	12	15	10	4	0

The interval '3–' means 3 minutes $\leqslant$ time $<$ 6 minutes, so any time including 3 minutes and up to (but not including) 6 minutes comes into this interval.

The class boundaries are 0, 3, 6, 9, 12, 18
The class widths are 3, 3, 3, 3, 6

(iv) Frequency distribution to show the masses of 40 packages brought to a particular counter at a post office

Mass (g)	–100	–250	–500	–800
Frequency	8	10	16	6

The interval '–250' means 100 g $<$ mass $\leqslant$ 250 g; so any mass over 100 grams up to and including 250 grams comes into this interval.

The class boundaries are 0, 100, 250, 500, 800
The class widths are 100, 150, 250, 300

(v) Frequency distribution to show the speeds of 50 cars passing a checkpoint

Speed (km/h)	20-30	30-40	40-60	60-80	80-100
Frequency	2	7	20	16	5

The class '30-40' means 30 km/h $\leqslant$ speed $<$ 40 km/h.

The class boundaries are 20, 30, 40, 60, 80, 100
The class widths are 10, 10, 20, 20, 20

(vi) Frequency distribution to show ages (in completed years) of applicants for a teaching post

Age (years)	21-24	25-28	29-32	33-40	41-52
Frequency	4	2	2	1	1

As the ages are in completed years (not to the nearest year) then '21-24' means 21 $\leqslant$ age $<$ 25. Someone who is 24 years and 11 months would come into this category. Sometimes this interval is written '21-' and the next is '25-', etc.

The class boundaries are 21, 25, 29, 33, 41, 53
The class widths are 4, 4, 4, 8, 12

HISTOGRAMS

Grouped data can be displayed in a histogram.

In a histogram rectangles are drawn so that the area of each rectangle is proportional to the frequency in the range covered by it.

We have area $\propto$ frequency

(a) Histograms with equal class widths

Example 1.2 The lengths of 30 Swiss cheese plant leaves were measured and the information grouped as shown. Measurements were taken correct to the nearest cm. Draw a histogram to illustrate the data.

Length of leaf (cm)	10-14	15-19	20-24	25-29
Frequency	3	8	12	7

Solution 1.2 The class boundaries are 9.5, 14.5, 19.5, 24.5, 29.5
The class widths are 5, 5, 5, 5

Now, area of rectangle = class width $\times$ height of rectangle

As the class width is 5 for each interval,

$$\text{area of rectangle} = 5 \times \text{height of rectangle}$$

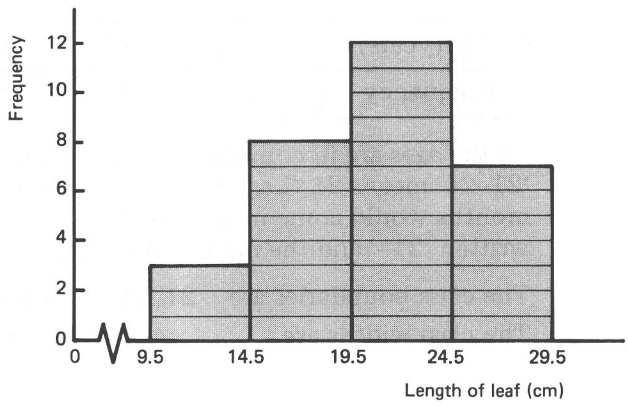
So $\text{area} \propto \text{height of rectangle}$

Now, if we make the height of each rectangle the same as the frequency,

we have $\text{area} \propto \text{frequency}$, as required.

When all the class intervals are of equal width the frequency can be used for the height of each rectangle.

Histogram to show the lengths of 30 leaves



(b) Histograms with unequal class widths

Example 1.3 The frequency distribution gives the masses of 35 objects, measured to the nearest kg. Draw a histogram to illustrate the data.

Mass (kg)	6–8	9–11	12–17	18–20	21–29
Frequency	4	6	10	3	12

Solution 1.3 The class boundaries are 5.5, 8.5, 11.5, 17.5, 20.5, 29.5

The class widths are 3, 3, 6, 3, 9

As the class widths are not equal we cannot make the height of each rectangle equal to the frequency.

So we choose a convenient width as a 'standard' and adjust the heights of the rectangles accordingly, as follows.

If we choose a class width of 3 as standard, then the first two rectangles can be 4 and 6 units high respectively. However, as the third interval is twice the standard width we must make the height of the rectangle equal to half the frequency.

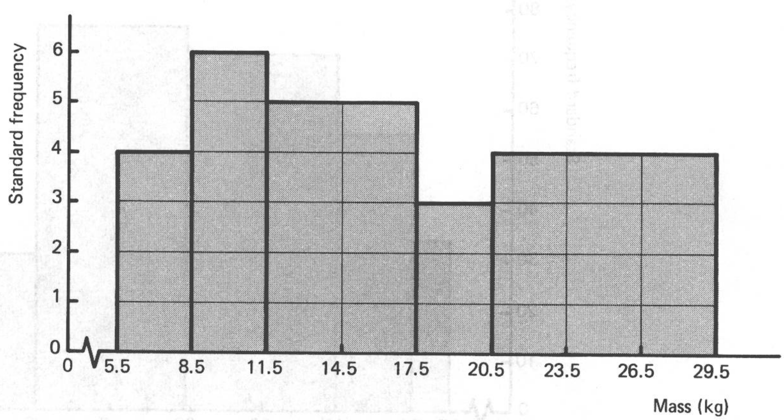
Similarly, as the last interval is $3 \times$ standard we must make the height of the rectangle equal to one-third of the frequency.

As the heights of the rectangles have been adjusted, we are considering frequency per standard width. We will write this as 'standard frequency'.

Mass (kg)	Class width		Frequency	Height of rectangle (standard frequency)
6–8	3	standard	4	4
9–11	3	standard	6	6
12–17	6	2 × standard	10	$\frac{1}{2}\times 10 = 5$
18–20	3	standard	3	3
21–29	9	3 × standard	12	$\frac{1}{3}\times 12 = 4$

We have now ensured that the area of each rectangle is proportional to the frequency, and the histogram is drawn as shown.

Histogram to show the masses of 35 objects



In general, choose a 'standard' width.

If class width $= n \times$ standard width

then height of rectangle $= \frac{1}{n} \times$ corresponding frequency

Example 1.4 The following table gives the distribution of the interest paid to 460 investors in a particular year.

Interest (£)	25–	30–	40–	60–	80–	110–
Frequency	17	55	142	153	93	0

Draw a histogram to illustrate this information.