

BRITISH STANDARD 600 : 1935

The application of
**STATISTICAL
METHODS**
to Industrial Standardization
and Quality Control

By E. S. PEARSON, D.Sc.

NO COPYING IN ANY FORM WITHOUT WRITTEN PERMISSION FROM BSI



THE APPLICATION OF
STATISTICAL METHODS
TO INDUSTRIAL
STANDARDIZATION
AND QUALITY CONTROL

by E. S. Pearson, D.Sc.

B.S. 600 : 1935

Incorporating amendments issued
August, 1957 and October, 1960

BRITISH STANDARDS INSTITUTION

Incorporated by Royal Charter

2 PARK STREET, LONDON, W1A 2BS

Telex: 266933

Telephone: 071-629 9000

UDC 311:519.2:389.6:658-56

CONTENTS

	Page
Foreword	7
Foreword to 1959 printing	9
Introduction	11

SECTION 1

<i>Scope of the Publication</i>	14
---	----

SECTION 2

Illustration of general problems by examples

(1) Example 1. Tensile strength of malleable iron castings	18
(2) Example 2. Breaking strength of cloth.	22
(3) Example 3. Ash-content in samples of coal	24

SECTION 3

Introduction to the simpler statistical tools

(4) Definitions	28
(5) The presentation of data	30
Dot diagram	30
Frequency distribution	30
Example A. Percentage ash-content in 250 test analyses of coal	30
Example B. Tensile strength in malleable iron castings	30
Example C. Depth of sapwood in telephone poles	30
(6) Use of the mean and the standard deviation in scaling a frequency distribution. The Normal curve	32

SECTION 4

Sampling and statistical theory

(7) Statistical uniformity and random sampling; levels of control	36
(8) Sampling methods in practice	40

SECTION 5

<i>Statistical theory and specification</i>	Page
(9) The statistical concept of a population	45
(10) The basis of securing conformity to specification	46
(a) By a system of tests of samples	46
(b) By requiring that records are kept which will provide statistical evidence of the level of control of the processes of manufacture	47
Considerations of importance to user	47
Considerations of importance to the producer	48

SECTION 6

<i>The statistical relationship between sample and batch</i>	
(11) The variation of the mean and the standard deviation in samples	50
Theoretical formulae	51
(a) Variation of means	51
(b) Variation of standard deviations	51
(12) The reliability of a mean estimated from a repre- sentative sample; duplicate sampling	55
Estimate of reliability of a mean obtained from a comparison of duplicate samples	57
(13) Illustration of the use of the mean strength, and the strength of the weakest strip, in a sample of pre- scribed size, in specifying the quality of cloth	58
(14) Fiducial or confidence limits for means, standard deviations and coefficients of variation	61
(15) Simultaneous variation in the mean and the standard deviation	64

SECTION 7

<i>Further examples of the use of statistical methods in securing conformity with specifications by consignment sampling</i>	
(16) Systems of grading or classifying	67
(i) Grading by the mean and the standard deviation	68
(ii) Grading by the mean and the coefficient of variation	69
(iii) Grading based on a lower limit	70
(17) The efficiency of a classification applied to random samples from consignments	70

SECTION 8

<i>Control charts</i>	Page
(18) Statistical analysis as a tool for locating sources of trouble; assignable and chance causes of variation	74
(19) The basis of the control charts; methods of estimating σ	76
Methods of obtaining σ_e	78
(a) Estimate of σ from the mean value of the squared standard deviations	78
(b) Estimate of σ from the mean value of the standard deviation	78
(c) Estimate of σ from the mean value of the range	78
(20) Description of the different forms of chart	79
(a) The chart for means	80
(b) The chart for standard deviations	80
(c) The chart for coefficients of variation	81
(d) The chart for range	84
(21) General remarks on the interpretation of the charts .	85
(22) Example 1. Ash-content in coal	87
(23) Example 2. Lamp quality	90
(24) Example 3. Quality of a manufactured chemical . .	93
(25) Example 4. Breaking strength of a roofing material .	95

SECTION 9

<i>Performance tests and indices of quality</i>	
(26) The statistical concept of correlation	98

APPENDICES

I. Definitions	103
II. Definitions in algebraic form	105
III. The calculation of the mean and the standard deviation	107
IV. The Normal curve; its equation and probability tables	113
V. Further constants and tables associated with the Normal distribution	116
VI. Methods of testing whether the variation in a quality characteristic is Normal or not	118
VII. Errors arising from the use of the tables of control limits and confidence limits owing to lack of Normality in the variation of a characteristic . .	124
VIII. Statistical reference books, tables and British Standards	126

FIGURES

	Page
1. & 2. Malleable iron castings	127
3. Breaking strength in pounds of cotton fabric strips . .	129
4. Ash-content in coal samples	129
5. Histograms of frequency distributions	131
6. Illustration of approach to limit	133
7. Diagram indicating types of variation in samples . .	135
8. Ash-content in coal. Simultaneous variation in means and standard deviations (samples of 10)	135
9. Types of grading	137
10. Specification levels	137
11. Diagram illustrating the use of the limits of a control chart to detect changes in the level of control . . .	139
12. Control charts: percentage of ash in coal	141
13. Control charts for length of life of lamps of marked voltage	143
14. Control chart, product D	145
15. Breaking strength of a roofing material	147
16. Correlation between hardness and strength	149
17. Comparison of the normal with an asymmetrical or skew curve	149

TABLES

1. Tensile strength in pounds per square inch of malleable iron castings	19
2. 75 test results of malleable iron castings corresponding to source 4 rearranged in order from min. to max. . .	19
3. Breaking strength, in pounds, of cotton fabric strips . .	23
4. Ash-content data: means in groups of 10 tests	26
5. Frequency distributions	31
6. Percentage of observations falling outside limits deter- mined by the standard deviation	33
7. Ash-content data: standard deviations and means in groups of 10 tests	53

	Page
8. Ash-content data: frequency distribution of means and standard deviations in groups of 10 tests	54
9. Cloth samples: testing rule based on mean of sample	60
10. Cloth samples: testing rule based on weakest strip in sample	60
11. Fiducial or confidence limits	65
12. Compressive strengths of sand lime bricks in lb per sq. in.	69
13. Factors for calculation of control limits for mean, standard deviation, and coefficient of variation	82
14. Factors for calculation of control limits for range	85
15. Percentage ash-content in two coals	88
16. Lamp data: length of life at marked voltage in hours. (5 lamps in a sample)	92
17. Lamp data: illustrating the calculation of σ_e	94
18. Product D: statistical constants relating to control charts	96
19. Calculation of standard deviation. Schemes A and B	108
20. Calculation of standard deviation. Scheme C	109
21. Squares of numbers, 1-300	110
22. Tables of the Normal probability curve. (1) Values of chances for intervals of 0.1 of t (2) Values of t corresponding to certain simple chances	114
23. Estimate of σ from the mean deviation of small samples	116
24. Further frequency distributions	119
25. Application of a rough test of Normality	123
26. Effect on control limits of non-normal variation	124

FOREWORD

At a Conference held on May 11th, 1932, the British Standards Institution invited representatives of manufacturing industries and others interested in the application of statistical methods to meet Dr. W. A. Shewhart, of the Research Laboratories of the Bell Telephone Company of America, for the purpose of discussing means whereby closer co-operation might be secured between workers engaged on the one hand in the theoretical development of statistical methods and on the other in their practical application to problems of standardization and particularly in the specification of quality. As a result of the deliberations of this conference it was decided to appoint a Committee charged with the following Terms of Reference:

1. To report on the application and use of statistical methods in standardization and specification of quality.
2. To draw up a short report which would serve to awaken interest in the application of statistical methods on the part of manufacturers and others concerned with problems of standardization and specification.
3. To consider what encouragement is necessary for the development of research on improved statistical methods and their application by industry.
4. To consider what steps should be taken to provide for co-operation with bodies in the United States of America and elsewhere instituted for similar objects.

The Committee was constituted as follows:

Mr. B. H. WILSDON (Chairman), Building Research Station.

Mr. L. R. CONNOR ...	... Imperial Chemical Industries, Ltd.
Mr. B. P. DUDDING...	... Research Laboratories, General Electric Co. Ltd.
Dr. E. S. GRUMELL ...	... Imperial Chemical Industries, Ltd.
Dr. E. S. PEARSON ...	... University College, London.
Mr. J. W. RYDE ...	... Research Laboratories, General Electric Co. Ltd.
Dr. E. C. SNOW ...	... United Tanners Federation, and Hon. Sec. Royal Statistical Society.

A review of statistical methods applicable in the problems considered, which was called for under Items 1 and 2 of the Terms of Reference, is contained in the present publication; this has been primarily the work of Dr. E. S. Pearson and on the recommendation of the Committee is published under his name. The treatment of the subjects and the methods advocated have been debated at length by the Committee, who are grateful to Dr. Pearson for his patient consideration and willing co-operation in the shaping of the publication into its present form.

Questions of terminology and notation, for which there is, as yet, no generally accepted practice have also been considered. Whilst it would be premature to raise the question of international agreement, it is hoped that the proposals contained in this publication may prove of practical interest to workers outside Great Britain.

The Committee desires to call attention to the formation, in 1933, by the Royal Statistical Society, of an Industrial and Agricultural Research Section specially devoted to the study of the problems of applied mathematical statistics involved both in scientific research and in routine technical control in manufacturing industries and in agriculture. The Section affords a medium for the publication of research into statistical methods and for discussion on the application of these methods to practical problems, and its formation is warmly welcomed by the Committee.

The workers in the statistical field are few at present, but our flourishing schools of applied statistics may be expected to furnish all immediate requirements for the recruitment of specialist mathematical statisticians. A wider field must however be cultivated if the hopes which the Committee entertain of the fruitful application of statistical methods in industry are to be fulfilled. The technician must himself learn to recognize, among the multifarious problems which arise, those which call for statistical treatment. It is to gain the interest of this latter class of worker that the form of presentation adopted by Dr. Pearson in this publication has been chosen. No attempt has been made to provide a comprehensive introduction to the methods of statistical analysis; rather the aim has been to demonstrate the necessity and usefulness of such methods in problems of standardization. It is the belief of the Committee that the simple treatment of the elementary technique of statistical analysis provided cannot fail to convince those concerned with the control of manufacturing processes or with the specification of quality of variable materials that a valuable tool is within the reach of all.

B. H. WILSDON,

*Chairman of Committee on Statistical
Methods in Standardization and
Specification.*

FOREWORD TO 1960 PRINTING

In the 25 years which have elapsed since this standard was first drawn up there has been an immense change in the outlook of British industry towards the use of statistical methods. It might well be thought, therefore, that there is no need to continue the publication of a book whose form of presentation so clearly dates from a time when it was felt necessary to show in convincing terms that a case existed for the use of these techniques. The Technical Committee of the British Standards Institution concerned with statistical methods has, however, decided that in view of the continued substantial demand for this standard, it should remain in print, and I have been asked to write a short foreword to the present re-issue of B.S. 600.

As stated in the original Introduction, the main object of this publication was to set out as clearly as possible why statistical techniques are needed in the control of quality of industrial products and what kind of assistance they can give. Further, the main attention was concentrated on the use of these methods in problems of specification, standardization and routine production work. With the passing of time a great deal more experience has accumulated than was available in the early 1930's as to which are the most effective statistical tools and how to illustrate their use most convincingly. New and more pertinent examples of applications could no doubt be found; indeed to bring the standard up to date, very substantial rewriting would be desirable. This has not been attempted and the book is reproduced in its original form, except in the following matters:

(1) All the minor amendments and corrections, issued from time to time by the B.S.I., have been incorporated in the text.

(2) Table 14, giving factors for calculating control limits for range, has been altered so as to bring the 'levels' into conformity with those used in Table 13 dealing with means, standard deviations and coefficients of variation. These factors $D_{0.001}$, $D_{0.025}$, $D_{0.975}$ and $D_{0.999}$ were not available in 1935. Some resulting modification of the text where these factors are used has been necessary.

(3) References to text books and statistical tables have been brought up to date.

This means, in particular, alteration to a number of the Appendices. There is one further matter to which attention should perhaps be drawn. The squared standard deviation, s^2 , in a sample of n observations $x_1, x_2, \dots, x_n$ with mean $\bar{x}$ is defined (pages 29, 103, 105 and 107) by the relation

$$s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / n$$

and this definition has been used throughout in the numerical illustrations. The

B.S. 600 : 1935

same definition is used in B.S. 600R. On the other hand in the most recent B.S.I. publication on Statistics, (B.S. 2846: 1957, 'The reduction and presentation of experimental results', by J. T. Richardson), s^2 has been defined in accordance with the now more generally accepted practice as $\sum_i (x_i - \bar{x})^2 / (n - 1)$.

No error was involved in the earlier definition, since the fact that on the average, in sampling, s^2 as originally defined would be less than the population variance, σ^2 , was pointed out on pages 55 and 83*. Factors were given to correct this bias where necessary, and appropriate adjustments were made in the table of fiducial or confidence limits (Table 11) and in the table of control limits (Table 13) for both the standard deviation and the coefficient of variation.

While there would undoubtedly have been some advantage in bringing the definition of B.S. 600 into line with that of B.S. 2846, extensive recalculation of standard deviation values given in the examples, as well as of the factors in the tables, would have been necessary. It was therefore decided to leave things unaltered, at the same time warning the reader that this difference in definition must not be forgotten.

E. S. PEARSON.

July, 1960

NOTE. Many of the examples given by Professor Pearson are taken from British Standards which have since been revised or withdrawn. The standards referred to may no longer be obtainable, although they can be referred to in the library at British Standards House. However, it was not considered necessary to alter the actual examples, although the references are out of date.

* In the 1935 printing, the equivalent pages in the 1960 printing are 52 and 78.

THE APPLICATION OF STATISTICAL METHODS TO INDUSTRIAL STANDARDIZATION AND QUALITY CONTROL

INTRODUCTION

A characteristic feature of modern industry is repetition work, that is, the reproduction of the same thing again and again. Nevertheless, anyone who gives the matter at all close examination will realize that in spite of the finest limits in manufacture, giving generally excellent practical results, it is impossible to achieve complete sameness whether the sameness of the thing produced is measured by linear dimensions as in the diameter of a shaft or the form of a screw, by tensile strength as in cloth or steel wire or cement, by electrical resistance as in parts of telephone equipment, or by any other method. In all cases the recorded measures will not be found identical and the differences between them are not merely due to errors in measurement, but exist in the things themselves.

An average value may be recorded, but the existence of variation about that average has to be recognized. That such is the case has been at least partially recognized in industry, where in specifications it is generally laid down that tests should be made *on more than one* article. In the course of routine production, repeated tests of quality are carried out by manufacturers but on the whole there is little evidence to show that sufficient attention has been given in the past to the question whether the results of these tests are used to the best advantage, how far they are adequate to serve the purpose of the control for which they are intended, or to what extent some of them may even be redundant. It is the object of this publication to discuss how such problems may be dealt with on a sound practical basis.

Clearly it is impossible to forecast the precise value of the characteristics of any one of a number of articles; we cannot foretell what the length of life of an electric lamp or the breaking strength of a railway axle will be. But given that the general conditions surrounding the production are stable, it is possible to define with a high degree of probability limits within which, in the long run, any desired percentage of the values will fall. For example, it may be possible, if a certain method of production continues unchanged, to have confidence that at least 990 lamps out of every 1000, if burnt continuously at their marked voltage, would show a life of between a_1 and a_2 hours. The methods of statistical technique dealt with in this publication have been devised to assist in the solution of problems of this nature; how to determine whether there is stability in variation, how to relate the ordinary concepts of confidence and risk with more precise numerical measures of probability, and how to calculate definite limits of variation with which such measures of probability can be associated.

Whilst there is a considerable basis of mathematical theory underlying the construction of these statistical methods or tools, yet in many cases the methods and the reasons for their application can be made readily intelligible to the non-mathematical reader, so that it is seen that the term 'statistics' is only 'common sense' under another name.

The main object of this publication has therefore been to set out as clearly as possible why statistical technique is needed and what kind of assistance it can give.* It is convenient to distinguish between the use of statistical methods,

- (a) in problems of specification, standardization, and routine production work,
- (b) in special research problems.

In this publication, which deals more particularly with the simpler problems falling under heading (a), no comprehensive treatment of statistical theory has been attempted. The most effective use of statistical tools in problems of type (b) cannot be made without some fuller training in the subject, and references to text books and useful tables are given in the Appendices. But whether the reader of this publication intends to employ these tools himself or not, there is little doubt that if he will persevere in the attempt to understand the statistical method of approach, he will gain a breadth of outlook valuable to him in many ways.

The purpose of the methods of analysis dealt with may be effectively summarized in the words of a well known writer on the subject ('Student'), who has stated that the functions of statistical method in the control of quality are:

- (a) To find out how much variation in the essential qualities of the finished product affects its value, and in certain cases even to assist in determining what these qualities are.
- (b) To compare different methods of measuring the essential qualities so as to be able to estimate the value as cheaply and accurately as possible.
- (c) To trace variation in these qualities back to the variations in the raw material or manufacturing process from which they originate.
- (d) To adjudicate between rival proposals for reducing these latter variations, and
- (e) To indicate how to keep the quality of the finished product under continuous observation so as to give warning of any new source of variation or of an unexpected increase in an old one.

All or some of these uses of statistical method have been recognized by manufacturers in the past, and the discussion which follows is not concerned so much with new ways for keeping production under control as with the advantage of using the most modern methods to assist in achieving that purpose.

* An outline of the scheme of development will be found summarized under the contents heading; certain points of more theoretical nature are dealt with in the Appendices.

The introduction of a different method of approach involving concepts hitherto unfamiliar in this field, with the necessary accompaniment of a new terminology and a new notation, naturally will not be accepted without some questioning nor achieved without some effort. The pages which follow have been submitted to very detailed criticism and revision, in an endeavour to set out the ideas involved in simple form, without at the same time omitting the fundamental because it is difficult. In this process many have rendered assistance, and foremost the members of the British Standards Institution Committee. There is much still to be done before the industrialist's grammar of statistics can be written, but what is a first essential is the stimulation of interest and of discussion. Only then will it be possible to discover the extent of the field of application of these methods. But at least it is clear that it is a field which covers the development on scientific lines of methods which seek to eliminate waste of energy in production, and to supply goods soundly made and such as are best fitted to satisfy human wants.

SECTION 1

SCOPE OF THE PUBLICATION

The main object of this publication is to set out as clearly as possible why statistical technique is needed in the solution of many industrial problems, and the kind of assistance it can give. Some if not all of the methods of controlling quality which are described have been recognized by manufacturers in the past, and the publication is not concerned so much with new ways of keeping production constant, as with the advantage of making use of modern statistical tools in order to achieve this purpose in as simple and efficient a manner as possible. As an introduction to the subject three examples have been given to focus attention on certain of the wider questions at issue.

The example of malleable iron castings has been used to suggest how statistical methods may help the producer in detecting causes of trouble in his processes; they will provide powerful analytical tools and at the same time will assist in determining, on a sound basis, simple routine checks on the efficiency of the technical control.

The example of tests on cloth has been used to illustrate a form of problem which arises in sampling a consignment of material in order to secure conformity with specification. It has suggested the importance of establishing the relationship between size of sample and the precise rules to be laid down for acceptance or rejection, based on the resulting tests.

The example of the analysis of coal for ash-content as discussed in this publication shows the need of determining on a sound basis the extent of sampling necessary to estimate the quality of a consignment. It has also introduced the concept of fluctuation of quality with time.

To proceed further in the subject it has next been necessary to introduce certain statistical terminology; to define and explain certain statistical measures; and to illustrate methods of graphical representation and tabulation. The fundamental importance of the mean, as a measure of central tendency, and the standard deviation, as a measure of variation has been pointed out, and the characteristics of the Normal distribution curve have been outlined.

The essential problem of the relationship between a sample of units and the larger batch or consignment from which it has been drawn has next been discussed, and the conditions necessary if statistical theory is to be used in estimating the quality of the whole from that of the part, have been indicated. The important concepts of random sampling, representative sampling, statistical uniformity, and level of control have been developed.

A section has been devoted to discussing the application of these principles in practical sampling, and some illustrations from British Standard specifications have been quoted.

In approaching the problem of the preparation of such specifications it may be necessary to consider the whole range of quality in the total output of a

particular article or material produced by different manufacturers. The user will sometimes require that the variation should fall between certain narrowly defined limits; sometimes he can allow greater latitude provided a certain minimum level is reached; and sometimes he is not so much concerned with the precise average level, provided this level remains constant from one consignment to another. These requirements must be taken into account in framing the specification.

An essential part of a specification is the clause defining the methods by which conformity is secured. Two principal methods have been described,

(a) by testing units sampled from consignments,

(b) by requiring that records be kept by the producer which, supplemented by a system of check testing, will provide statistical evidence of the level of control maintained during manufacture.

The publication points out that both methods can form the basis of a system of marking to show conformity to a specification, and that the assistance of statistical theory is essential both in providing assurance to the user of the adequacy of the sampling, and in giving warning to the producer if unexpected variations in processes are affecting the quality of his product. The relation of these two methods to both user and producer is shown and it has been pointed out that in many cases there are clear advantages in the second, namely, that of requiring that definite evidence of control be furnished during the process of manufacture. These advantages have been summarized as follows:

(1) The difficulty which often arises of determining how to draw a representative sample from a consignment is avoided.

(2) The cost of sampling on the large scale often necessary to give adequate assurance to the user is saved.

(3) Even where it is necessary to carry out occasional tests on samples from consignments to gain assurance that the control in production remains effective, the amount of sampling necessary would be far less than that required to provide definite protection in the face of erratic control.

(4) The form of routine statistical analysis necessary to provide the basis of a system of marking would be the same as that which the producer would employ in attempting to increase the efficiency of his production processes.

In order to illustrate more fully the steps necessary to put either method (a) or (b) into operation it has next been necessary to describe in rather greater detail the elements of statistical sampling theory. The most important of these are connected with the relation between (1) the mean and the standard deviation calculated from a limited number of units contained in a sample, and (2) the values of the corresponding statistical measures which define the much larger batch or consignment from which the sample has been drawn. Definitions of the standard error of a mean and the standard error of a standard deviation have

been given, and the application of theory to measuring the reliability of the mean of a representative sample and duplicate samples has been described.

As a practical illustration of some of these results the example of testing cloth already mentioned has been dealt with at greater length.

A convenient method of expressing the accuracy of the information deduced from a sample consisting of relatively few observations has also been described, and tables of factors used in calculating the so-called 'confidence limits' are given. This method of expressing the reliability of a statistical estimate (such as a mean or standard deviation) has very definite advantages over the traditional method of expression in terms of 'probable error'.

Other problems which arise in the attempt to secure conformity with a specification by the sampling of consignments have then been discussed. Such a problem is that of devising a system of grading or classification in order to limit the range of variation in quality of certain characteristics of a manufactured article or material. Here the fundamental difficulty is that the classification must be based, not on 100 per cent inspection of the consignment, but on tests carried out on sampled units the number of which may be small owing to the expense of testing. As a result the testing clauses in the specification may be inefficient in the sense that owing to the inevitable waywardness of chance, certain samples from a consignment of quality below the desired level will be passed as satisfactory (and consequently the consignments will be passed), while others will fail to pass and the consignment be rejected, although its quality is in fact above the required level. The concept of the efficiency of a classification applied to random samples from consignments has been introduced, and methods by which this efficiency may be increased have been indicated.

The importance has been emphasized of carrying out preliminary investigations on statistical lines before the conditions for sampling and testing in a specification of this type are laid down. As a result it may often be found that when there is considerable variation among the individual units, it will not be an economic proposition for the user to test sufficient units to give him the desired degree of information about the quality of the consignment. In such a case it would be natural to consider whether conformity with specification would not be better assured by requiring that statistical evidence of the level of control be furnished by the producer under a guarantee or mark system.

The next problem has therefore been to consider the simpler methods of statistical analysis which may be applied by the producer during the course of manufacture to assess the level of his control of quality. These methods consist essentially in the collection and arrangement of test records in a manner planned to detect any lack of statistical uniformity which may be present, and it has been shown how the results may be usefully presented in the form of simple control charts. The theoretical basis of these charts has been described, the ideas to be borne in mind when interpreting them have been discussed, and tables to assist

in rapid calculation of the 'control limits' have been provided. Finally four illustrative examples have been given, namely:

(1) Ash-content in successive deliveries of coal. Two coals were compared, and it was possible to trace time fluctuations in the level of control, for the coal with the higher ash content.

(2) Quality of electric lamps. Control limits were based on an old standard. Afterwards changes in manufacture were introduced, and it was possible in a short time to detect the resulting improvement in quality from examination of the charts.

(3) Quality of a manufactured chemical. This example provided an illustration of the close interlocking of statistical analysis and technical skill. The former suggests that a cause of trouble is present; the latter is needed in examining the production processes to identify this cause; and finally the former must again be applied to show whether the trouble has been successfully removed.

(4) Breaking strength of a roofing material. In this case weekly changes in quality were due to the regular replacement of a machine part. An assignable cause of variation, which it was in fact possible to control, would here have been overlooked if a rational form of grouping the test records had not been adopted.

Finally a brief reference has been made to an important type of problem on which further investigation is needed. That is the specification of tolerance limits for a characteristic A as an index of a second more important characteristic B of which a direct performance test is difficult or costly to obtain. The value of using such indices of quality depends upon the degree of correlation between the characteristics A and B. The statistical concept of correlation has been touched on very briefly, the example of the use of a measure of hardness as an index to tensile strength being quoted.

SECTION 2

ILLUSTRATION OF GENERAL PROBLEMS BY EXAMPLES

The term 'statistics' is so commonly associated with an idea of lists of figures, whether relating to output, to costs, to sales, to prices, or to operatives' hours and wages, that it is advisable to make clear at the outset of this publication what in fact are these 'statistical methods' that may be used in problems of industrial standardization and quality control. It is important to give some preliminary answer to the questions, Why is this technique needed? In what does it consist? What kind of assistance can it give? For this purpose it has seemed best to deal first with the particular rather than the general, using the following three examples to focus attention on the wider questions at issue:

- (1) Tensile strength of malleable iron castings.
- (2) Breaking strength of cloth.
- (3) Ash-content of samples of coal.

1. *Example 1. Tensile strength of malleable iron castings*

The following clauses are from a specification* issued by the American Society for Testing Materials:

'(a) The tension test specimens specified in Section 5 shall conform to the following minimum requirements as to tensile properties:

Tensile strength, lb/sq. in. ...	...	...	...	...	50 000
Yield point, minimum, lb/sq. in. ...	...	...	...	...	32 500
Elongation on 2 in., per cent ...	...	...	...	...	10.0

'(b) The yield point may be determined by the drop of the beam of the testing machine or by the divider method.'

At a symposium on this subject, held under the auspices of the American Society for Testing Materials, some 20 000 test results were contributed from 17 different sources. The results as recorded provide only (a) the maximum, (b) the minimum, and (c) the average or mean test value, and these figures for tensile strength are reproduced in Table 1.

Before making any observations on the differences between the figures, an examination may be made of the detailed results from one source, No. 4; these are shown in Table 2.

It will be seen that 8 of the 75 test specimens have failed to pass the 50 000 lb/sq. in. limit, although the average value is well above this; that is because there is a large amount of variation about the average. In order to improve the quality of the malleable castings, of which these are sample pieces, in such a way that the test results can be depended upon always to satisfy the minimum

*A.S.T.M. Designation A. 47—30, Section 3.

**TABLE 1. TENSILE STRENGTH IN POUNDS PER SQUARE INCH OF
MALLEABLE IRON CASTINGS**

Source	Number of tests	Maximum	Minimum	Average
1	6000	59 000	45 000	54 000
2	4000	58 500	53 000	56 250
3	200	56 880	50 000	52 460
4	75	55 850	47 850	52 890
5	1210	62 140	54 400	57 920
6	300	62 860	52 150	56 350
7	50	56 000	50 000	53 000
8	—	58 000	50 000	55 000
9	3200	61 300	49 000	55 000
10	289	59 800	50 000	53 970
11	36	60 600	46 600	52 670
12	500	58 000	50 000	53 000
13	—	62 000	51 000	53 000
14	107	56 640	45 500	51 170
15	3300	61 500	45 000	53 710
16	540	58 000	50 500	55 500
17	59	56 160	50 480	52 830

**TABLE 2. 75 TEST RESULTS OF MALLEABLE IRON CASTINGS
CORRESPONDING TO SOURCE 4 RE-ARRANGED IN ORDER
FROM MINIMUM TO MAXIMUM**

47 850	51 600	52 950	53 600	54 400
48 550	51 750	52 950	53 700	54 450
48 850	51 850	53 000	53 750	54 550
49 300	51 900	53 000	53 750	54 650
49 550	51 900	53 150	53 800	54 750
49 700	52 100	53 200	53 800	54 800
49 850	52 250	53 200	53 800	54 900
49 950	52 250	53 200	53 850	54 950
50 150	52 400	53 250	53 900	55 150
50 450	52 500	53 300	53 950	55 300
50 650	52 550	53 300	54 050	55 350
50 850	52 700	53 350	54 150	55 400
51 110	52 700	53 400	54 150	55 450
51 450	52 750	53 450	54 350	55 600
51 550	52 900	53 600	54 400	55 850
Maximum 55 850.		Minimum 47 850.	Average 52 890.	

It will be assumed that readings were taken to the nearest 50 lb.

requirement of the specification, it will be necessary to make preliminary investigations into the causes to which the extreme variations may be assigned.

This investigation would require a consideration of such questions as the dependence of the quality of casting upon the choice of the materials, homogenization of the mixture, control of temperature, avoidance of accidental chilling, etc., and it would involve the division of the test records into a number of groups, within each of which all or some of these possible factors were constant. This grouping, which is essential in any process of analysis, has been described by W. A. Shewhart* as division into *rational sub-groups*. Clearly, at this first stage, sound judgment, coupled with a knowledge of the production process, is needed as distinct from a mere knowledge of statistical technique.

Suppose now for the sake of illustration that the 75 tests in the present example fall naturally into 5 sub-groups of 15, which it is thought might be differentiated owing to changes in one or other of the factors suggested above. Two assumed groupings are shown diagrammatically in Fig. 1 (page 127), where each spot represents a test value.

Case (a). There is here no obvious differentiation in the spots between the 5 sub-groups; strengths of less than 50 000 lb/sq. in. occur in each case. The analysis has not in fact provided any evidence that there was a source of trouble present in one group which did not also exist in the other four.

Case (b). Here the tests in sub-group (2) clearly differ from the rest. That is to say, the failure of 8 out of 75 tests to pass the minimum specification has been due to some cause of trouble which it has been possible to locate as operating only on the material from which the 15 tests of sub-group (2) were taken.

The difference between the situations illustrated in the two cases suggests considerations of some importance. If, whatever methods of sub-division based on knowledge of manufacturing processes are tried, the test values are always scattered more or less without differentiation as in case (a), the only method of raising the minimum level may be to raise the strength value of the whole output, which would mean an increase in the upper limit, as well as in the lower. For whilst it may be possible to increase the *average* strength it is impossible to reduce excessive *variation* without some clue to its main cause or causes. If, on the other hand, some assignable cause of variation can be located as the result of the analysis shown for case (b), it may be possible to remove the trouble without having to increase the strength throughout. In this way a higher average strength may be secured with a smaller variation, as shown for the tests belonging to sub-groups (1), (3), (4) and (5) of case (b).

The user's point of view may be also illustrated from these data. Thus, besides the attainment of a specified standard, it is often important to the user:

* Economic control of quality of manufactured products. Macmillan, 1931.

- (1) that the quality of the product received should remain uniform from one consignment to another;
- (2) that the amount of test sampling necessary should be reduced to the minimum consistent with safety.

On the assumption that the sub-groups in Fig. 1 (case (b)) represent samples tested from 5 consignments, the fact that the tensile strength of all 15 test-pieces belonging to the first sample fell above the 50 000 lb/sq. in. limit clearly gives no justification for confidence in the quality of material in a later consignment. Indeed, it is possible that the user would prefer case (a), in which a certain margin of strength is sacrificed for confidence in uniformity.

It may be questioned: What has statistical method to do with this? Why is not a mere inspection of diagrams such as these adequate? The answer is that in practice all kinds of intermediate situations between those suggested in cases (a) and (b) are possible; situations in which the interpretation is not clear from casual inspection. It is here that a more exact method, often expressed in the form of a simple routine working rule, can give valuable assistance in deciding whether apparent differences in the sub-groups are large enough to warrant the expense and time involved in a search for assignable causes of trouble.

A comparison of the results of the tests on material supplied by the different producers given in Table 1 can be made from an examination of Fig. 2, in which the horizontal lines represent the range of variation among the test specimens from each of the 17 sources given in Table 1. The average strengths are shown by the circles. A very considerable difference is evident; in one direction lie the results from sources Nos. 1 and 15, with a medium average strength, and a great range of variation between the extremes; in the other those from sources Nos. 2 and 5 showing a high average and a small variation. Each of these 4 cases is based on more than 1000 test specimens.

Certain questions are at once suggested. Is it economic that material of such great variety should be sold under the same specification? Does it in fact cost more to produce the castings from source 5 than those from source 1? Supposing that such variety is not desirable, what should be the best standard of quality to aim at, having regard to the needs of the user and the difficulties to be overcome by the producer? Could the specification be modified so as to encourage the attainment of this standard?

In addition to these points there is one more closely connected with statistical theory. It is this: the mere statement of minimum and maximum test strengths does not really provide an adequate measure of variation. It is possible, for example, that a *single* faulty specimen among the 6000 pieces tested from source 1 may have pulled down the minimum strength from over 50 000 to 45 000 lb/sq. in. If a single measure of variation is to be used, it should be one dependent on the whole distribution of individual tests, and not only upon the two extreme

values. Statistical theory provides such a measure in the 'standard deviation', which will be defined in a later section.

2. Example 2. Breaking strength of cloth

In specifications for cloth a minimum strength is commonly required. Now, however uniformly cloth is manufactured, it cannot be supposed that the strength, as tested by breaking, will be the same from one length of cloth to another, or for different strips cut from the same length. Consequently to obtain any adequate appreciation of the quality of a certain type of material, the breaking strength of a number of test pieces must be found. Further it is not only the average strength which is of importance, but also the uniformity as measured by the variation about that average from one area of the cloth to another.

It follows that in fixing any specification which is to be based on tests on sample pieces, it is essential to take into account the relationship between average values, minimum values, range of variation, etc. and further to pay regard to the manner in which these quantities are dependent upon the number of pieces tested. It is possible to illustrate the nature of the problem on the data given in Table 3 (p.23). The figures represent the breaking strength in pounds of 128 strips of cotton fabric each 3 yd long and 2 in. wide*. They have been grouped for the purposes of illustration into 32 samples each containing 4 strips.

The results are presented in diagrammatic form in Fig. 3, where the straight lines represent, to scale, the range of variation in each sample†, and the circles the corresponding average or mean strength for that sample. From (a) the 32 samples of 4 strips, (b) 16 samples of 8 and (c) 8 samples of 16 have been formed. For instance, the first sample of 4 tests gave breaking strengths of 177, 176, 183 and 180 lb: the vertical line under Sample No. 1 in the Fig. 3 (a) therefore runs from 176 to 183, and the circle is centred at the mean, namely $\frac{1}{4}(177 + 176 + 183 + 180) = 179.0$ lb. The first sample of 8 in Fig. 3 (b) is made up of samples Nos. 1 and 2; it ranges from 164 to 188 lb with a mean at 179.25 lb. The first sample of 16 in Fig. 3 (c) is made up of samples Nos. 1-4; it ranges from 162-188 lb and has a mean at 176.2 lb.

Attention is drawn to the following points, among others, brought out by an inspection of the diagram:

(1) The variation in mean strength is considerable, but it becomes smaller the greater the number of tests upon which the mean is based, for example,

for samples of 4 pieces, the means fall between 158.5 and 180.5 lb,

for samples of 8 pieces, the means fall between 162.9 and 179.3 lb,

for samples of 16 pieces, the means fall between 165.0 and 176.9 lb.

The limits of variation of the mean are shown by dotted lines.

* A. J. Turner. 'Random and systematic selections of Warp Specimens in Cloth-Sampling.' *Journal of the Textile Institute*, XXII, 1930, T.77. Actually the strips were taken from the same length of cloth, but this does not affect the present argument.

† For the statistician's definition of a 'sample', see page 28.

TABLE 3. BREAKING STRENGTH, IN POUNDS, OF COTTON FABRIC STRIPS

Sam- ple No.	Strength	Sam- ple No.	Strength	Sam- ple No.	Strength	Sam- ple No.	Strength	Sam- ple No.	Strength	Sam- ple No.	Strength	Sam- ple No.	Strength	Sam- ple No.	Strength	Sam- ple No.	Strength
1	177	5	158	9	166	13	156	17	167	21	176	25	180	29	174		174
	176		175		174		166		175		172		155		168		168
	183		157		166		164		171		186		168		173		173
	180		174		173		177		166		182		168		166		166
2	182	6	165	10	164	14	183	18	166	22	184	26	178	30	174		174
	188		167		181		178		167		172		180		172		172
	164		169		170		181		166		182		178		160		160
	184		172		158		170		178		178		186		182		182
3	180	7	172	11	163	15	142	19	157	23	182	27	177	31	186		186
	168		152		154		158		167		174		173		172		172
	178		172		166		171		161		174		162		166		166
	175		154		174		163		174		166		175		172		172
4	176	8	163	12	178	16	176	20	182	24	184	28	172	32	165		165
	162		175		178		177		172		170		182		164		164
	178		161		173		156		160		166		171		162		162
	168		154		158		167		165		182		178		150		150

(2) In general the range of variation in strength within a sample—the length of the vertical lines—increases with the number of pieces in the sample. These ranges have the following average values:

- (a) for samples of 4 pieces, 16.4 lb;
- (b) for samples of 8 pieces, 23.1 lb;
- (c) for samples of 16 pieces, 28.6 lb.

(3) Suppose that a minimum required strength limit were to be fixed by the specification at 164 lb.

If this minimum criterion were applied to the *mean* strength in a sample, we would find that:

- | | |
|------------------------------|-------------------------------|
| (a) 4 out of 32 samples of 4 | } fail to meet the criterion. |
| (b) 2 out of 16 samples of 8 | |
| (c) 0 out of 8 samples of 16 | |

On the other hand, if it were applied to the *weakest* strip in a sample, we would find that:

- | | |
|-------------------------------|-------------------------------|
| (a) 16 out of 32 samples of 4 | } fail to meet the criterion. |
| (b) 12 out of 16 samples of 8 | |
| (c) 7 out of 8 samples of 16 | |

Without placing undue stress on figures which are based on a single sampling experiment, the following conclusions can be established by statistical theory and turned to practical account:

(1) The larger the number of observations or tests, the greater the range of variation to be expected among them, since exceptional individual cases are more likely to be met.

(2) The larger the number of observations or tests, the smaller the variation between the mean of one set of tests and that of another, since the stability of an average increases with the number of items on which it is based.

A statement by way of specification that 'the strength must be greater than 164 lb' would therefore be inadequate unless supplemented by information as to the number of pieces to be tested, and to whether the limit referred to the mean or to the minimum strength among the tests. Without this it would be impossible to know how to proceed to ascertain whether the material did in fact conform to the requirements.

3. Example 3. Ash-content in samples of coal

In a Report issued by the British Standards Institution on the 'Sampling of small fuel up to 3 in.*', figures were given of the percentage ash-content found from the chemical analysis of samples of coal, each drawn from a different

* B.S.: 403 1930. Report by Dr. E. S. Grumell and Dr. A. C. Dunningham. (This Report is no longer available.) The first 100 of the tests used below are those given for Fuel F in Table 3, page 17 of that Report.

wagon. It has been possible to supplement these data and to make use of the results of some 550 successive tests made as an ordinary routine practice by the same firm of consumers on the 'rough slack' F, during the course of about 10 years. For the present purpose, the results of the 550 tests have been combined into groups of 10 consecutive tests, and Table 4 (page 26) shows:

- (a) The 55 mean percentage ash-content values for these groups.
- (b) The 'progressive' means, *i.e.*, the means of all tests up to a given one.*

The sampling is, of course, carried out in order that the user may obtain an appreciation of the quality of the bulk of the fuel he is purchasing. Consignments of coal may be either single deliveries such as for the export trade or they may be deliveries under a contract, extending over a period of months. In the latter case the question will arise of possible changes in the quality of the coal during the period. In the former case the problem of sampling resolves itself into the question of determining how many wagons should be sampled in order to obtain, with any required degree of accuracy, an appreciation of the quality of the whole consignment.

It is not necessary to enter here into details of the conclusions reached regarding the number and weight of samples to be drawn, but to illustrate certain points of importance by examining the results of the first 250 tests (in 25 groups of 10) given in Table 4. These are shown diagrammatically in the left hand side of Fig. 4; above, (a), are the means of the 25 groups of 10, and below, (b), the progressive means.

There is considerable variation about the mean of 16.987 per cent for the whole 250 tests. Suppose 17 per cent is taken to represent the true measure of the ash-content in the whole consignment sampled. A practical question would then be to ascertain how many tests must be made before it would be possible to estimate this measure within ± 1 per cent of its value, (*i.e.*, within the limits 17.17 and 16.83).

The lower half of the diagram shows that, as the number of samples is increased from 10, 20, 30 to 250, the progressive mean approaches the limiting value of 17. The shaded area represents the belt lying between 17.17 and 16.83. From the 14th group onwards the fluctuation in the progressive mean has steadied down to fall within the belt. That is to say, in this particular case, it appears that 140 wagons would have to be sampled before a stable estimate lying within ± 1 per cent of 17 per cent could have been made of the true ash-content. This, however, must not be taken as a general rule; much will depend upon the homogeneity of the fuel, the weight of the sample, the method of sampling, etc., and it is in the determination of the relationships between these quantities that the methods of statistical analysis are called for.

* For example, the means of the first three groups are 15.62, 18.45 and 15.90 respectively; the corresponding progressive means are 15.62, $\frac{1}{2}(15.62 + 18.45) = 17.04$ and $\frac{1}{3}(15.62 + 18.45 + 15.90) = 16.66$.

TABLE 4. ASH-CONTENT DATA (PER CENT): MEANS IN GROUPS OF 10 TESTS

Group No	Mean	Progressive mean	Group No	Mean	Progressive mean	Group No	Mean	Progressive mean
1	15.62	15.62	20	17.90	17.03	39	12.80	16.71
2	18.45	17.04	21	16.35	17.00	40	13.57	16.63
3	15.90	16.66	22	17.18	17.01	41	14.31	16.57
4	16.63	16.65	23	15.91	16.96	42	15.40	16.54
5	17.56	16.83	24	18.07	17.00	43	14.34	16.49
6	16.55	16.78	25	16.57	16.99	44	14.11	16.44
7	19.71	17.20	26	21.74	17.17	45	13.55	16.37
8	17.48	17.24	27	16.22	17.13	46	13.23	16.30
9	16.54	17.16	28	16.79	17.12	47	15.23	16.28
10	14.81	16.92	29	15.26	17.06	48	17.02	16.30
11	17.04	16.93	30	15.09	16.99	49	13.27	16.24
12	15.22	16.79	31	16.22	16.97	50	12.01	16.15
13	16.82	16.79	32	17.93	17.00	51	12.59	16.08
14	17.65	16.86	33	16.06	16.97	52	11.66	16.00
15	16.95	16.86	34	18.32	17.01	53	11.86	15.92
16	17.87	16.92	35	14.99	16.95	54	11.35	15.83
17	17.33	16.95	36	14.72	16.89	55	9.53	15.72
18	18.09	17.01	37	15.87	16.86			
19	16.46	16.98	38	14.87	16.81			

Again what is meant by the 'true measure of the ash-content of the whole consignment'? If the coal were purchased in a single delivery then it might mean the percentage of ash in the whole consignment, even if this fluctuated from one part of the consignment to another. But if the order were extending over several months or even years, a new factor would have to be taken into account. The right hand side of Fig. 4 shows the results of tests made later on 100 further wagons containing the same fuel (arranged in the diagram in 10 groups of 10 tests). It is to be noted that:

- (a) The means of the individual groups all fall well below the 17 per cent line;
- (b) the progressive means fall to an increasing extent outside the shaded belt.

The conclusion drawn above that the fluctuations in the progressive means were steadying down to a stable value of 17 per cent has been falsified. The quality of the coal, in fact, began to change, owing to altered methods of mining, at about the time when the 200th wagon was tested, and the ash-content was gradually reduced. This was a satisfactory change, but an unsatisfactory one might also have occurred. This is evidence of an obvious but often disregarded truth: no prediction of the reliability of an estimate is of any value unless the conditions governing the production at the time when the estimate was formed remain essentially constant for the future period.

No analysis, however careful, will enable a change in conditions to be detected before it has begun to occur; but precise methods have often provided data from which the experienced investigator recognizes the first signs of a change some time before this could be discovered by cruder methods, such as the mere inspection of columns of figures. This, for example, might have been important had the quality of the Fuel F deteriorated instead of improved, and had the change in ash-content been an index of this deterioration.

SECTION 3

INTRODUCTION TO THE SIMPLER STATISTICAL TOOLS

4. *Definitions**

The illustrations in the preceding section give a general idea of the function of statistical methods in the analysis and control of variation. Before developing in greater detail the application of these methods to problems of specification and standardization, it will be necessary to describe rather more fully some of the simpler tools of the statistician.

Suppose that a single characteristic has been measured and recorded for each of a number of individual articles or units. The units may be similar parts of some instrument, strands of wire, pieces of cloth, wagons of coal, sections of timber, etc.; the characteristics may be tensile strength, absorptive power, chemical composition, specific gravity, etc. The measured values of a characteristic will be termed observed values or observations.

If a group of units has been selected from a larger whole, it is defined in statistical terminology as a sample; it is also common to speak of the observations themselves as forming the sample. Thus a sample may consist of 1, 2, 3 . . . n units or observations. A sample which is drawn without bias will be termed a random sample; the larger whole or population† of units from which the sample has been taken may be the output of a particular machine, the day's output of a factory, the consignment which a user is purchasing, etc. In the sections that follow it will be necessary to discuss various aspects of the relationship between the sample and the population, but it is first necessary to introduce certain statistical measures which from the descriptive view-point may be equally applied, whether the group of units under consideration forms the sample or the population.

There are various methods in use for measuring the central tendency and the variation within a group of observations. In general the arithmetic mean has been most used as a measure of central tendency, and variation can be expressed in terms of the departures or 'deviations' from that mean. It is important that a descriptive measure should not be at the mercy of a single erratic observation; for this reason the arithmetic mean is to be preferred to the midpoint between the maximum and minimum observations as a measure of central tendency. In the same way a measure of variation based on all the deviations, (such as the mean deviation and standard deviation referred to below) will be preferable to the difference between the maximum and minimum value (defined as the range). In certain problems which are discussed in a later section the range is a useful measure, but in general for accurate work the more stable measures dependent upon all the individual observations are necessary, and the arithmetic mean and standard deviation are used. The calculation of the latter may at first

* These and other definitions are collected in Appendix I, page 103.

† See page 45 for a discussion of this term and Appendix I for definition.

appear difficult, but if the methods of procedure described in the appendices are studied, it will be realised that this disadvantage is not of a serious nature.

Arithmetic Mean (generally called *Mean*), is the sum of the observations divided by their number.

Median, is the value of a variable characteristic which is greater than half of the observations and less than the other half. There will be a difference in definition according as the number of observations is odd or even. For example, if 5 observations are arranged in order of magnitude, the median value is that of the 3rd observation. If, however, there are 6 observations, the median may be defined as the value mid-way between the 3rd and 4th. The median is sometimes used as a central measure in place of the mean, and is in a sense dependent on the whole of the observations.

Mid-point is the value midway between the extreme (maximum and minimum) observations.

Range is the difference between the maximum and the minimum observations (i.e. it represents the distance between extreme observations).

Deviation (in the special sense used here) is the difference between an observation and the mean of all the observations.

Mean Deviation is the mean of the deviations when all are given a positive sign.

Standard Deviation is the square root of the mean of the squares of the deviations of all the observations.

To illustrate these terms* the following figures will be used, representing the percentage ash-content found from 10 test analyses of coal, each test based on the contents of a different wagon†:

16.70, 16.20, 18.31, 14.46, 12.01, 17.58, 14.87, 12.61, 15.52, 17.93.

the *mean* of the above values = $\frac{1}{10} (16.70 + 16.20 + \dots + 17.93) = 15.619$

the *median* of the above values = $\frac{1}{2} (15.52 + 16.20) = 15.86$, since the 5th and 6th values in order of magnitude are 15.52 and 16.20

the *mid-point* of the above values = $\frac{1}{2} (12.01 + 18.31) = 15.16$

the *range* of the above values is $18.31 - 12.01 = 6.30$

the *deviations* of the above values are respectively + 1.08, + 0.58, + 2.69, - 1.16, - 3.61, + 1.96, - 0.75, - 3.01, - 0.10, + 2.31.

The sum of these will be zero.

the *mean deviation* = $\frac{1}{10} (1.08 + 0.58 + 2.69 + 1.16 + \dots + 0.10 + 2.31)$
= $\frac{1}{10} (17.25) = 1.725$.

the *standard deviation* = $\sqrt{\left[\frac{1}{10} \{ (1.08)^2 + (0.58)^2 + \dots + (2.31)^2 \} \right]}$
= $\sqrt{4.19269} = 2.048$.

* The measures are expressed in algebraic form in Appendix II.

† These data are the tests forming the first group in Table 4.

5. *The presentation of data*

Dot Diagram. When only a few observations are available, these may be conveniently listed individually as was done in Tables 2 and 3, and in such cases 'dot diagrams' of the type shown in Fig. 1 will often give a useful preliminary picture of the situation. Indeed for certain purposes the consideration of such a diagram may be all that is needed.

Frequency Distribution. When dealing with a larger number of observations, however, it is generally found convenient to group the data. The total observed range of variation in the characteristic measured is then usually divided into convenient equal intervals, perhaps 12–25 in all, and the number of observations falling into each interval counted up. This number is termed the frequency for that interval and the resulting tabulated series of numbers shows the frequency distribution.

Three examples of such distributions are shown in Table 5.

Example A. Percentage ash-content in 250 test analyses of coal; (for review of data see Example 3, pages 24–27). The results were given to the nearest 0/100th of 1 per cent, and they have been divided into 33 groups each of 0.50 per cent width. The number of groups is rather more than is generally necessary. The table shows for example that out of the 250 observations, 14 were recorded as having values 14.50, 14.51, ... or 14.99 per cent. If the measurements were taken with sufficient accuracy, this group should include all analyses for which the ash-content lay between 14.495 and 14.995 per cent. What may be termed the 'central value' for this group is therefore 14.745 per cent, and for simplicity it is often this quantity that is entered in the table instead of the upper and lower limits of the group (see examples B and C). In presenting data, it is particularly important that the table should make clear what the group limits are; for this purpose the use of central values, if correctly made, is always sufficient.

Example B. Tensile strength of malleable iron castings. These are the 75 values reviewed in Example 1 (see pages 18–22), and given individually in Table 2. In this case central values have been used for classification; for example, in 6 tests the tensile strength lay between 51 500 and 52 000 lb/sq. in.*

Example C. Depth of sapwood in telephone poles. These data were collected in the Bell Telephone Laboratories in connection with an investigation into the effect of creosote and other treatments. Central values have again been used in tabulation, the group breadth being 0.3 in., and the groups ranging as follows:

0.85 — 1.15, 1.15 — 1.45, ... 5.35 — 5.65.

A grouped frequency distribution may be represented diagrammatically in a number of ways. Probably the most satisfactory of these is that shown in Fig. 5, where the three frequency distributions of Table 5 have been reproduced. Here

* In cases where a value recorded lies at the boundary between two groups, it is sometimes the custom to enter a $\frac{1}{2}$ frequency in each group. But the group intervals can usually be chosen to avoid this difficulty.

TABLE 5. FREQUENCY DISTRIBUTIONS.

Example A Samples of coal		Example B Malleable iron castings		Example C Telephone poles	
Percentage ash-content	Frequency	Tensile strength (central values)	Frequency	Depth of sapwood (central values)	Frequency
		lb/sq. in.		in.	
9.00—9.49	1	47 750	1	1.0	2
9.50—9.99	—	48 250	—	1.3	29
10.00—10.49	2	48 750	2	1.6	62
10.50—10.99	1	49 250	1	1.9	106
11.00—11.49	1	49 750	4	2.2	153
11.50—11.99	2	50 250	2	2.5	186
12.00—12.49	5	50 750	2	2.8	193
12.50—12.99	4	51 250	2	3.1	188
13.00—13.49	7	51 750	6	3.4	151
13.50—13.99	6	52 250	4	3.7	123
14.00—14.49	13	52 750	8	4.0	82
14.50—14.99	14	53 250	12	4.3	48
15.00—15.49	15	53 750	11	4.6	27
15.50—15.99	13	54 250	7	4.9	14
16.00—16.49	24	54 750	6	5.2	5
16.50—16.99	15	55 250	5	5.5	1
17.00—17.49	19	55 750	2		
17.50—17.99	23				
18.00—18.49	22				
18.50—18.99	12				
19.00—19.49	12				
19.50—19.99	7				
20.00—20.49	6				
20.50—20.99	8				
21.00—21.49	6				
21.50—21.99	4				
22.00—22.49	2				
22.50—22.99	2				
23.00—23.49	—				
23.50—23.99	3				
24.00—24.49	—				
24.50—24.99	—				
25.00—25.49	1				
Total	250	Total	75	Total	1 370

the horizontal scale corresponds to the variable characteristic, and the frequency of observations in a given interval is represented by a rectangle, of height proportional to this frequency, standing on the appropriate base element. For example for the case C, 153 observations are shown in Table 5 with a depth of sapwood between 2.05 and 2.35 in.; these are represented in the diagram by the shaded rectangle of height 153, and breadth equal to 0.30 in. on the horizontal scale.

Using this method, *frequency* in the table corresponds to *area* in the diagram. Such diagrams are termed *histograms*.

6. Use of the mean and the standard deviation in scaling a frequency distribution. The normal curve

It is now possible to illustrate a simple but useful property of the mean and the standard deviation. The following results are calculated from the distributions given in Table 5.*

	<i>Mean.</i>	<i>Standard deviation.</i>
Example A	16.993 per cent	2.660 per cent
Example B	52.877 lb/sq. in.	1.779 lb/sq. in.
Example C	29.14 in.	0.798 in.

Suppose that on either side of the mean, distances of $\pm \sigma$, $\pm 2\sigma$ and $\pm 3\sigma$ are plotted off, the common practice of using σ to denote the standard deviation being followed. The result is shown in Fig. 5.† Certain points are evident from inspection.

(1) In each case almost the whole of the observations lie within the limits, $\text{Mean} \pm 3\sigma$.

(2) Distribution A appears approximately symmetrical; in B there are no observations above $\text{mean} + 2\sigma$, and there is a long drawn-out 'tail' in the negative direction; distribution C appears to be slightly asymmetrical, but in the opposite direction.

The situation may be appreciated more exactly by calculating in each case the percentage of the observations lying beyond each of the six limits. An approximate method of calculation gives the results shown in Table 6.‡

* It will be noticed that the means for A and B do not exactly correspond with those quoted on page 25 for A (16.987) and page 19 for B (52.890). The difference arises from the fact that the present figures are based on grouped data.

† For example for A.

$\text{Mean} + \sigma = 16.993 + 2.660 = 19.653$, $\text{Mean} - \sigma = 16.993 - 2.660 = 14.333$.

$\text{Mean} + 2\sigma = 16.993 + 5.320 = 22.313$, etc.

‡ The contents of frequency groups crossed by a limiting line have been divided in proportional parts.

TABLE 6. PERCENTAGE OF OBSERVATIONS FALLING OUTSIDE
LIMITS DETERMINED BY THE STANDARD DEVIATION

Limits from mean at	$+\sigma$	$-\sigma$	$+2\sigma$	-2σ	$+3\sigma$	-3σ
Example A	14.7	15.1	2.7	2.3	0.4	0.0
	29.8		5.0		0.4	
Example B	14.8	16.5	0.0	4.8	0.0	0.1
	31.3		4.8		0.1	
Example C	17.0	17.0	3.0	1.3	0.1	0.0
	34.0		4.3		0.1	

The lack of symmetry in case B is noticeable, but what is perhaps more remarkable is the agreement between the percentages, when the figures for positive and negative deviations are combined. We see that in each case less than a half of 1 per cent of the observations lie outside the range (mean -3σ) to (mean $+3\sigma$); that about 4 per cent to 5 per cent lie outside the range (mean -2σ) to (mean $+2\sigma$); and about $\frac{1}{2}$ of the total lie outside (mean $-\sigma$) to (mean $+\sigma$).

Figures not differing very greatly from these will be found again and again when handling frequency distributions based upon homogeneous material. This result is clearly of importance because it shows that provided there is no very great lack of symmetry, these two measures alone, the mean and standard deviation, are likely to provide a useful indication of the limits within which variation may be expected to occur. They lead in fact to a crude scaling adequate for many practical purposes.

It will be noticed in Fig. 5 that the distribution for example C (based on 1370 observations), is smoother than those for either A (250 observations) or B (75 observations). This is a typical result. Experience in fact shows that when dealing with material where the variation is uniform or stable*, the greater the number of observations the more regular the figures become, so that ultimately as the numbers increase the frequency distribution will be represented by a smooth curve. This concept is illustrated diagrammatically in Fig. 6.

* This characteristic of stability will be discussed fully later.

Starting with a sample of 10, the observations have been gradually increased to 20, 50, 100 and 500. For the first two cases the frequencies have been shown only in dot form; afterwards histograms are drawn, each having the same total area. The gradual smoothing of the distributions as the sample size increases is evident. At the bottom of the diagram the limiting form is shown as a smooth symmetrical curve which has a mean of 50.00 and standard deviation of 3.33 units.

In this illustration the curve shown is that known as the *Normal Curve* (sometimes termed the *Gaussian*)*. It has a definite mathematical equation which depends only on the values of mean and standard deviation, and it has been found to represent approximately a great number of the symmetrical frequency distributions met with in practical experience. For this curve the percentage of observations lying outside certain limiting deviations from the mean are as follows:

Outside range, mean $- \sigma$ to mean $+ \sigma$, 31.73 per cent.

Outside range, mean $- 2 \sigma$ to mean $+ 2 \sigma$, 4.55 per cent.

Outside range, mean $- 3 \sigma$ to mean $+ 3 \sigma$, 0.27 per cent.

It will be seen from Table 6 that these theoretical figures show close agreement with those obtained on combining the two extremes of the three observational distributions. But owing to the lack of symmetry evident in Fig. 5, the distribution of example B, taken as a whole, clearly could not be represented by this particular curve.

Statistical methods and tables are available,

- (a) To test whether the assumption is justified that the variation in a certain measured characteristic may be represented by the Normal curve,
- (b) To calculate on the basis of this assumption, and from the mean and standard deviation, the expected frequency of individual units that will be found within any desired range of the scale of measurement.

Taking for example the ash-content distribution from Table 5, investigation as in (a) above shows that there is no reason to question the assumption that the variation in this characteristic from wagon to wagon may be represented by a Normal curve. Consequently if we take the mean and standard deviation quoted on page 32,

mean = 16.993 per cent; standard deviation = 2.660 per cent

the appropriate tables enable us to estimate that the ash-content in

15 test samples out of 250 (6.0 per cent) should lie outside the limits 12 per cent to 22 per cent,

65 test samples out of 250 (26.0 per cent) should lie outside the limits 14 per cent to 20 per cent.

It will be found from Table 5 that the observed numbers are actually 15 and 61.

* See Appendices IV, V and VI.

Such close agreement between prediction and actual experience cannot of course always be expected. The example however suggests the important descriptive part that this curve may play, always provided that its suitability to represent the type of variation in question has been first established.

It would seem well to take this opportunity to discourage a common belief. There is no magic about the Normal curve in the sense that if a distribution follows this law, that is proof that the variation of the product is stable and under control. Nor is it only possible to use statistical methods when variation is Normal. It should, however, be noted:

- (a) When present, this characteristic has undoubtedly great practical convenience, since it makes possible the use of the published tables, both those referred to on page 34 and others such as Tables 11, 13 and 14 discussed later.
- (b) As mentioned in pages 51 and 80 the distribution of means, $\bar{x}$, of samples of n observations, will be very closely Normal even when n is as small as 4 or 5, in cases where the distribution of the individual observations is not Normal.

SECTION 4

SAMPLING AND STATISTICAL THEORY

7. *Statistical uniformity and random sampling; levels of control*

When 100 per cent inspection is not practicable, which is in general the case, the characteristics of a batch or consignment of material must be derived by both producer and user from the inspection of a limited number of units, or what has been defined in statistical terms as a sample. Having admitted that there is variation between similar units, it becomes essential to take this into account when considering the relation of the sample to the batch from which it is drawn.

The product may be a component part of an article, an assembly of component parts making up a finished article, or a bulk of material in solid, liquid or gaseous form, but in all cases where sampling is necessary, the vital question of determining the extent to which the batch that has been sampled conforms to the specification can only be answered satisfactorily with the aid of statistical treatment.

This relation of the sample to the batch is the centre of the problem. It is impossible in discussing it to avoid the introduction of certain ideas which may appear difficult because they are unfamiliar, but it is hoped that the following paragraphs if studied in conjunction with the diagrams in Fig. 7 will convey to the reader the essential features of the statistician's method of approach.

The figure may be supposed to represent the results of tests made at a factory on similar articles or component parts withdrawn five at a time, at regular intervals during the course of production. For example the data might relate to measurements on the diameter of steel bolts taken from the lathe, or to life tests of electric lamps taken from the flow of finished lamps coming from the making machine. The result of each set of five tests forming a sample is shown in the Figure in the dot form which has been used above (Figs. 1 and 6), the scale of measurement being supposed horizontal. Six different cases numbered I, II, . . . VI are presented, and in each case 12 consecutive sample results are shown. Beneath, curves have been drawn which may be supposed to represent the distribution of the measured characteristics that would be found were it possible to test the whole of the batch that has been sampled.

Consider Case I; there is of course considerable difference between the dot patterns of the 12 samples, yet a certain stability or uniformity in the variation is evident which is clearly absent in Case V. In Case I, the differences are in fact no more than would be met if a large number of the bolts had been mixed together in a box and successive samples of five drawn out in a haphazard manner and tested. Such samples will then be random samples.

If the variation satisfies this condition of stability it becomes possible,

- (a) when a characteristic has been measured only on a sample, to form from these measurements an estimate of the distribution curve of the

characteristic in the batch from which the sample has been drawn at random;

- (b) when the distribution curve is known from past experience, to predict the nature of the variation to be expected from one random sample to another.

It will be seen that problems (a) and (b) are the inverse of each other.

In forming the estimate described in (a), greater reliability will be secured the larger the number of units selected. Thus a better appreciation could be formed of the mean and the standard deviation of the distribution curve at the bottom of Column I from the tests of four samples of 5 than from a single sample of 5. This is common sense reasoning, but as soon as an attempt is made to form, from the test records, an estimate of the distribution curve for the batch, in terms, let us say, of hours of lamp life, the assistance of statistical theory is necessary. For example it is this theory which provides the information contained in the Table of confidence limits (sub-section (14), page 61).

To illustrate problem (b), the following example may be taken. Suppose that it is known on the basis of extensive past experience, that the distribution curve of length of life (at marked voltage) of a certain type of lamp follows the Normal law, with a mean of 1000 hours and standard deviation of 300 hours. Then statistical theory enables such statements as the following to be made.*

- (i) The chance that the mean length of life in a random sample of 20 lamps will fall below 800 hours is 0.0014, or such a result may be expected only once in about 700 samples.

- (ii) The chance that *at least one lamp* in this sample of 20 will have a life of less than 800 hours is 0.997, or this result may be expected in all but 3 samples in every 1000.

That these two chances would be entirely different is an expected result; statistical theory however is able to give precision to the expected.

When this characteristic of stability (represented pictorially by Fig. 7, Case I) is present, the variation will be described as *statistically uniform*.† It is then possible to make use of the methods of statistical theory described in the following sections for such purposes of inference and prediction as were referred to under headings (a) and (b) above.

Although it is not easy to give any precise non-mathematical definition of statistical uniformity, the concept is not difficult to grasp. It has been already illustrated by contrast in Fig. 1 (a) and (b) (discussed on page 20); some further comparisons will now be drawn with the help of Fig. 7.

* See Appendices IV and V for reference to appropriate Tables.

† After much consideration this term has been adopted in the place of W. A. Shewhart's expression '*statistically controlled*'. This decision was made largely in order to avoid the confusion which has been found to occur when using the words 'statistical control' between the concept of statistical uniformity and the pragmatic interpretation of the phrase as the act, for example, of a process manager. For Shewhart's definition his book should be studied; see for example pages 6 and 145-149.

While the variation in Case I is statistically uniform that for Case V is definitely not so. There can hardly be uniformity in a production process which leads without assignable cause both to samples No. 4 and No. 11. The dot pattern suggests indeed that there may be factors at work leading to two centres of variation, of which sometimes one, sometimes the other, and sometimes both are in operation. This lack of homogeneity is suggested in the distribution curve for the whole batch shown at the bottom of the column.

Another situation is presented by Case III; here it appears that samples No. 3, 7 and 11 are differentiated from the others. But there is a systematic repetition in the irregularity which was not evident in Case V, and it may be supposed that it is possible to assign the cause for these differences. If this were so, statistical methods could still be usefully employed, for it would be possible to divide the material into homogeneous sub-portions within each of which there was statistical uniformity. Thus properly representative samples could be obtained, and the form of variation in the complete batch of material could still be predicted. For example if samples 1, 5, 9, . . . were from the output of one machine, samples 2, 6, 10, . . . from a second, 3, 7, 11, . . . from a third and 4, 8, 12, . . . from a fourth, and machine 3 differed somewhat from the others, the samples of 20 shown in Column IV could be described as representative of the total output.

As an illustration of a situation of this type, take the case of the manufacture of sheet brass, the thickness of which at the edges is less than at the centre owing to the nature of the rolling process. If then a sheet is cut into narrower widths, the thickness will vary according to the position from which the strip has been cut. If this variation is recognized, the product may be divided accordingly into parts which will be homogeneous for sampling purposes. But if it is not recognized it is possible that test pieces will sometimes all be taken from the edges, sometimes from the centre and sometimes from both; the situation will then be as in Case V and no reliable inference can be drawn from the sample measurements.

Case VI represents another type of situation. Here there is a systematic change with time taking place in the quality of the material produced. If it is possible to assign causes for these fluctuations—as for example known changes in temperature or humidity (perhaps in some textile process), or periodic fluctuation due to the introduction of fresh raw material into the production routine at regular intervals; for example, fresh supplies of raw cotton arriving weekly—then again it may be possible to divide the data into groups within which the variation may be statistically uniform. But if there are irregular fluctuations in time without known cause, prediction of the characteristics of the batch from one or two samples will be impossible. Thus samples Nos. 1 and 2, or again Nos. 8 and 9 would not be representative of the batch distribution shown at the foot of the column.

The distinction typified by the difference between Cases I and III is an important one. In the former case any one of the samples may be used to give information regarding the batch or consignment of material; in the latter, care must be exercised in choosing representative samples. In the one case, as has been suggested, the distribution among the samples is that which would be found among units drawn at random from a well mixed source of supply; in the other it is as though the units were drawn from a number of sources of different constitution, the proportion taken from each source being carefully specified though the drawing from within a source was at random. The second case may be described as that of representative sampling.

It should be clear that if the total output of a particular article is made up from a number of sources, the variation associated with each source being statistically uniform and the proportion of the total coming from each being known, then statistical methods can be used to estimate the quality of the total from properly drawn samples. If on the other hand there is not sufficient information to enable properly representative samples to be drawn, the variation must be described as statistically indeterminate. This indeterminacy may be due either to changes in space or changes in quality with time, e.g. from one machine or factory to another, or in the product of one factory due to seasonal influences or changes in raw material.

It is desirable to mention at this point that variation which is statistically uniform is not *necessarily* represented by the Normal distribution curve. It is true that in a very large number of cases met with in industrial experience this has been found to be so; it is also the case that the Tables and methods described below are only valid provided the variation is approximately of the Normal form. But examples can be found (as that of breaking strength in cotton fibres given in Appendix 6) where the distribution curve of a measured characteristic is definitely not symmetrical, and yet the concept of statistical uniformity is still appropriate. This point is discussed more fully in that Appendix.

There is one further concept of importance that must be introduced before the significance of these ideas to user and producer is dealt with. The variation in Case II is statistically uniform as well as in Case I, and for both the complete batch variation has been represented by Normal curves. But the variation among the individual units is greater in II than in I. It has been pointed out that a Normal curve is completely defined by its mean ($\bar{X}$) and its standard deviation (σ); it follows that the difference between I and II lies in the fact that σ_1 (Case I) is less than σ_2 (Case II). Thus, if I and II represent the variation in length of life of similar lamps made by two different producers, it might be supposed that $\sigma_1 = 300$ hours and $\sigma_2 = 400$ hours. At the same time the mean length of life might differ, say $\bar{X}_1 = 1800$ hours and $\bar{X}_2 = 1600$ hours (although this is not shown in the diagram since no scale is given). It would then be said that the levels at which production was controlled were different. In both cases statistical

theory might be used to predict the proportion of lamps to be expected to have a life within any given limits, but clearly as far as this lamp characteristic is concerned the first producer would be turning out a more homogeneous and more durable article than the second.

This distinction between the concepts of *statistical uniformity* and the *level of control* is important. Statistical uniformity is a term used to describe the characteristics of variation which appear stable, but the stability is relative to the sampling technique employed. For example, in a material which is being continuously produced, sampling at short intervals might detect lack of statistical uniformity which sampling at longer intervals would pass over.

Evidence of stability is not, however, enough; information regarding mean value and range of variation is required. As in the case of the lamps just quoted, this is provided by the mean and standard deviation, and it is a knowledge of these measures that is necessary in assessing the level of control. Statistical uniformity may be associated with a low as well as a high level of control. Where however lack of uniformity can be detected by a statistical analysis of test records, the technician may be enabled to lay his hands on defects in the production process, with the result that the level of control is raised. Methods of analysis that may be used for this purpose are described in the section on control charts, page 74.

8. *Sampling methods in practice*

It is important to consider the bearing of the principles discussed in the preceding section on problems of practical sampling. How far are samples, in fact, generally drawn in such a way that statistical theory can be profitably employed in this connection? To what extent and in what manner is it possible to improve existing sampling procedures? No single answer can be given to these questions, since the methods of sampling which are practicable must vary enormously from one type of material to another, but certain illustrations may be used with profit to point out some of the difficulties inherent in the problem and to show the manner in which these can be overcome.

In the first place consider the case where the material sampled consists of a number of similar units, either component parts or finished articles. In certain instances it will be relatively easy to secure a random sample from a single well mixed source of supply; such for example is true in sampling small engineering parts such as ball bearings, bolts, screws, etc. A simple method of securing such a sample which is open to the producer who has confidence that his level of control is remaining constant, is to arrange that every 50th, 100th or 500th unit (as the need may be) which comes from the making machines, is set aside for inspection purposes. There is of course one danger in this procedure which should not be overlooked; if there is a periodic fluctuation in quality owing perhaps to wear and subsequent renewal of parts of a machine,* to temperatures

* For an example of this, see sub-section (25), page 95.

changing with the time of day, or to reduction of attention on the part of the operative, it is essential that this should not be in step with the time interval between withdrawal of units for test purposes. Were this to be the case the sample might be definitely biased and conclusions drawn from it misleading.

In cases where inspection for defect is carried out on a large scale by the producer, and random samples are easily secured, routine methods have been developed on a statistical basis which aim at combining any desired degree of assurance of detecting serious defect with a minimum of total inspection. An interesting description of such methods introduced in the Bell Telephone System has been given by Dodge and Romig, (Bell System Technical Journal VIII, 1929, pages 613-631).

Far more often, however, the problem is not so simple; this is particularly the case when sampling must be carried out to determine the quality of a consignment before acceptance by the user. Since there is generally no evidence available that the producer's level of control remains constant, it becomes important to plan a sampling procedure which will provide a reliable estimate of the quality of a consignment, even if the variation within that consignment is not statistically uniform. Care must, in fact, be taken to draw a representative sample.

A good illustration of this problem may be found in the sampling clauses of the British Standard Specification for Tungsten Filament Electric Lamps (B.S. 161 : 1934). The following brief extracts from pages 8, 9 and 13 of the specification will explain the position:

The term *Batch* denotes all the lamps of one class ordered at one time, or, at the option of the purchaser, all the lamps of one class put forward in one lot for acceptance; or all the lamps of one class which are subjected upon one occasion to test for compliance with this specification.

The 'Inspection Quantity' consists of sample lamps selected for the purpose of determining the acceptability of a batch of lamps as to mechanical and physical requirements.

12 (b). Alternatively (i.e., when all the lamps of a batch are not tested), an 'Inspection Quantity' comprising not less than 5 per cent of the batch (with minimum of 25 lamps) shall be taken for test in approximately equal numbers from selected containers.

At least every second container in a batch of 1000 lamps, or every third container in a batch of more than 1000, shall be selected for this purpose, with a minimum of 10 containers; or all the containers if the batch occupies less than 10. The containers shall be selected evenly, so that no undue proportion is taken from any one portion of the batch.

Should 5 per cent plus one lamp* of the 'Inspection Quantity' fail to pass the requirements of any one of Clauses 2-8, or, alternatively, should the sum of the failures under Clauses 2-8 exceed 20 per cent of the number of lamps* tested the whole batch of lamps represented by the 'Inspection Quantity' shall be deemed not to comply with the specification.

It will be seen that these directions for sampling clearly envisage the possibility of lack of statistical uniformity in the variation of lamp quality, or else it would have been adequate to sample a single container. It is recognized

* NOTE. In the case of a fraction resulting from this calculation, the next higher whole number shall be taken.