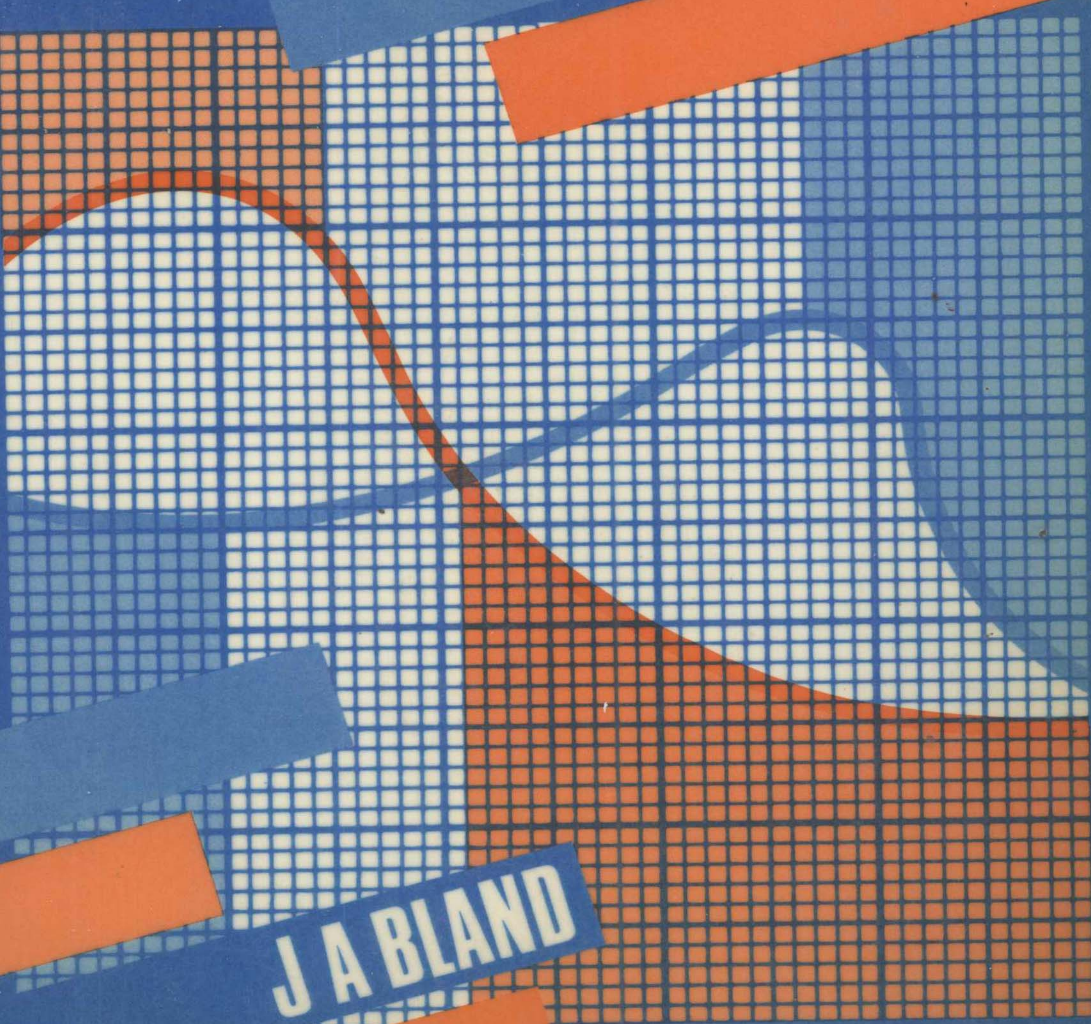
STATISTICS FOR CONSTRUCTION STUDENTS



J A BLAND

Statistics for construction students

J A Bland B.Sc., Ph.D., AFIMA
Trent Polytechnic
Nottingham

Construction Press
London and New York

General Editor: Colin Bassett BSc, FCIOB, FFB

Construction Press
an imprint of:

Longman Group Limited
Longman House, Burnt Mill, Harlow
Essex CM20 2JE, England
Associated companies throughout the world

Published in the United States of America
by Longman Inc., New York

First published 1985

British Library Cataloguing in Publication Data
Bland, J.A.
Statistics for construction students.
1. Building – Statistical methods
I. Title
519.5'024624 TH153

ISBN 0-86095-043-3

Library of Congress Cataloging in Publication Data
Bland, J.A., 1951–
Statistics for construction students.

Bibliography: p.
Includes index.
1. Engineering—Statistical methods. 2. Probabilities.
I. Title.
TA340.B58 1985 624'028 84–19946
ISBN 0–86095–043–3

Printed in Great Britain by
The Bath Press, Avon

Preface

Statistics for Construction Students is written for students of building in colleges, polytechnics and universities. It may also be of use to students of civil engineering and surveying.

The aims of this book are:

(a) To explain statistical ideas and methods;
(b) To show construction applications of statistics.

With regard to the first aim, this book concentrates on probability distributions and so is not intended to be a complete and definitive work on statistics.

Chapter 1, which gives only brief coverage of descriptive statistics, serves to introduce (in the book) terms such as mean and standard deviation, sample and population. Chapter 2, as the title states, is merely an introduction to probability and goes to no greater depth than is required for subsequent chapters. The book proper, in terms of probability distributions, commences with Chapter 3.

To achieve the second aim the statistical backgrounds of topics unique to construction (e.g. the compliance criteria, limit state design) are presented and explained. The use of these topics to integrate statistics and construction is a feature of this book. Furthermore, the overall content (i.e. explanations, worked examples, exercises) has a construction bias.

It is expected that this book will be of use to construction students studying for the following qualifications; HND(BTEC), B.Sc., and CIOB parts I and II. Although statistics is a component of construction courses it is recognized that construction students are not 'main-stream' mathematicians; the mathematical level of this book is no higher than that required of construction students studying for the above qualifications.

In conclusion I should like to thank the following people; Mrs J. Frampton, for the excellent typing of the manuscript; my wife Christine, to whom this book is dedicated, for her support and encouragement throughout the preparation stages; our daughter Victoria, who came along during the writing of Chapter 8 (a small sample?), for not slowing the preparation process any more than she could help.

J A Bland
Trent Polytechnic
Nottingham

Acknowledgements

The author wishes to thank the following for permission to reproduce copyright material:

(a) The McGraw-Hill Book Company (UK) Limited, for the normal distribution table (Table A.1).

(b) John Wiley and Sons Ltd, for the t-distribution table and χ^2-distribution table (Table A.2 and Table A.3, respectively).

(c) The British Standards Institution.

Extracts from British Standards are reproduced by permission of the British Standards Institution, 2 Park Street, London, W1A 2BS, from whom complete copies of the standards can be obtained.

Contents

Preface v
Acknowledgements vi

Chapter 1. Presentation of data 1
1.1 Introduction 1
1.2 Pictorial presentation of data 1
1.3 Graphical presentation of data 5
1.4 Grading of aggregates 9
1.5 Averages and dispersion 10
1.6 Population and sample 13

Chapter 2. Introduction to probability 16
2.1 Introduction 16
2.2 Definitions 16
2.3 Laws of probability 18
2.4 Probability distributions 20

Chapter 3. The binomial and Poisson probability distributions 23
3.1 Introduction 23
3.2 The binomial distribution 23
3.3 The Poisson distribution as an approximation 26
3.4 The Poisson distribution in its own right 27

Chapter 4. The normal probability distribution 30
4.1 Introduction 30
4.2 The normal distribution 30
4.3 The standard normal distribution 32
4.4 Strength of concrete 36
4.5 Normal probability paper 40
4.6 Structural design philosophy 43

Chapter 5. Normal sampling distributions 48
5.1 Introduction 48
5.2 The distribution of sample means 49
5.3 The compliance criteria 54
5.4 The distribution of sample mean differences 58

Chapter 6. Statistical estimation	64
6.1 Introduction	64
6.2 Confidence intervals	64
6.3 Confidence interval for the population mean	66
6.4 Accuracy in building	69
6.5 Confidence interval for the population mean difference	72
6.6 Building design	75
Chapter 7. Statistical decisions	79
7.1 Introduction	79
7.2 Hypothesis tests	79
7.3 Acceptance sampling	85
7.4 Operating characteristic curve for concrete	92
7.5 Repeatability and reproducibility	94
Chapter 8. The t-distribution	100
8.1 Introduction	100
8.2 The t-distribution	100
8.3 Estimating with small samples	103
8.4 Test statistics for small samples	105
Chapter 9. The χ^2-distribution	109
9.1 Introduction	109
9.2 The χ^2-distribution	109
9.3 Confidence interval for the population variance	111
9.4 Test statistic for the population variance	113
9.5 To test goodness of fit	114
9.6 Contingency tables	118
Chapter 10. Quality control	124
10.1 Introduction	124
10.2 Statistical control	124
10.3 Quality-control chart for sample means	125
10.4 Quality-control chart for sample ranges	130
10.5 Cusum charts	132
Chapter 11. Correlation and regression	138
11.1 Introduction	138
11.2 Correlation	139
11.3 Regression	146
11.4 Abrams' law for concrete strength	154
11.5 Confidence intervals	156
Appendix. Statistical tables	162
References	166
Index	167

Presentation of data

1.1 Introduction

Many situations occur in construction which involve a large amount of data (e.g. compressive strength values obtained from test cubes of concrete, the number of various building components used in a construction project). Often each individual item of data is unimportant by itself and must be considered in relation to the other data. In other words, when large amounts of data are considered their *relative* values are sometimes more important than their *absolute* values.

In order to investigate and analyse a large amount of data so that values, or groups of values, can be compared, statistical methods in the form of pictures and graphs are used. Data presented in this manner enables comparisons to be made 'at a glance' and information to be readily obtained.

1.2 Pictorial presentation of data

A large amount of data which has not been organized numerically is called **raw data**. As an example consider the numbers of houses built by a local authority on 100 housing schemes over a five-year period. The raw data is displayed in Table 1.1.

Table 1.1 Raw data; houses built during the period 1975–79, year of completion given in brackets

100(75)	38(75)	44(76)	35(76)	85(77)	78(77)	76(78)	280(78)	56(79)	184(79)
152(75)	86(75)	82(76)	100(76)	143(77)	121(77)	156(78)	122(78)	360(79)	63(79)
68(75)	73(75)	20(76)	64(77)	155(77)	87(77)	29(78)	78(78)	235(79)	45(79)
104(75)	62(75)	84(76)	156(77)	82(77)	61(77)	188(78)	46(78)	36(79)	226(79)
50(75)	100(75)	96(76)	94(77)	262(77)	37(77)	76(78)	56(78)	99(79)	169(79)
45(75)	64(75)	89(76)	21(77)	74(77)	63(77)	154(78)	98(78)	310(79)	80(79)
84(75)	36(75)	32(76)	91(77)	23(77)	81(77)	40(78)	74(79)	58(79)	98(79)
122(75)	55(75)	68(76)	50(77)	158(77)	19(77)	272(78)	28(79)	127(79)	30(79)
168(75)	28(75)	20(76)	172(77)	140(77)	100(77)	50(78)	127(79)	72(79)	36(79)
65(75)	35(76)	45(76)	28(77)	55(77)	93(78)	36(78)	68(79)	385(79)	34(79)

In order to make sense of the raw data it must be organized into groups. The grouping or **tabulation** of the raw data depends on what comparisons are required (or desirable). For instance, with the housing schemes, if the number of houses built per year is being investigated, then the raw data can be tabulated on a yearly basis, as shown in Table 1.2.

Table 1.2

Year	*Number of houses built*
1975	1500
1976	750
1977	2500
1978	1850
1979	3000

Several methods of presenting the tabulated data pictorially are now given. The motivation for this pictorial presentation of data is that columns of numbers can be uninspiring and do not readily indicate trends 'at a glance'.

(a) Pie diagrams

A **pie diagram** represents groups of data (here the number of houses in a given year) as a sector of a circle. The angle of the sector (and hence the area) for each group being in proportion to the value of the group.

The total number of houses built in the five-year period is

$$\begin{aligned} N &= 1500 + 750 + 2500 + 1850 + 3000 \\ &= 9600 \text{ houses} \end{aligned}$$

Hence 9600 represents 360° and the sector angles can be calculated as shown in Table 1.3.

Using the data in Table 1.3 a pie diagram can be constructed and is shown in Fig. 1.1.

Table 1.3 Calculation of sector angles

Year	*Number of houses, n*	*Fraction of total,* $\frac{n}{N}$	*Sector angle* $\left(\frac{n}{N}\right)360°$
1975	1500	0.156	56°
1976	750	0.078	28°
1977	2500	0.260	94°
1978	1850	0.193	69°
1979	3000	0.313	113°

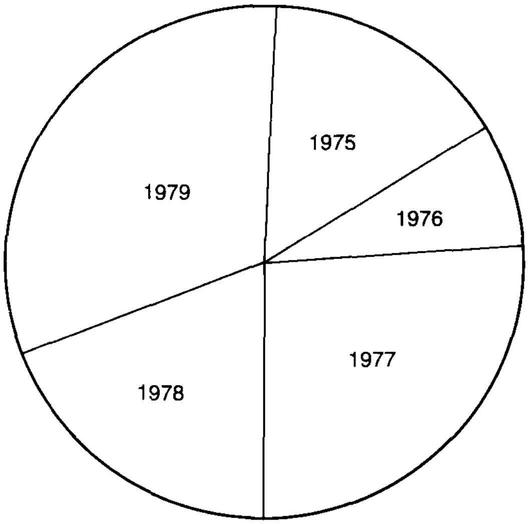


Fig 1.1 Pie diagram

A pie diagram is used to compare groups and not to determine their values. For instance, by comparing sector areas, the pie diagram in Fig. 1.1 indicates that more houses were built in 1977 than the year before but it does not reveal how many houses were built in those years.

(b) Ideographs

An **ideograph** represents each group of data as a number of suitable pictures (ideographs are sometimes called **pictograms**). The pictures are all of the same size and denote a fixed value, as shown in Fig. 1.2.

Ideographs are used to compare groups of data by observing the difference in the number of pictures representing each group. They can also give the value of the group since each picture denotes a known fixed value. However, precision is lost when a proportion of a picture is encountered.

(c) Picturegrams

Unlike ideographs, **picturegrams** represent each group of data as a single picture. The picture for each group is of the same object but its size is in proportion to the value of the group. A picturegram for the local authority housing schemes is shown in Fig. 1.3.

A picturegram is used to compare groups by observing the difference in the sizes of associated pictures. The picturegram in Fig. 1.3 can also be used to indicate the value of each group since the reference base length denotes 1000 houses. However, if no key were given then the picturegram could not indicate the group values and even comparison becomes difficult because there would be confusion as to whether lengths or areas were being compared.

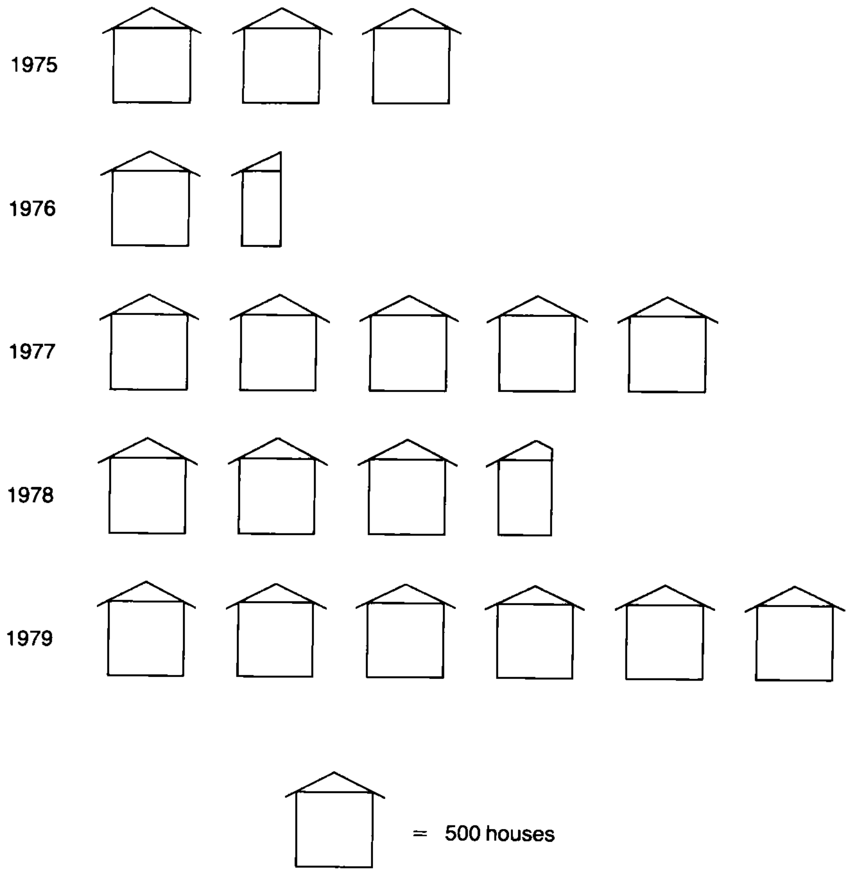


Fig 1.2 Ideograph

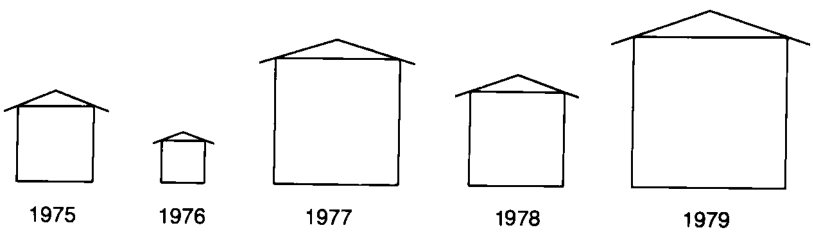


Fig 1.3 Picturegram

(d) Bar charts

Bar charts represent the value of each group of data as a length of bar, as shown in Fig. 1.4.

A bar chart is one of the most useful ways of presenting data since groups are easily compared by observing the differences in bar heights and, with reference to the given vertical scale, their values can be determined.

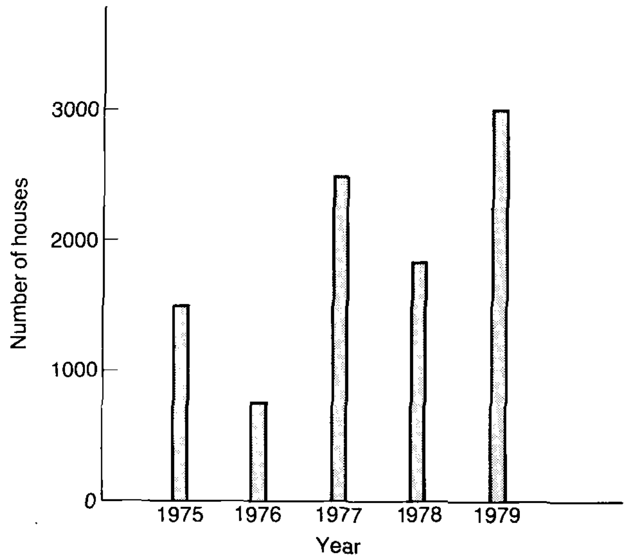


Fig 1.4 Bar chart

1.3 Graphical presentation of data

In the previous section the presentation of data concentrated mostly on comparison; in this section numerical values will be equally important.

If, for example, 100 pieces of timber are delivered to a building site and their lengths are measured and recorded (to the nearest centimetre), then the results can be summarized in a **frequency distribution** (or **frequency table**) as shown in Table 1.4.

Table 1.4 Frequency distribution

Length (cm)	*Number of pieces of timber*
60–62	5
63–65	18
66–68	42
69–71	29
72–74	6

The first column of Table 1.4 contains the length **classes** and the second column gives the **class frequencies** (i.e. the number of pieces of timber in each class).

With reference to Table 1.4, the interval denoted by 60–62 is called the **class interval**; 60 and 62 being the **class limits**. Now since the timber lengths are recorded to the nearest centimetre the class frequency includes all measurements between exactly 59.5 cm and just less than 62.5 cm (a piece of timber with length exactly 62.5 cm would be included in the class interval 63–65). Hence 59.5 and 62.5 are called the **class boundaries** (see Fig. 1.5) and the **class width** is 62.5−59.5 = 3 cm. The mid-point between the boundaries is called the **class mid-point** and is given by (59.5 + 62.5)/2 = 61 cm.

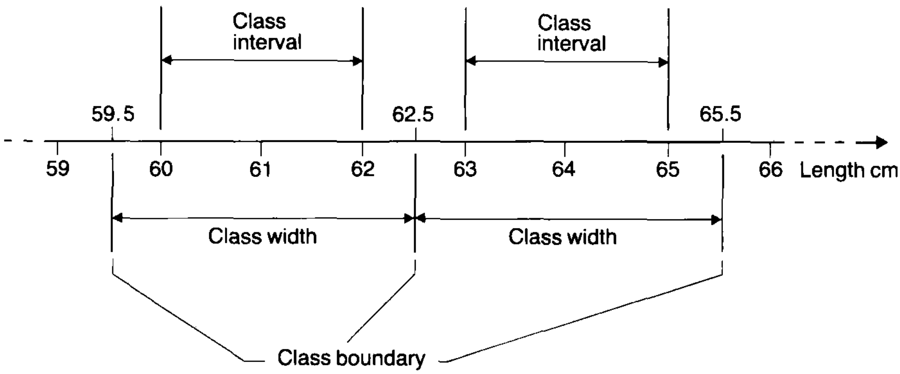


Fig 1.5 Class boundary, class interval and class width

The data in the frequency distribution can be presented graphically in the following ways.

(a) Histogram and frequency polygon

A **histogram** for the data in Table 1.4 is shown in Fig. 1.6. It comprises vertical columns drawn side by side on a horizontal axis denoting length. The columns meet the horizontal axis at the class boundaries and their heights denote class frequency as measured by the vertical scale.

A **frequency polygon** corresponding to Table 1.4 is shown by the broken lines in Fig. 1.6; the mid-points of the tops of the columns are connected by straight lines. By considering class intervals 57–59 and 75–77 with zero frequency the frequency polygon is 'anchored' to the base at class mid-points 58 cm and 76 cm, respectively (this could not be achieved if the class intervals in Table 1.4 were unequal).

(b) Relative frequency histogram

The **relative frequency** of a class is the frequency of the class divided by the total frequency of all classes (and is sometimes expressed as a percentage). For example, the relative frequency of the first class in Table 1.4 is 5/100 = 0.05 (or 5%). Clearly the sum of all relative frequencies is 1 (or 100%).

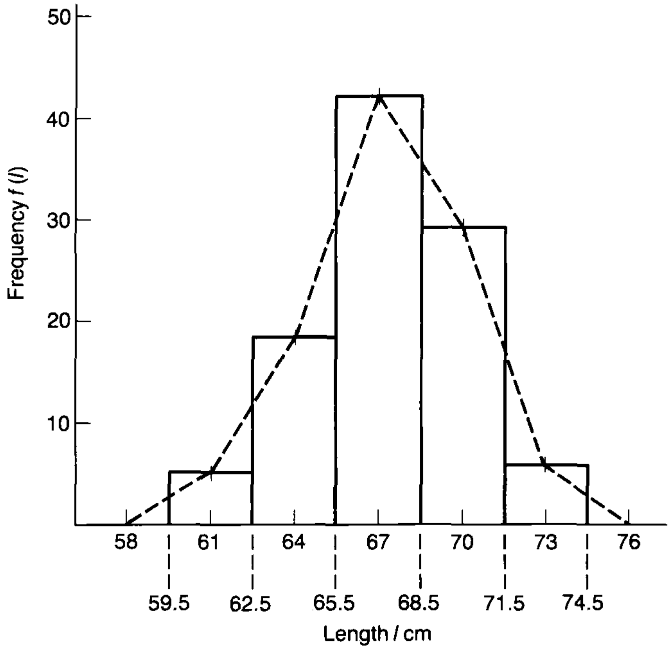


Fig 1.6 Histogram and frequency polygon

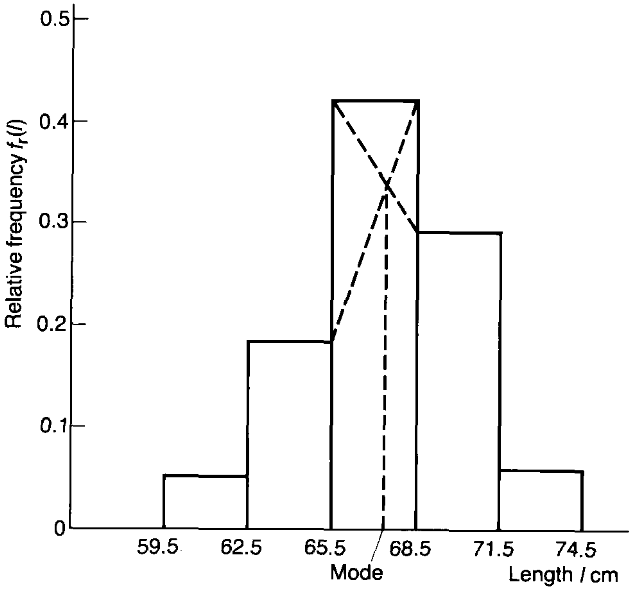


Fig 1.7 Relative frequency histogram

A **relative frequency histogram** can be obtained from the histogram in Fig. 1.6 by simply changing the vertical scale from frequency, $f(l)$, to relative frequency, $f_r(l)$, keeping exactly the same diagram; see Fig. 1.7.

(c) Cumulative frequency curve

The total frequency of all values less than the upper class boundary of a given class interval is called the **'less than' cumulative frequency**. For example, the 'less than' cumulative frequency for the class interval 66–68 in Table 1.4 is 5 + 18 + 42 = 65 (i.e. 65 pieces of timber have length less than 68.5 cm). Table 1.5 gives the 'less than' cumulative frequencies for the classes in Table 1.4.

A graph of the cumulative frequency less than an upper class boundary plotted against the upper class boundary is called a **'less than' cumulative frequency curve** (or **'less than' ogive**). The 'less than' cumulative frequency curve for the data in Table 1.4 and Table 1.5 is shown in Fig. 1.8.

Although less usual, an 'or more' cumulative frequency curve (or 'or more' ogive) can be obtained by plotting the cumulative frequency equal to or more than a lower class boundary against the lower class boundary.

Table 1.5 Cumulative frequency distribution

Length (cm)	*Cumulative frequency*
less than 59.5	0
less than 62.5	5
less than 65.5	23
less than 68.5	65
less than 71.5	94
less than 74.5	100

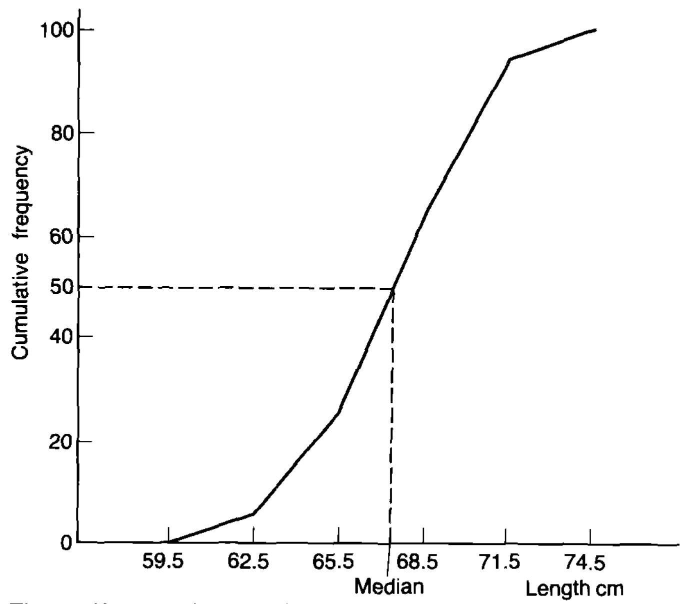


Fig 1.8 'Less than' cumulative frequency curve

1.4 Grading of aggregates

The term 'aggregates' is used to describe the gravels, crushed stones and other materials which are mixed with cement and water to make concrete.

Aggregates usually constitute between 50% and 80% of the volume of concrete and so may greatly influence properties such as strength and shrinkage. Hence the selection of suitable aggregates (with particular attention given to particle size) is important in concrete production.

The proportions of the difference sizes of particles that constitute an aggregate are found by sieving and are known as the *grading* of the aggregate.

As specified in BS 410 (1976) the aperture sizes of the test sieves used in the grading of aggregates are 50.0, 37.5, 20.0, 14.0, 10.0, 5.0 mm made with perforated plate and 2.36, 1.18 mm, 600, 300, 150 μm made with woven wire.

In BS 882 (1983) distinction is made between aggregates;

(a) coarse aggregate is mainly retained on a 5.0 mm BS 410 test sieve
(b) fine aggregate mainly passes a 5.0 mm BS 410 test sieve
(c) all-in aggregate is a mixture of fine and coarse aggregate.

Also, BS 882 specifies grading limits for the above classes of aggregate; those for fine aggregate are given in Table 1.6.

As shown in Table 1.6 a wide range of gradings of fine aggregate is acceptable for concrete (C, M and F indicate a progressively finer material).

Table 1.6 Grading limits for fine aggregate; after BS 882

	Percentage by mass passing BS sieve			
		Additional limits for grading		
Sieve size	*Overall limits*	*C*	*M*	*F*
10.00 mm	100	—	—	—
5.00 mm	89–100	—	—	—
2.36 mm	60–100	60–100	65–100	80–100
1.18 mm	30–100	30–90	45–100	70–100
600 μm	15–100	15–54	25–80	55–100
300 μm	5–70	5–40	5–48	5–70
150 μm	0–15	—	—	—

The data in Table 1.6 becomes more informative when the grading limits are presented graphically as 'percentage passing' cumulative frequency curves. The overall grading limits for fine aggregate are shown in broken line in Fig. 1.9 together with grade C additional limits in continuous line.

Similar cumulative frequency curves can be constructed for the grading limits of coarse and all-in aggregate.

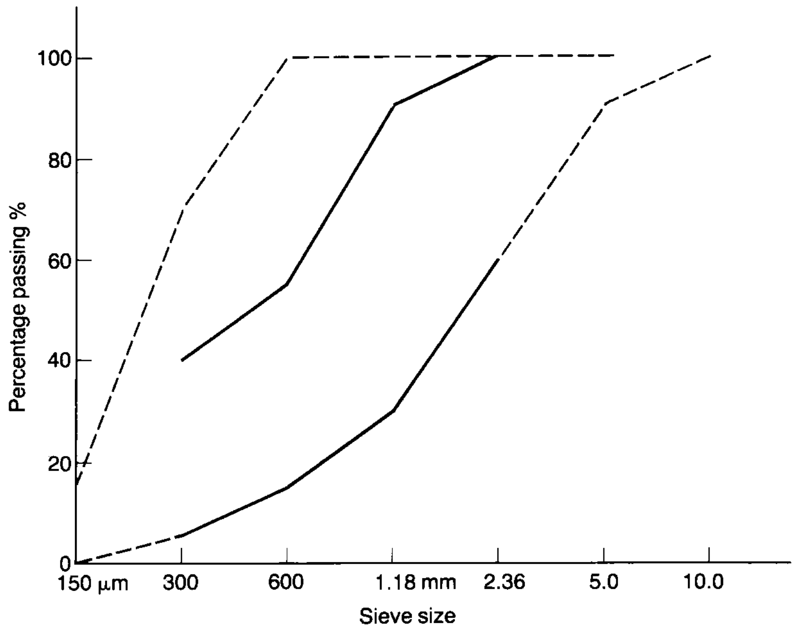


Fig 1.9 Overall and additional grading limits for grade C fine aggregate

1.5 Averages and dispersion

An **average** is a general term used to denote a value which is typical or representative of a set of data. There are several types of average; the **mean** (i.e. arithmetic mean), the **median** and the **mode**.

The degree to which numerical data tend to spread about an average value is called the **dispersion** of the data, the most common measures of dispersion being the **range** and the **standard deviation**.

The statistical methods used to determine an average and dispersion of a set of data differ according to whether the data is considered as individual values or as tabulated data in a frequency distribution, that is, **ungrouped** or **grouped data**, respectively.

(a) Ungrouped data

For a set of n values $x_1, x_2, ..., x_n$ the mean, range and standard deviation are defined mathematically as

$$\text{mean} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\text{range} = \max\{x_i\} - \min\{x_i\}$$