

First edition
2009-04-15

Guidance on the application of statistical methods to quality and to industrial standardization

*Lignes directrices pour l'application des méthodes statistiques à la
qualité et à la normalisation industrielle*



Reference number
ISO/TR 18532:2009(E)

AVA Documents
Document Sharing Hub
© ISO 2009

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.



COPYRIGHT PROTECTED DOCUMENT

© ISO 2009

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword	ix
Introduction.....	x
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Illustration of value and role of statistical method through examples	1
4.1 Statistical method.....	1
4.2 Example 1: Strength of wire	2
4.2.1 General	2
4.2.2 Overall test results and lower specification limit.....	2
4.2.3 Initial analysis	3
4.2.4 Preliminary investigation	3
4.2.5 General discussion on findings	6
4.2.6 Explanation of statistical terms and tools used in this example.....	6
4.3 Example 2: Mass of fabric	7
4.3.1 General	7
4.3.2 Test results and specification limits	7
4.3.3 Discussion of specific results.....	10
4.3.4 Discussion on general findings	11
4.4 Example 3: Mass fraction of ash (in %) in a cargo of coal	11
4.4.1 General	11
4.4.2 Test results (reference ISO 11648-1: <i>Statistical aspects of sampling from bulk materials</i>).....	12
4.4.3 Initial graphical analysis of specific results	12
4.4.4 Benefits of a statistically sound sampling plan	14
4.4.5 General conclusions	16
5 Introduction to basic statistical tools.....	16
5.1 General	16
5.2 Basic statistical terms and measures	16
5.3 Presentation of data	19
5.3.1 Dot or line plot	19
5.3.2 Tally chart.....	19
5.3.3 Stem and leaf plot.....	19
5.3.4 Box plot	20
5.3.5 Multi-vari chart.....	22
5.3.6 Position-Dimension (P-D) diagram	23
5.3.7 Graphical portrayal of frequency distributions	25
5.3.8 The normal distribution	31
5.3.9 The Weibull distribution.....	35
5.3.10 Graphs	41
5.3.11 Scatter diagram and regression.....	41
5.3.12 Pareto (or Lorenz) diagram.....	43
5.3.13 Cause and effect diagram.....	44
6 Variation and sampling considerations	45
6.1 Statistical control and process capability	45
6.1.1 Statistical control	45
6.1.2 Erratic variation	47
6.1.3 Systematic variation.....	47
6.1.4 Systematic changes with time	48
6.1.5 Statistical indeterminacy	49

6.1.6	Non-normal variation	49
6.1.7	Quality level and process capability	49
6.2	Sampling considerations	50
7	Methods of conformity assessment	54
7.1	The statistical concept of a population	54
7.2	The basis of securing conformity to specification	55
7.2.1	The two principal methods	55
7.2.2	Considerations of importance to the customer	56
7.2.3	Considerations of importance to the supplier	56
8	The statistical relationship between sample and population	57
8.1	The variation of the mean and the standard deviation in samples	57
8.1.1	General	57
8.1.2	Variation of means	58
8.1.3	Variation of standard deviations	60
8.2	The reliability of a mean estimated from stratified and duplicate sampling	64
8.2.1	Stratified sampling	64
8.2.2	Duplicate sampling	66
8.3	Illustration of the use of the mean mass, and the lowest mass, in a sample of prescribed size of specimens of fabric	67
8.4	Tests and confidence intervals for means and standard deviations	69
8.4.1	Confidence intervals for means and standard deviations	69
8.4.2	Tests for means and standard deviations	71
8.4.3	Equivalence of methods of testing hypotheses	77
8.5	Simultaneous variation in the sample mean and in the sample standard deviation	77
8.6	Tests and confidence intervals for proportions	80
8.6.1	Attributes	80
8.6.2	Estimating a proportion	80
8.6.3	Confidence intervals for a proportion	81
8.6.4	Comparison of a proportion with a given value	82
8.6.5	Comparison of two proportions	82
8.6.6	Sample size determination	83
8.7	Prediction intervals	84
8.7.1	One-sided prediction interval for the next m observations	84
8.7.2	Two-sided prediction interval for the next m observations	85
8.7.3	One and two-sided prediction intervals for the mean of the next m observations	85
8.8	Statistical tolerance intervals	86
8.8.1	Statistical tolerance intervals for normal populations	86
8.8.2	Statistical tolerance intervals for populations of an unknown distributional type	87
8.8.3	Tables for statistical tolerance intervals	87
8.9	Estimation and confidence intervals for the Weibull distribution	87
8.9.1	The Weibull distribution	87
8.10	Distribution-free methods: estimation and confidence intervals for a median	88
9	Acceptance sampling	89
9.1	Methodology	89
9.2	Rationale	90
9.3	Some terminology of acceptance sampling	91
9.3.1	Acceptance quality limit (AQL)	91
9.3.2	Limiting quality (LQ)	91
9.3.3	Classical versus economic methods	92
9.3.4	Inspection levels	92
9.3.5	Inspection severity and switching rules	92
9.3.6	Use of “non-accepted” versus “rejected”	93
9.4	Acceptance sampling by attributes	93
9.4.1	General	93
9.4.2	Single sampling	94
9.4.3	Double sampling	96
9.4.4	Multiple sampling	96
9.4.5	Sequential sampling	99

9.4.6	Continuous sampling	100
9.4.7	Skip-lot sampling	101
9.4.8	Audit sampling	102
9.4.9	Sampling for parts per million	102
9.4.10	Isolated lots	103
9.4.11	Accept-zero plans	103
9.5	Acceptance sampling by variables — Single quality characteristic	104
9.5.1	General	104
9.5.2	Single sampling plans by variables for known process standard deviation — The “ σ ” method	105
9.5.3	Single sampling plans by variables for unknown process standard deviation — The “ s ” method	106
9.5.4	Double sampling plans by variables	109
9.5.5	Sequential sampling plans by variables for known process standard deviation	110
9.5.6	Accept-zero plans by variables	110
9.6	Multiple quality characteristics	111
9.6.1	Classification of quality characteristics	111
9.6.2	Unifying theme	111
9.6.3	Inspection by attributes for nonconforming items	111
9.6.4	Inspection by attributes for nonconformities	112
9.6.5	Independent variables	113
9.6.6	Dependent variables	113
9.6.7	Attributes and variables	113
10	Statistical process control (SPC)	113
10.1	Process focus	113
10.2	Essence of SPC	116
10.3	Statistical process control or statistical product control?	117
10.4	Over-control, under-control and control of processes	118
10.4.1	General	118
10.4.2	Scenario 1: Operator reacts to each individual sample giving rise to process over-control	119
10.4.3	Scenario 2: Operator monitors a process using a run chart giving rise to haphazard control	120
10.4.4	Scenario 3: Monitoring using SPC chart with a potential for effective control	121
10.5	Key statistical steps in establishing a standard performance-based control chart	122
10.5.1	General	122
10.5.2	Monitoring strategy	122
10.5.3	Construction of a standard control chart	125
10.6	Interpretation of standard Shewhart-type control charts	127
10.7	Selection of an appropriate control chart for a particular use	128
10.7.1	Overview	128
10.7.2	Shewhart-type control charts	129
10.7.3	Cumulative sum (cusum) charts	129
11	Process capability	130
11.1	Overview	130
11.2	Process performance versus process capability	131
11.3	Process capability for measured (i.e. variables) data	132
11.3.1	General	132
11.3.2	Estimation of process capability (normally distributed data)	132
11.3.3	Estimation of process capability (non-normally distributed data)	133
11.4	Process capability indices	138
11.4.1	General	138
11.4.2	The C_p index	138
11.4.3	The C_{pk} family of indices	139
11.4.4	The C_{pm} index	142
11.5	Process capability for attribute data	145
12	Statistical experimentation and standards	148
12.1	Basic concepts	148
12.1.1	What is involved in experimentation?	148

12.1.2	Why experiment?	148
12.1.3	Where does statistics come in?	149
12.1.4	What types of standard experimental designs are there and how does one make a choice of which to use?	149
13	Measuring systems	164
13.1	Measurements and standards	164
13.2	Measurements, result quality and statistics	165
13.3	Examples of statistical methods to ensure quality of measured data	166
13.3.1	Example 1: Resolution	166
13.3.2	Example 2: Bias and precision	169
13.3.3	Precision — Repeatability	171
13.3.4	Precision — Reproducibility	172
Annex A (informative)	Measured data control charts: Formulae and constants	177
Bibliography		181
Index		188

Figure 1	— Dot plot of breaking strength of 64 test specimens	2
Figure 2	— Basic cause and effect diagram for variation in wire strength (due to possible changes of material and process parameters within specified tolerances)	3
Figure 3	— Line plots showing patterns of results after division into rational groups	4
Figure 4	— Diagram indicating the effect of the interrelationship between oil quench temperature and steel temperature on wire strength	5
Figure 5	— Means of masses plotted against sample number (illustrating decreasing variation in the mean with the sample size increase)	9
Figure 6	— Ranges of masses within each sample vs sample number [illustrating increasing (range) variation within a sample with sample size increase]	9
Figure 7	— Averages of mass fraction of ash (in %) of coal by lot from cargo	13
Figure 8	— Progressive averages of mass fraction of ash (in %) in terms of lot	13
Figure 9	— Schematic diagram showing plan for sampling percentage ash from cargo of ship	14
Figure 10	— Mass fraction of ash (in %) plotted against test number for lots 19 and 20 (illustrating relative consistency of percentage ash within each of these lots)	15
Figure 11	— Mass fraction of ash (in %) plotted against test number for lots 9 and 10 (illustrating rogue pairs in both lots)	15
Figure 12	— Line plot of breaking strength of wire (Table 1 data)	19
Figure 13	— Typical tally charts	19
Figure 14	— Stem and leaf plot for data	20
Figure 15	— Box plot	21
Figure 16	— Box plot for Delta E panel shade variation between supply sources	21
Figure 17	— Multi-vari chart as a tool for process variation analysis	23
Figure 18	— Measurements on cylinder to determine nominal size, ovality and taper	23
Figure 19	— Measurement on cylinder — P-D diagrams showing ideal diameter values, pure taper and pure ovality	24
Figure 20	— Measurement on cylinder — P-D diagrams indicating progressive decrease of mean and increase in geometric form variation and the beneficial effects of overhaul	25
Figure 21	— Frequency histogram for immersion times in Table 6	27
Figure 22	— Percentage frequency histogram for immersion times in Table 6	27
Figure 23	— Cumulative percentage frequency histogram for immersion times in Table 6	28
Figure 24	— Cumulative percentage frequency diagram for immersion times in Table 6	29
Figure 25	— Normal curve overlaid on the immersion time histogram (mean = 6,79; standard deviation = 1,08)	30
Figure 26	— Straight line plot on normal probability paper indicating normality of data in Table 6	31
Figure 27	— Percentages of normal distribution in relation to distances from the mean in terms of standard deviations	32
Figure 28	— Standard normal probability density with indications of percentage expected beyond a value, U or L , that is z standard deviation units from the mean	33
Figure 29	— Comparison with Weibull distributions, all with $\alpha = 1$	37

Figure 30 — Q-Q plot to assess the fit of days between accidents (data in column one of Table 8) to a Weibull distribution	39
Figure 31 — Weibull probability plot of days between accidents (data in column one of Table 8)	40
Figure 32 — Scatter diagrams of four data sets that all have the same correlation coefficients (r) and fitted regression lines	43
Figure 33 — Relative contribution of different types of in-process paint faults	44
Figure 34 — Process cause and effect diagram for cracks in a casting	45
Figure 35 — Diagram indicating types of variation in samples	47
Figure 36 — Contrast of the capabilities of two filling machines	50
Figure 37 — Illustration of one-sided test	73
Figure 38 — Scatter chart for sample means and standard deviations in canned tomatoes data	78
Figure 39 — Standardized control chart for mean and standard deviation	79
Figure 40 — Type A and B OC curves for $n = 32$, $A_c = 2$, $N = 100$	94
Figure 41 — Type B OC curves for $A_c = 0, 1/3, 1/2$ and 1	95
Figure 42 — OC curves for single, double and multiple sampling size code letter L and AQL 2,5 %	97
Figure 43 — Average sample size (ASSI) curves for single, double and multiple sampling plans for sample size code letter L and AQL 2,5 %	98
Figure 44 — Curves for the double and multiple sampling plans for sample size code letter L and AQL 2,5 % showing the probability of needing to inspect significantly more sample items than under single sampling	99
Figure 45 — Example of sequential sampling by attributes for percent nonconforming	100
Figure 46 — Acceptance chart for a lower specification limit	106
Figure 47 — Acceptance charts for double specification limits with separate control	107
Figure 48 — Standardized acceptance chart for sample size 18 for double specification limits with combined control at an AQL of 4 % under normal inspection	107
Figure 49 — Standardized acceptance chart for sample size 18 for double specification limits with combined control at an AQL of 1 % for the upper limit and an AQL of 4 % overall under normal inspection	108
Figure 50 — ISO 9001:2008 Model of a process-based quality management system	114
Figure 51 — Control chart for nonconforming underwear	117
Figure 52 — Outline of process of applying a topcoat to a photographic film	118
Figure 53 — Probability of setter/operator observing a single mass value when mean = 45	119
Figure 54 — Example of process run chart with variation, but with no guidance on how to interpret and deal with variation	121
Figure 55 — Example of process control chart with criteria for “out-of-control” signals	122
Figure 56 — A two factor nested design is the basis of an $\bar{X}R$ chart (illustrated with a subgroup size of 3)	123
Figure 57 — Effect of subgroup size on ability to detect changes in process mean (process nominal = 5,00, process standard deviation = 0,01)	124
Figure 58 — Mean and range chart for masses of standard specimens of fabric	126
Figure 59 — Graphical comparison of process capability with specified tolerance	133
Figure 60 — Illustration of the estimation of capability with a skew distribution (equivalent to a range of $\pm 3\sigma$ in a normal distribution)	134
Figure 61 — Dot plot for percent of silicon data showing overall pattern of variation	135
Figure 62 — Probability plot for percent of silicon data showing overall pattern of variation	136
Figure 63 — Probability plot for the logarithm of percent of silicon data showing overall pattern of variation	137
Figure 64 — Individuals control chart of ln percent of silicon with limits	137
Figure 65 — Relationship between C_p and C_{pkU} and C_{pkL} for two sets of process variability and locations of specification limits	141
Figure 66 — Comparison of conformance to toleranced specification with optimal value approach ..	144
Figure 67 — Printed circuit board faults SPC chart and cumulative faults per unit (FPU) chart	147
Figure 68 — Effect of lubrication, speed, surface finish and density on push-off strength	154
Figure 69 — Interaction between squeegee speed and ink viscosity	155
Figure 70 — Central composite design of the face-centred cube variety for 3 factors	156
Figure 71 — Computer-generated contour plot for oxide uniformity in terms of power and pulse of the etching process for gas ratio fixed at its coded level 0	158
Figure 72 — Illustration of the fundamental difference in designs for two independent factors as compared with a two-component mixture	159

Figure 73 — Illustration of the fundamental difference in designs for three independent factors as compared with a three-component mixture.....	160
Figure 74 — Ten-point augmented simplex centroid three-component design.....	160
Figure 75 — Response surface contours for mean burn rate in terms of fuel (X1), oxidize (X2) and binder (X3) blend components.....	161
Figure 76 — Response surface contours for standard deviation of burn rate in terms of fuel (X1), oxidize (X2) and binder (X3) blend components	161
Figure 77 — Factors A and B set at nominal to give a process yield of 68 %.....	162
Figure 78 — First stage optimisation using Box EVOP	163
Figure 79 — First stage Box EVOP as local optimum has been found.....	163
Figure 80 — Incomplete 5 stage simplex maximization experiment for two factors in terms of yield .	164
Figure 81 — Recommended resolution for process control and determination of compliance with specified tolerance	167
Figure 82 — Range charts showing adequate and inadequate resolutions.....	168
Figure 83 — Bias and precision	169
Figure 84 — Effect of measuring systems uncertainty on compliance decision.....	170
Figure 85 — Establishing bias and precision for a pressure gauge	171
Figure 86 — Individuals and moving range chart for pressure to check for stability of results prior to performing a bias and precision test.....	172

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

In exceptional circumstances, when a technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide by a simple majority vote of its participating members to publish a Technical Report. A Technical Report is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO/TR 18532 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*.

Introduction

This Technical Report demonstrates the advantages in the application of statistical methods in as simple and efficient a manner as possible so that they become accessible to the many rather than to the few.

As an introduction to the subject, three examples are given in Clause 4 to focus attention on some of the wider questions at issue. These examples suggest how statistical thinking coupled with the use of simple statistical tools and technical and operational knowledge of the process can help in improving designs, process efficiency and performance and product conformity to specification.

- Example 1, relating to the strength of wire, illustrates the role and value of division of data into so-called rational subgroups coupled with the use of cause and effect diagrams and line plots. It also shows how to exploit interrelationships between process parameters to achieve robust designs. The need is emphasized to treat numerical data not just as a set of figures but as potentially meaningful information on a process. It demonstrates clearly that an enquiring mind and sound judgement, coupled with an understanding of the actual process producing the numerical data, are required as distinct from a mere knowledge of statistical method. This indicates the need for non-statisticians to become more aware of the role of statistical method and to become more involved in their actual application to secure the maximum possible benefits to any organization.
- Example 2, on fabric mass, illustrates key aspects that need to be considered when sampling to establish conformance of an entity to specification. In this example, general conclusions are established by statistical theory and are turned to practical use.
- Example 3 concerns the mass fraction of ash (in %) in coal. Specifically, it demonstrates four principal concepts: how to handle apparent fluctuation of quality within a quantity of material; the need to determine, on a sound basis, the amount of sampling necessary to estimate the quality of a commodity; the necessity to establish, in advance, a well designed sampling procedure; and the value of progressive analysis of results, in a simple graphical manner, as they become available.

More generally, example 3 illustrates the importance of the application of statistical thinking and design method to a numerical study prior to it being undertaken. It also indicates that, to gain full benefit from such a study, persons familiar with the activity under scrutiny should be involved throughout.

Clause 5 introduces basic statistical terms and measures, and a wide range of simpler statistical tools used to present and analyse data. Emphasis has been placed on a pictorial approach that can most readily be communicated to, and understood by, the many.

Clause 6 describes the fundamentals of sampling on a statistical basis and distinguishes between statistical uniformity (stability of a process) and quality level (process capability). Clause 7 introduces sampling with reference to a product requirement. It draws out the two principal methods, viz. that of after the event acceptance sampling and that of the ongoing control of inherently capable processes. Clause 8 provides a detailed treatment of the statistical relationship between sample and batch. Clause 9 describes the methodology, terminology and rationale of acceptance sampling. Single, double, multiple, sequential, continuous, skip-lot, audit, parts per million, isolated lot and accept-zero plans for acceptance sampling by attributes are dealt with. Acceptance sampling by variables covers the following plans for individual quality characteristics: single sampling plans for known and for unknown standard deviation; double sampling plans; sequential sampling plans for known standard deviation and accept-zero plans. Multiple-quality characteristic plans are also described.

Clause 10 covers the fundamentals of statistical process control. It distinguishes between statistical process control and the use of statistical process control techniques for statistical product control. Over-control, under-control and control are discussed. The key steps in establishing and interpreting performance-based control charts that are intended primarily to differentiate between special and common causes of variation and

provide a basis for capability and performance assessment are covered. The principal types of Shewhart-type control charts and the role and application of cumulative sum (CUSUM) charts are dealt with.

Clause 11 deals with performance benchmarking of stable processes under the heading of process capability assessment. Three very pertinent business questions are answered by a control chart: 1) is the process in control?; 2) what is the performance of the process?; and 3) is there evidence of significant improvement in process performance? Clause 11 focuses on answering the second question regarding process capability/performance of both measured data and attribute processes. It introduces the use of the internationally standardized capability indices, C_p , C_{pkL} and C_{pkU} . It also discusses the business implications, in terms of aiming at preferred value and minimizing variation, with the quotation of minimum C_{pm} values, rather than the convention of tolerating maximum use of specified tolerances in determining whether or not an entity conforms to requirements.

Clause 12 begins by illustrating the role and value of simple economic experimental designs where the mathematical content is such that all the necessary calculations can readily be done manually. It then continues to exploit the development of computer software programs in the design and analysis of experiments. Nowadays the need for computational skills has become so minimal that the practitioner can concentrate his attention on choosing the right kind of design for a particular application, how to perform the experiment and how to interpret the computer outputs. In both cases, pictorial outputs are encouraged to facilitate understanding.

Clause 13 deals with the capability of measuring systems. Following a resumé of the basic statistical requisites of a measuring system that ensures the integrity of the data output, examples are given of the application of statistical method to the evaluation of resolution, bias and precision, uncertainty, repeatability and reproducibility.

Guidance on the application of statistical methods to quality and to industrial standardization

1 Scope

This Technical Report describes a broad range of statistical methods applicable to the management, control and improvement of processes.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ISO 3534-2, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO 3534-3, *Statistics — Vocabulary and symbols — Part 3: Design of experiments*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO 3534-2 and ISO 3534-3 apply.

4 Illustration of value and role of statistical method through examples

4.1 Statistical method

The term “statistics” is commonly associated with an idea of lists of numbers, whether relating to output, costs, sales, prices or wages. It is thus advisable to make clear at the outset what in fact this statistical method is that may gainfully be applied in the field of quality and standardization. It is important to give some preliminary answers to certain questions. Why is statistical method needed at all? What does it consist of?

What kind of assistance can it give? Where can, when can, and should, it be applied? For this purpose, it has seemed best to deal first with the particular rather than the general, using specific examples to focus attention on the wider issues involved.

4.2 Example 1: Strength of wire

4.2.1 General

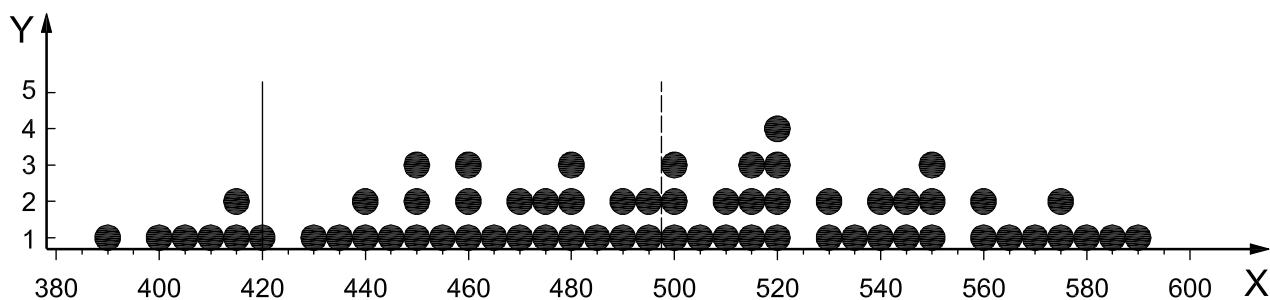
This example illustrates the role and value of the division of data into so-called *rational subgroups* coupled with the use of *cause and effect diagrams* and *dot plots*. It shows their applicability to both problem solving and process and product enhancement. It also indicates the need to treat numerical data, not just as a set of figures, but as potentially meaningful information on a process. It demonstrates clearly that an enquiring mind and sound judgement, coupled with an *understanding of the actual process* producing the numerical data, are required, as distinct from a mere knowledge of statistical method. Hence, there is a need for technologists, technicians, and operational, administrative, marketing and management personnel to become more aware of the role of statistical method and become more involved in their actual application to secure the many benefits possible to any organization.

4.2.2 Overall test results and lower specification limit

Suppose 64 test results were obtained on the breaking strength of wire where the lower specification limit is set as 420 units. The results are shown in ascending order reading down the columns in Table 1 and as a dot plot in Figure 1.

Table 1 — 64 test results of wire breaking strength arranged in order from minimum to maximum (measurements were made to the nearest 5 units)

390	415	435	450	460	470	480	490	500	510	515	520	540	550	560	575
400	415	440	450	460	470	480	490	500	510	520	530	540	550	565	580
405	420	440	450	460	475	480	495	500	515	520	530	545	550	570	585
410	430	445	455	465	475	485	495	505	515	520	535	545	560	575	590
Mean = 495; median = 497,5; minimum = 390; maximum = 590.															



Key

X strength

Y number of observations

The dashed line at 497,5 represents the sample median and the full line at 420 represents the lower specification limit.

Figure 1 — Dot plot of breaking strength of 64 test specimens

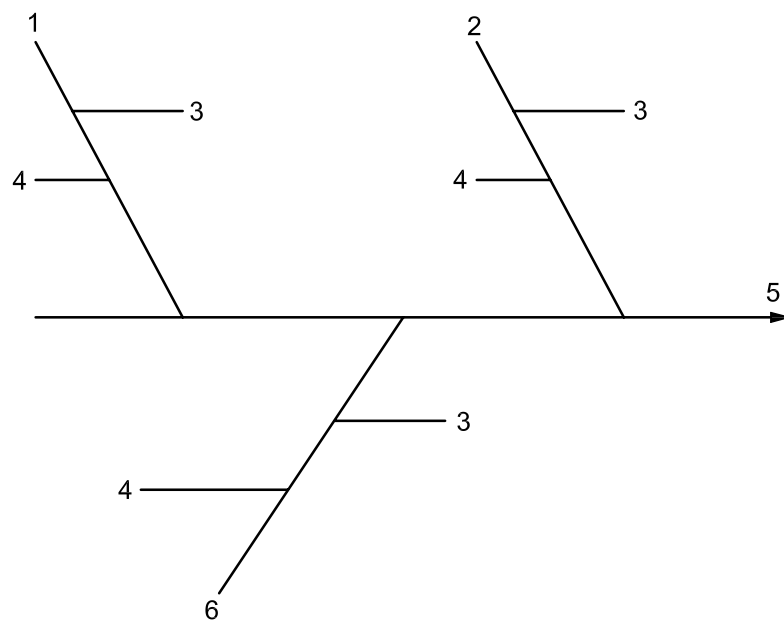
4.2.3 Initial analysis

It can be seen that 6 of the 64 test specimens have failed to achieve the 420 lower limit, although the mean is well above this at nearly 500; that is because there is a large amount of variation about the average. This is best indicated graphically in the form of the *dot plot* of Figure 1. (For the corresponding *line plot*, see Figure 12.)

It is obviously necessary to improve the quality of the wire, of which these are sample pieces, if the breaking strength is to be depended upon to always satisfy the minimum requirements of the specification. The pattern of variation is fairly symmetrical with a relatively large scatter. Whilst it may be possible to increase the mean strength, it is impossible to reduce excessive variation without some clue as to its main causes. If, on the other hand, some assignable (special) cause of variation can be located, it may be possible to take specific action both to increase the mean strength and reduce the overall variability. This will call for preliminary investigations into the causes to which the extreme variations may be assigned.

4.2.4 Preliminary investigation

This investigation would first require a consideration of such questions as the possible causes of variation in the wire strength. The outcome from a multi-disciplined team was the simple *cause and effect diagram* (see 5.3.13) as shown in Figure 2, which suggests a dependence of the wire strength on material composition and levels of steel and oil quench temperatures.

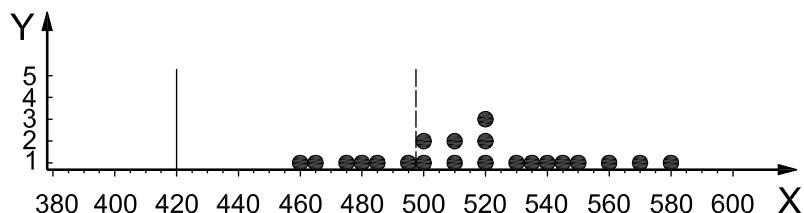


Key

- | | | | |
|---|------------------------|---|----------------|
| 1 | steel temperature | 4 | low |
| 2 | oil quench temperature | 5 | strength |
| 3 | high | 6 | carbon content |

**Figure 2 — Basic cause and effect diagram for variation in wire strength
(due to possible changes of material and process parameters within specified tolerances)**

The next stage involves the division of the test records into a number of groups, within each of which all or some of these possible factors are roughly constant. This grouping, which is essential in any process of analysis, is described as division into *rational subgroups* (see 10.5.2). Suppose now that the 64 tests in the present example fall naturally into 4 subgroups, which is thought might be differentiated owing to changes in one or other of the factors suggested in Figure 2. The result is shown in the *dot plots* of Figure 3.

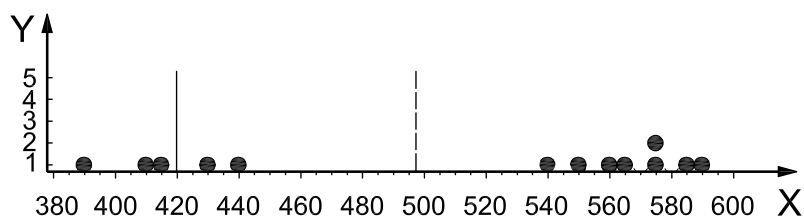


Key

X strength

Y number of observations in group 1

a) Group 1: Low oil quench temperature — High steel temperature

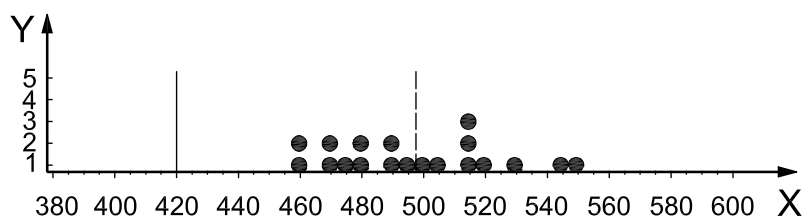


Key

X strength

Y number of observations in group 2

b) Group 2: High oil quench temperature — High steel temperature

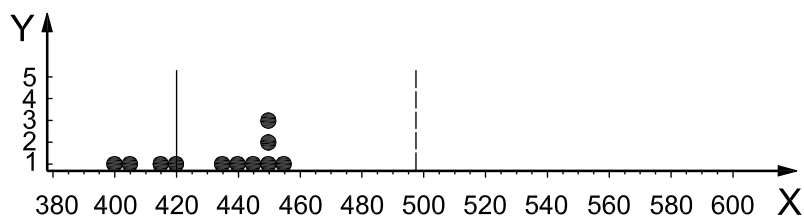


Key

X strength

Y number of observations in group 3

c) Group 3: Low oil quench temperature — Low steel temperature



Key

X strength

Y number of observations in group 4

d) Group 4: High oil quench temperature — Low steel temperature

The dotted line in each subfigure at 497,5 represents the sample median of all 64 test results and the full line at 420 represents the lower specification limit.

Figure 3 — Line plots showing patterns of results after division into rational groups

A study of Figure 3 indicates the following.

- a) Group 1 results are similar to those of group 3. This suggests that strength does not appear to vary a lot at low oil quench temperatures even if the steel temperature varies. The technical expression for this is that the process is *robust* to steel temperature variation at low oil quench temperatures. The means are of the order of 500, or greater, and the minimum sample values about 460, compared with the minimum specification value of 420.
- b) A study of group 2 and group 4 results appear to indicate a very different situation. Group 4 results are consistently low at, or around, the minimum specification limit. Group 2 results, on the other hand, are in two sets: one low set, with a mean below the specification limit, comparable with those of Group 4 and another contrasting set with an extremely high mean at about 570 with a relatively low variation.

A comparison of the records of these two sets indicated that the low set corresponded with operating conditions where the preset high steel temperature had inadvertently dropped to a low value for a short period. At a high quench temperature, the wire strength is extremely sensitive to variation in steel temperature and extremely low results, with a high proportion below the specification limit, may be expected at low steel temperatures, whereas at high steel temperatures the high quench temperature appears to yield a far superior strength performance with a mean of the order of 570 with relatively low scatter. The relationship (which was later confirmed by statistical experimentation, but is not reported in this Technical Report) is shown diagrammatically in Figure 4. Note that Figure 4 is not exclusively a summary of Figure 3 a) to d) but also draws on the results of extra experimentation.

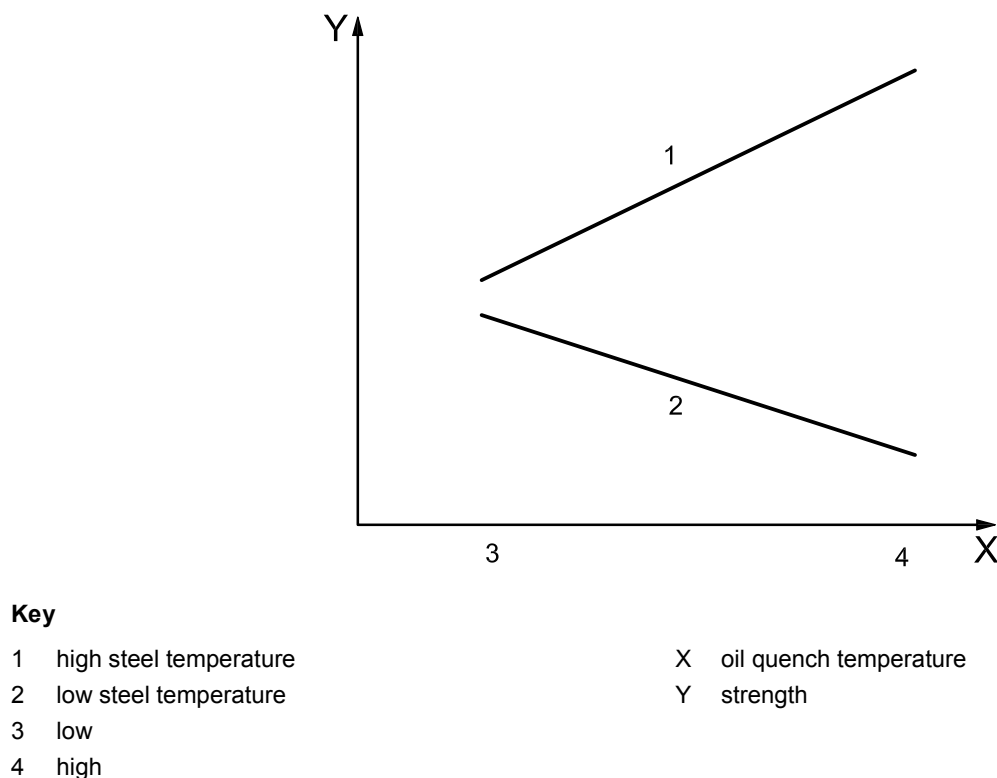


Figure 4 — Diagram indicating the effect of the interrelationship between oil quench temperature and steel temperature on wire strength

Now is decision time. How should this process be run to be likely to ensure uniform strengths of wire which do not contravene the lower specification limit? The choice depends on operational, economic, marketing and statistical considerations.

Option 1 is to run at low quench temperatures, which would be expected to give results similar to those of groups 1 and 3. Due to the predicted value of the mean and the pattern of variation, there would be some chance that occasionally the minimum specified strength may not be achieved. Variation in steel temperature between high and low would then be anticipated to have little impact on wire strength. Certain economies might be achieved using this option, by running with a lower steel temperature or a lower level of control of steel temperature.

Option 2 is to grasp the opportunity to achieve a relatively high mean wire strength with low variability by running the process with both a high steel and oil quench temperature. This may increase the process cost but it would ensure wire strength conformance to specification. It would also, perhaps, be appropriate to seek marketing advantage by improving the grade and increasing the price of the wire. However, the wire strength is seen to be particularly vulnerable to drops in steel temperature at high settings of oil quench temperatures. It is vitally important if this option is chosen to place strict controls on steel temperature.

4.2.5 General discussion on findings

This example has been used to suggest how simple statistical tools, coupled with technical and operational knowledge of the process, may help in improving process efficiency and performance and product conformity to specification. They provide powerful analytical and communicating tools and, at the same time, assist in determining, on a sound basis, simple routine checks on the efficiency of technical control.

Certain questions are posed. Should material of such great variation in strength be sold under the same specification? What is the relative cost to produce wire under some process parameter settings rather than others? Supposing that such variety is not desirable, what should be the best standard to aim at, having regard to the needs of the user and the obstacles to be overcome by the producer? Should the strength specification be modified, either downwards to encourage the attainment of the standard, or upwards following improvement in process settings and control, to increase the grade and price? To what extent are other product characteristics, such as hardness and brittleness, related to strength? Are trade-offs between one and the other involved?

In addition to these points, there is one more closely connected with statistical theory. The mere statement of means and minimum and maximum sample test strengths and the graphical display of the results, in the form of a dot or line plot, do not really provide measures of variation adequate for numerical prediction of the ability of the process to produce wire strengths conforming to standard. A number of other statistical aspects need to be considered, such as the stability of the process in relation to wire strength and the fitting of a probability distribution to the pattern of variation of the results (such as is shown later for a different example in Figure 26).

4.2.6 Explanation of statistical terms and tools used in this example

Rational subgroup: a group in which data is so organized through classifying, grouping or stratifying as to ensure the greatest similarity among the data within each subgroup and the largest difference between subgroups. The aim of rational subgroups is to include only *common causes of variation* within a subgroup with all *special causes of variation* occurring between subgroups. The object is to discriminate more readily between common and special cause variation in sets of data.

Knowledge and information, obtained through theory, experimentation or experience of the process, typically form the basis of the selection of rational subgroups. For example, in the administrative area, historical data on late payments could be grouped by account, account supervisor, product or by intervals of time. In a production process, the maximum homogeneity within a subgroup is frequently obtained by making up rational subgroups from consecutively produced parts taken from the same location or machine. For example, five consecutively produced parts from one machine may be taken every hour. It is then possible to segregate special causes of hour to hour variation, identified from subgroup to subgroup variation, from the inherent sources of common cause variation within a subgroup.

Common causes of variation: source of variation that is inherent in a process. It relates to those sources of *natural* variation in a particular process. For example, a turret capstan may produce to 0,25 mm, a grinder to 0,025 mm and a hand lapper to 0,002 5 mm; an investment casting may produce to 0,2 mm per metre and a sand casting to 0,8 mm per metre. Hence, only people responsible for the system can often reduce common cause variation. The variation is predictable in a process subject only to common cause variation.

Special causes of variation: source of intermittent variation in a process. A special cause arises because of specific circumstances that are not always present. For instance, it could be irregular (e.g. power surge), progressive (e.g. tool wear) or stepwise (e.g. change in datum of a gauge, or change in setting). As such, in a process subject to special causes, the magnitude of the variation from time to time is unpredictable. The presence of special cause variation is found using a statistical process control (SPC) chart by operational people, those who work in the system.

Dot/line plot: the frequency of readings at each measurement is shown by dots/lines built up vertically on a horizontal axis representing the scale of measurement. It can be used to compare or contrast, graphically, the pattern of variation of data both within a rational subgroup and between subgroups. It is particularly useful when working with limited sets of data.

Mean (arithmetic mean): sum of the values of the observations divided by the number of observations.

Median: value of a variable characteristic, which is greater than one half of the observations and less than the other half (the middle, or mid-value). In the case of an even number of observations, it is usual practice to set the sample median equal to the arithmetic mean of the central two observations after arranging the observations in ascending order.

Cause and effect diagram: frequently called a fishbone diagram (because of its shape) or an Ishikawa diagram (named after its creator, see Reference [101]). It applies where it is required to explore and display causes of a specific concern, problem or condition. The concern (effect) is shown on the right of a main horizontal spine. Possible categories of causes of the concern are shown on main branches from the spine. Subcategories are indicated on subbranches.

4.3 Example 2: Mass of fabric

4.3.1 General

This example illustrates a form of problem that arises in sampling anything, for example, a product or material, to determine whether or not it conforms to specification. It suggests 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 or measurements.

Specifications may be one-sided, with either a minimum (e.g. strength) or maximum value (e.g. eccentricity) quoted, or two-sided, with both a minimum and maximum given (e.g. assembly component dimension). When measurements are taken of successive results from however stable and precise a process, it cannot be supposed that the results will be identical. Some variation will be evident if the resolution of the measuring device is appropriate. Consequently, to obtain any adequate appreciation of the quality of the particular characteristic in question, a number of results need to be obtained. Furthermore, it is not only the resulting average that is of importance, but also the uniformity as measured by the variation about that average.

It follows that, in using a series of sets of measurements to check for conformity to a specification, it is helpful to take into account the following:

- a) the relationship between average values, minimum values, range of variation, etc.; together with
- b) the manner in which these are dependent upon the actual number of measured values taken.

4.3.2 Test results and specification limits

It is possible to illustrate the nature of the problem using the data given in Table 2. The figures represent the masses of specimens taken from a roll of fabric. They have been grouped for purposes of illustration into 32 samples of 4 specimens each.

**Table 2 — Masses of 128 specimens from a roll of fabric —
Minimum specification limit = 98 dg = 9,8 g**

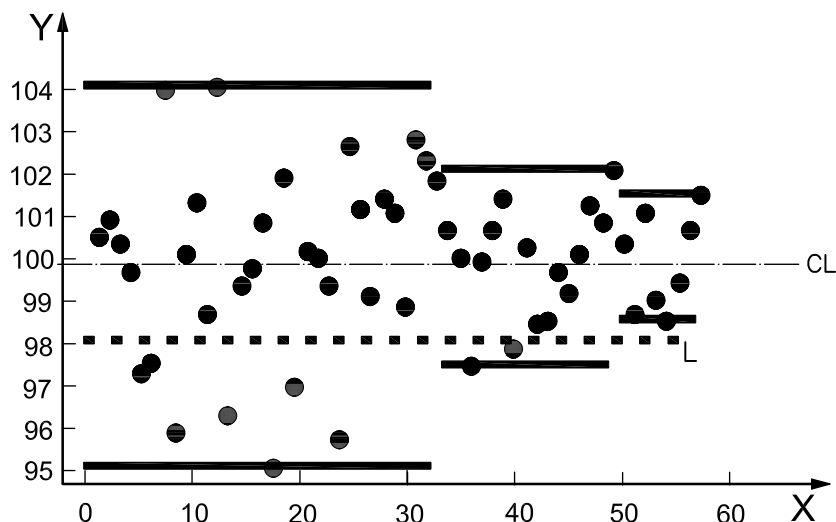
Sample number	Mass of specimens (1 dg)				Sample number	Mass of specimens (1 dg)			
	1	2	3	4		1	2	3	4
1	101	99	100	102	17	100	97	91	92
2	106	98	101	99	18	106	100	102	100
3	98	101	102	100	19	97	97	94	99
4	103	104	95	96	20	99	101	100	101
5	96	97	100	96	21	100	101	95	103
6	101	96	97	97	22	101	99	99	99
7	109	100	106	101	23	94	96	94	98
8	92	97	100	95	24	99	100	104	108
9	104	102	95	100	25	102	100	105	98
10	98	101	99	107	26	99	98	103	97
11	99	98	99	99	27	97	98	106	104
12	109	101	105	102	28	97	101	108	99
13	95	94	97	100	29	100	97	100	98
14	102	100	100	95	30	104	103	104	100
15	97	101	102	98	31	105	99	103	103
16	103	101	99	100	32	98	104	102	103

Figure 5 illustrates the data of Table 2 in 3 ways, in relation to sample means:

- the first 32 samples each have 4 masses (as shown in Table 2);
- samples 33 to 48 relate to the same data in Table 2, which has now been combined into 16 samples each of 8 masses with sample 33 being the union of samples 1 and 2, etc.; and
- samples 49 to 56 relate to the same data in Table 2, which has now been combined into 8 samples each of 16 masses with sample 49 being the union of samples 1 to 4, etc.

Figure 6 illustrates the data of Table 2 in 3 ways, in relation to sample ranges (see 5.2 for definition of range) rather than sample means. As in Figure 5:

- the first 32 samples each have 4 masses (as shown in Table 2);
- samples 33 to 48 relate to the same data in Table 2, which has now been divided into 16 samples each of 8 masses with sample 33 being the union of samples 1 and 2, etc.; and
- samples 49 to 56 relate to the same data in Table 2, which has now been divided into 8 samples each of 16 masses with sample 49 being the union of samples 1 to 4, etc.

**Key**

X sample number

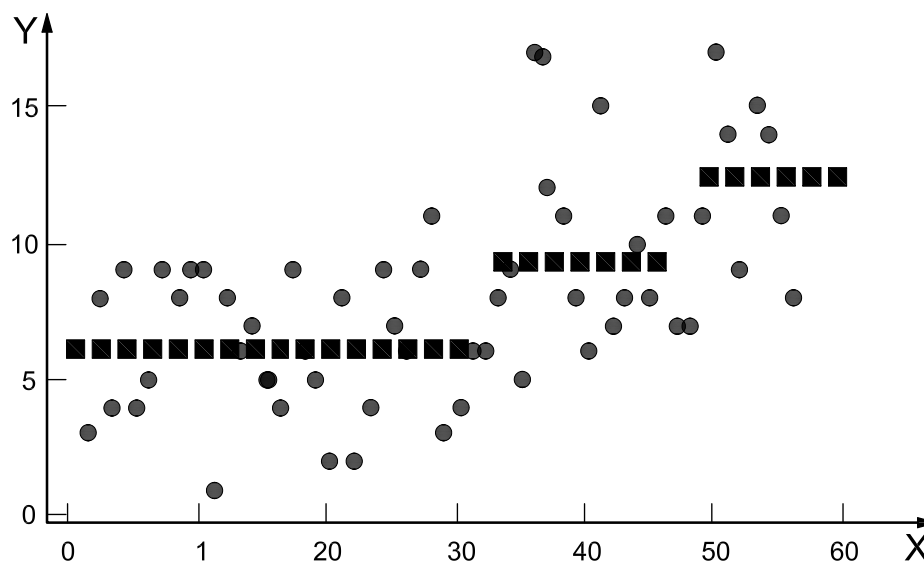
Y mass in units of 1 dg

NOTE 1 Bold horizontal lines indicate maximum ranges of means for sample sizes of 4, 8 and 16.

NOTE 2 L = lower specification limits

NOTE 3 CL = centreline

**Figure 5 — Means of masses plotted against sample number
(illustrating decreasing variation in the mean with the sample size increase)**

**Key**

X sample number

Y range in units of 1 dg

NOTE Dotted lines show increase with sample size in mean range within a sample.

**Figure 6 — Ranges of masses within each sample vs sample number
[illustrating increasing (range) variation within a sample with sample size increase]**

4.3.3 Discussion of specific results

Attention is drawn to the following points, brought out by examination of Figure 5.

a) The mean (Figure 5).

Sample size	Range of means (in units of 1 decigram)
4	95 to 104 = 9
8	97 to 102 = 7
16	99 to 101 = 2

The conclusion is that as the number of values upon which the mean is based becomes larger, the variation in the mean becomes smaller.

b) The range (Figure 6).

Sample size	Average range (in units of 1 decigram)
4	6
8	10
16	12

The conclusion is that the range of variation within a sample *increases* with the number of values in the sample.

c) Conformance to specification.

Suppose that the minimum mass, the lower specification limit (LSL), were to be set at 98.

1) If this criterion is applied to the *mean value* in a sample, then:

- i) 7 of 32 samples of 4;
- ii) 1 of 16 samples of 8;
- iii) 0 of 8 samples of 16

would fail to meet the criterion.

2) On the other hand, if this criterion is applied to the *smallest value* in the sample, then:

- i) 15 of 32 samples of 4;
- ii) 12 of 16 samples of 8;
- iii) 8 of 8 samples of 16

would fail to meet the criterion.

This example is discussed further in 8.3.

4.3.4 Discussion on general findings

Without placing undue emphasis on figures that are based on a single sample, the following general conclusions can be established by statistical theory and turned to practical account.

- a) As the number of values or tests becomes larger, the (absolute value of the) expected difference between the mean of one set of tests and that of another becomes smaller.
- b) As the number of values or tests becomes larger, their expected range of variation also becomes larger.
- c) A statement by way of specification that the lower specification limit = 9,8 g, say, is inadequate unless supplemented by
 - 1) information concerning the number of specimens to be tested,
 - 2) whether or not something conforms to specification.

Without this information, it would not be possible to know whether something conforms or does not conform to specification.

4.4 Example 3: Mass fraction of ash (in %) in a cargo of coal

4.4.1 General

This example illustrates four principal concepts:

- the handling of apparent fluctuation of a quality characteristic within a quantity of material (or alternatively with time);
- the need to determine, on a sound basis, the extent of sampling necessary to estimate the quality characteristic of a commodity;
- the necessity to establish, in advance, a well designed sampling procedure based on sound, but basic, statistical principles;
- the value of progressive analysis of results, in a simple graphical manner, as they become available.

These concepts are explained by reference to bulk sampling. However, they are applicable generally to all kinds of sampling.

Sampling of bulk commodities, for example, particulates, liquids and gases, can be classified into two types:

- a) sampling to make a decision on lot acceptance/rejection (see ISO 10725 ^[39]);
- b) sampling to make an estimation of the average quality of a particular characteristic of the bulk commodity. (See ISO 11648^[43].)

Illustrations of the application of b) include sampling of chemical products such as those in liquid state, cokes, ferroalloys and cements; agricultural products such as grains and flours; minerals and liquid state petroleum products. Sampling may take place on moving streams or in stationary situations such as stockpiles, silos, wagons and holds of ships and barges.

For this particular example, the quality characteristic chosen is the mass fraction of ash (in %) in a ship's cargo of coal. The aim is to estimate the average (arithmetic mean) value. The prime purpose of such sampling is, typically, to obtain an appreciation of the quality characteristic of the bulk of the fuel as a basis for determining the price to pay for the consignment.

4.4.2 Test results (reference ISO 11648-1: *Statistical aspects of sampling from bulk materials*)

Table 3 shows the test results from a series of 20 lots of coal being unloaded from a ship. For each lot, eight samples of coal were drawn and the mass fraction of ash (in %) measured.

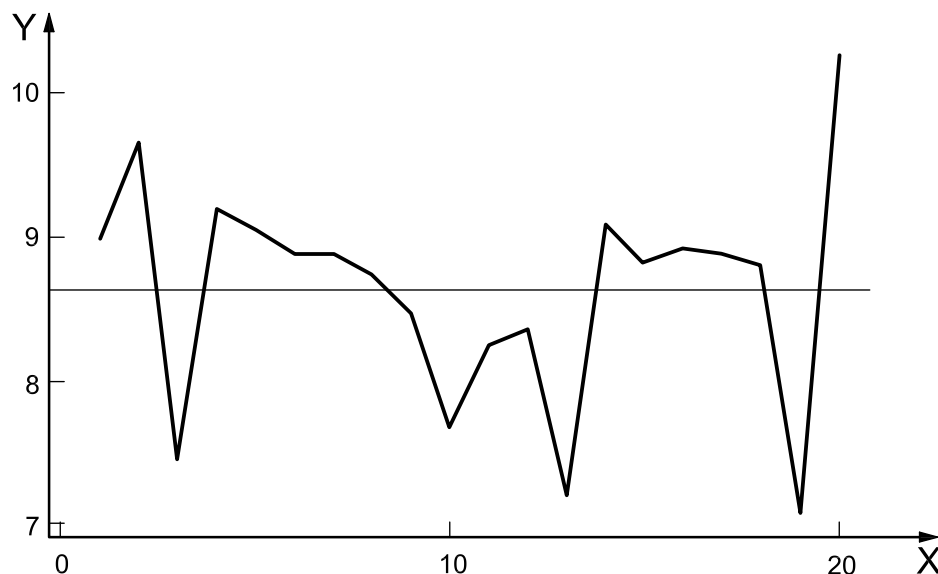
Table 3 — Mass fraction of ash (in %) measurement results by lot from ship's cargo on unloading

Lot no.	Result 1	Result 2	Result 3	Result 4	Result 5	Result 6	Result 7	Result 8
1	9,38	9,24	9,02	8,98	9,22	9,32	8,40	8,38
2	9,76	9,80	9,92	9,92	9,36	9,36	9,72	9,54
3	7,40	7,26	7,32	7,40	7,55	7,61	7,57	7,49
4	8,62	8,76	8,82	8,84	9,20	9,34	10,00	10,00
5	9,16	9,18	8,72	8,68	8,89	8,75	9,51	9,47
6	9,08	9,08	9,06	8,86	8,80	8,84	8,76	8,60
7	8,77	8,69	8,77	8,75	9,16	8,92	9,06	8,94
8	8,62	8,68	8,80	8,42	8,78	9,02	8,62	8,94
9	8,60	8,74	7,10	7,22	8,88	9,10	9,08	9,00
10	6,96	7,20	7,32	7,40	8,59	8,89	7,55	7,43
11	8,44	8,26	7,92	7,70	8,65	8,45	8,37	8,15
12	8,24	8,00	8,38	8,12	8,42	8,26	8,78	8,72
13	7,21	7,25	6,85	7,03	7,21	7,31	7,31	7,39
14	8,84	9,00	8,96	8,90	9,24	9,16	9,20	9,38
15	8,45	8,51	8,91	8,79	9,00	9,06	8,86	8,96
16	9,02	9,08	9,16	9,08	8,75	8,83	8,65	8,75
17	8,71	8,77	8,75	8,75	8,98	8,96	9,00	9,18
18	8,77	8,92	9,24	9,32	8,82	8,64	8,32	8,42
19	7,37	7,39	7,13	7,25	7,10	6,92	6,64	6,74
20	10,12	10,02	9,96	9,94	10,72	10,78	10,30	10,30

4.4.3 Initial graphical analysis of specific results

A plot of the averages of mass fraction of ash (in %) of the coal by lot is shown in Figure 7. The points have been joined by straight lines to aid visualization of the variation.

Fluctuation is observed about the overall average of 8,63 % ash content. Suppose 8,6 % is taken to represent the ash content in the whole consignment. A practical question would then be how many sets of tests need to be made before it would be reasonable to estimate this measure within, say, ± 1 % of its value (i.e. approximately 8,5 to 8,7). This question can be answered by reference to the progressive average plot in Figure 8.



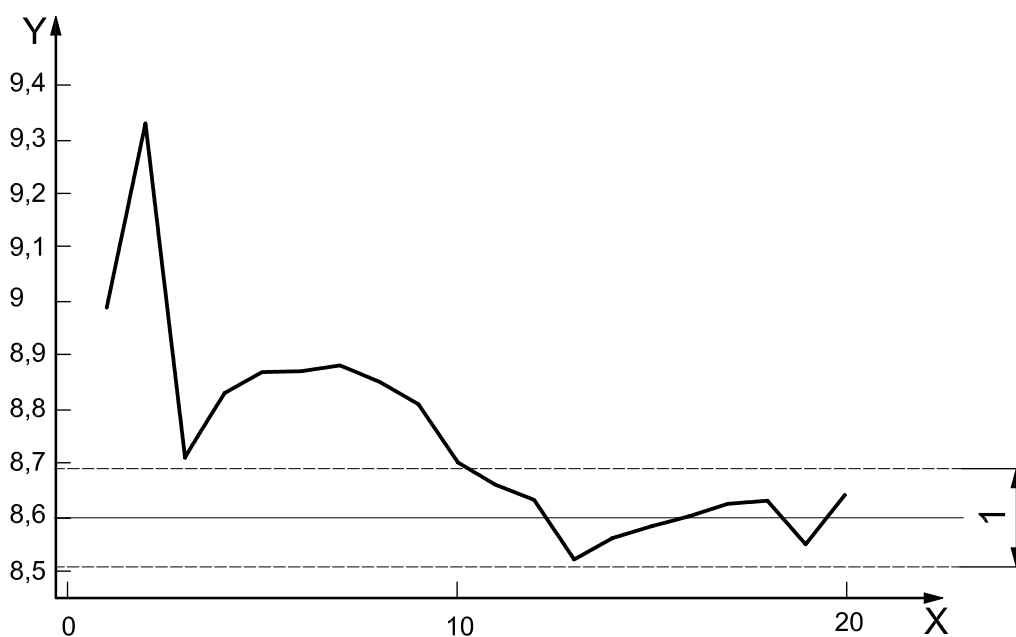
NOTE The value of mass fraction of ash (in %) in the figure for each lot represents the average of 8 results for the lot. Namely, from Table 3, the value for lot 1 = $(9,38 + 9,24 + 9,02 + 8,98 + 9,22 + 9,32 + 8,40 + 8,38) / 8 = 8,99$, and so on.

Key

X lot number

Y average of mass fraction of ash (in %)

Figure 7 — Averages of mass fraction of ash (in %) of coal by lot from cargo



NOTE The value of mass fraction of ash (in %) plotted by lot number represent progressive, or cumulative, averages of all measured values up to that lot. For example, the averages of the first three lots are: 8,99, 9,67 and 7,45. The corresponding progressive means are 8,99, $(8,99 + 9,67) / 2 = 9,33$ and $(8,99 + 9,67 + 7,45) / 3 = 8,70$.

Key

1 ± 1 % band

X lot number

Y progressive average of mass fraction of ash (in %)

Figure 8 — Progressive averages of mass fraction of ash (in %) in terms of lot

Figure 8 shows that, as the number of sampling lots increases, the progressive average appears to approach a limiting value of approximately 8,6. From the 10th sampling lot, the fluctuation of the progressive average has stabilized to fall within the ± 1 % bounds. That is to say, in this particular case, some 10 sampling lots would be required before a stable estimate lying within ± 1 % of 8,6 % could have been made of the ash content. However, it is important that this is not taken as a general rule. Much will depend upon the homogeneity of the consignment, the mass of the sample, the sampling and sample preparation procedures, sampling plan design, instrument resolution, etc. It is in the determination of the relationships between these factors that the methods of statistical analysis are useful.

4.4.4 Benefits of a statistically sound sampling plan

The benefits of a statistically sound sampling design become evident on analysis and attempting to draw conclusions from the results. For example, it is noted from purely cursory observation of Table 3 that:

- a) there is noticeable variation within each column of results;
- b) the rows of results (i.e. within lots) for the main peaks and troughs of Figure 7, lots 2 (high), 3, 13 and 19 (low), and 20 (high), have similar patterns of variation;
- c) there are adjacent column pairs of very low values in lot 9 (7,10 and 7,22 for results 3 and 4, respectively) compared with the six other values ranging from 8,6 to 9,1;
- d) there are adjacent column pairs of very high values in lot 10 (8,59 and 8,89 for results 5 and 6 respectively).

What do these results indicate? To answer this question, it is necessary to refer to the plan for sampling percentage ash from the ship's cargo. This is shown in Figure 9.

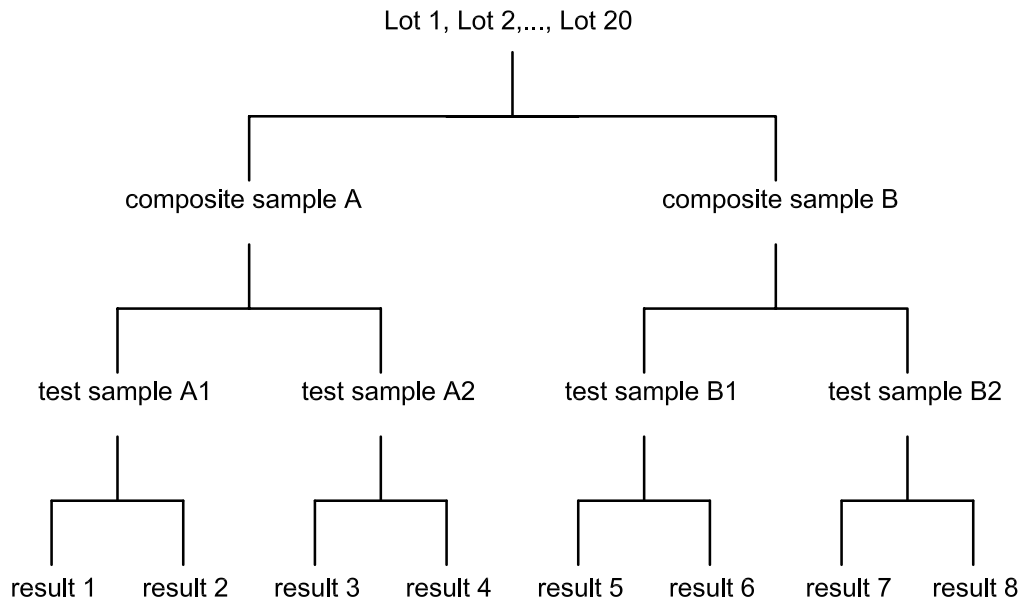
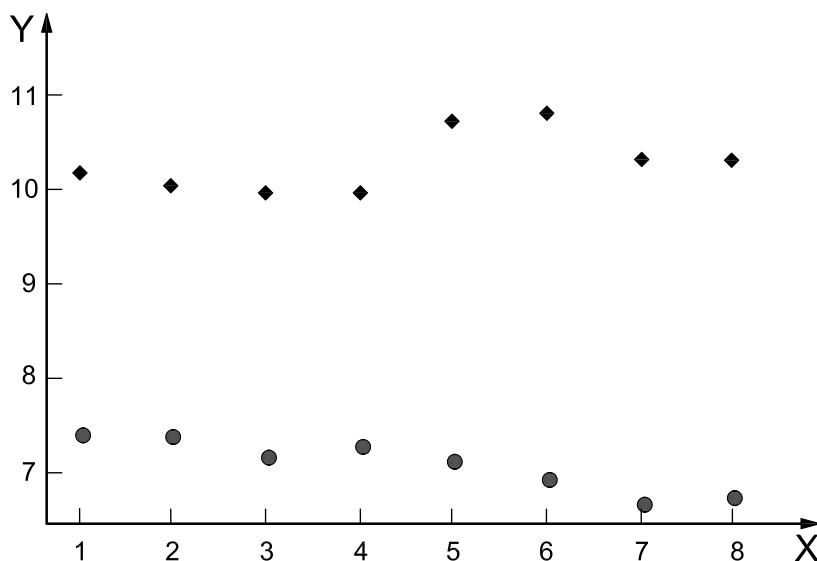


Figure 9 — Schematic diagram showing plan for sampling percentage ash from cargo of ship

This statistical design permits the isolation of lot to lot, composite sample to composite sample, test sample to test sample and measurement variation. This type of design is recommended (see ISO 11648-1) when there is no, or little, prior knowledge about the sampling situation.

In this particular case, the conveyor belt unloading coal from the ship was stopped at uniform time intervals.

A prespecified mass increment of coal was shovelled from the conveyor belt. Individual consecutive increments were placed alternately into two containers, A and B. Each of the two containers ultimately contains 30 such increments, which make up so-called composite samples. Two test samples are then prepared from each composite sample. Ash content is then analysed in duplicate on each test sample.



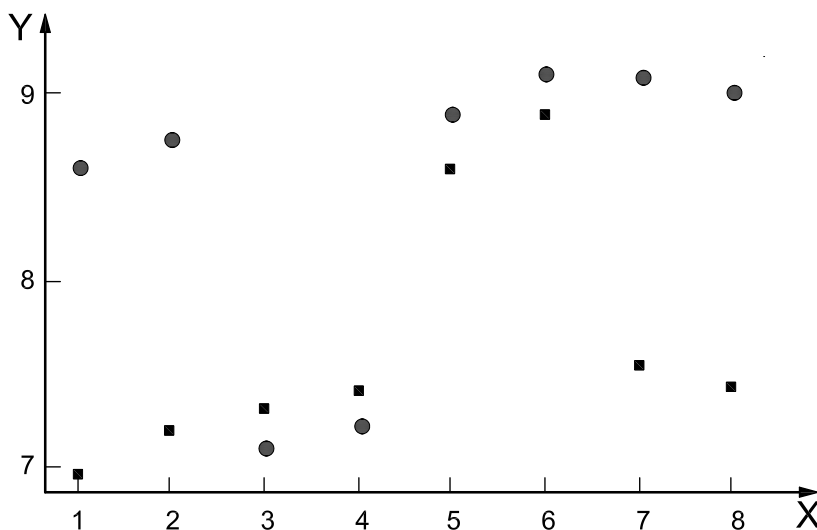
NOTE Circle: lot 19; Diamond: lot 20.

Key

X test number

Y mass fraction of ash (in %)

Figure 10 — Mass fraction of ash (in %) plotted against test number for lots 19 and 20 (illustrating relative consistency of percentage ash within each of these lots)



NOTE Circle: lot 9; Square: lot 10.

Key

X test number

Y mass fraction of ash (in %)

Figure 11 — Mass fraction of ash (in %) plotted against test number for lots 9 and 10 (illustrating rogue pairs in both lots)

Because of the design of the statistical sampling plan, certain conclusions may now be drawn, for example:

- i) in general, there is more lot to lot variation (row to row variation in Table 3) than within lot variation (within row variation);
- ii) Figure 10 confirms, with respect to lots 19 and 20, the relative consistency of test results within these lots;
- iii) Figure 11 indicates two pairs of rogue values in lots 9 and 10. Reference to the sampling plan indicates that these pairs are associated with test sample A_2 in lot 9 and B_1 in lot 10. The latter phenomenon could be due to problems in sample preparation or, perhaps, an abrupt change in calibration level of the measurement system. In retrospect, it is not possible to assign this special cause specifically. However, if simple graphical analysis such as this is ongoing, as results become available, and is not left to be done retrospectively, it is more likely that the specific cause of events such as these could be identified, at the time and place of the sampling activity. Operational or technical personnel could then take action, to remove the cause and its effect by eliminating the rogue values or substituting more representative ones.

4.4.5 General conclusions

This example illustrates:

- a) the importance of the deployment of statistical thinking and design method to a numerical study *prior* to its being undertaken;
- b) the value of the progressive application of simple, mainly graphical, statistical tools to any numerical study at the time and place of the particular activity rather than *just* applying more sophisticated statistical methods retrospectively;
- c) that to gain full benefit from b) it is essential that personnel who are technically and operationally familiar with the activity under scrutiny are involved in the progressive statistical analysis because it will facilitate the early identification and removal of any special cause variation that may be found to be present.

A more sophisticated retrospective statistical study of the results in this example included the use of analysis of variance (ANOVA). This confirmed that most of the overall variation in mass fraction of ash (in %) (84 %) was attributable to lot to lot variation, indicating variability in the ash content of the cargo. About 7 % to 8 % was attributed to each of composite and test sample variation and less than 1 % to measurement variation.

5 Introduction to basic statistical tools

5.1 General

The examples in Clause 4 give a general idea of the function of statistical methods in the analysis, control and reduction of variation and the usefulness of simple graphical presentation of data. Before developing in greater detail the application of these methods to *quality*, *specification* and *standardization*, it is necessary to describe more fully some of the basic tools.

Suppose that a single *quality characteristic* has been measured and recorded for each of a number of *objects*. The objects/characteristics of interest may be teeming temperature or vacuuming time in a steel mill; lateness of trains; time to pay invoices; length, diameter, surface finish or eccentricity of a component; hardness or silicon content of a material; time to answer a telephone; time to failure; noise level; emission level of engines. This partial listing gives some impression of the broad applicability of these tools. The measured values of characteristics will be termed values or observations.

5.2 Basic statistical terms and measures

If a group of units or quantities have been selected from a larger whole, it is defined in statistical terminology as a *sample*. It is also common to speak of the individual observations themselves as forming a sample. Thus, a sample may consist of 1, 2, 3, ..., n units or observations.

A sample that is drawn without bias is termed a random sample. The larger whole of units, which is to be the subject of sampling (e.g. all students at a college at the time of a survey), is called a *population*. (Further discussion of this concept is provided in 7.1 and 8.1.1.) A sampling *frame*, on the other hand, is a list of sampling units from which the sample is taken (e.g. college register). In the subclauses 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 viewpoint, may be equally applied whether the group of units under consideration forms the sample or the population.

With a group of observations, three aspects are of prime importance. These are:

- a) a measure of central tendency;
- b) a measure of the magnitude of the variation;
- c) the pattern of variation.

There are various methods for measuring the central tendency and the magnitude of the variation within a group of observations. An important feature that is, unfortunately, frequently not taken into consideration in the application of these measures is the *pattern of variation*. For example, too often normality (symmetrical bell-shaped distributions) is assumed in process capability studies, and constant failure rates in the specification of, and performance claims for, equipment reliability.

Central tendency is most commonly expressed in terms of:

- 1) *arithmetic mean* (or just *mean* or *average*): the total of the values divided by the number of values;
- 2) *median*: the central value when the data are ranked in order of size and when the number of observations is odd; if the number of observations is even, then the median is usually taken to be the average of the two central values;
- 3) *mode*: the most frequently occurring value.

The two most frequently used measures of *variability* are:

- i) *range*: the difference between the smallest and largest values in the data;
- ii) *standard deviation*: measures the variation of the data around the mean. The less the variation, the smaller the value. When derived from a sample, the value is given by the expressions:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n - 1)}}$$

where

$\sum$ is the sum of;

x is the individual value;

$\bar{x}$ is the arithmetic mean of the individual values;

n is the number of values;

s is the sample standard deviation.

A summary of the relative advantages and disadvantages of these measures is given in Table 4.

Table 4 — Advantages and disadvantages of various statistical measures

Measure	Advantages	Disadvantages
Mean	Easy to understand Commonly used	Affected by very high or low values Need all the data to calculate
Median	Unchanged by very high or very low values	Slow and tedious to calculate for large sample sizes unless a suitable calculator or computer facility is available
Mode	Unchanged by very high or low values	May be multi-modal
Range	Easy to calculate	Uses extreme values only
Standard deviation	More efficient than range	Less easy to calculate manually

To illustrate these terms, a set of five values is used: 7, 5, 10, 7 and 6. Using these values, the various statistical measures are as follows.

$$\text{Arithmetic mean} = (7 + 5 + 10 + 7 + 6)/5 = 7$$

$$\text{Median} = \text{central value of ordered set, } 5, 6, 7, 7, 10 = 7$$

$$\text{Mode} = \text{most frequent value} = 7$$

$$\text{Range} = \text{maximum value} - \text{minimum value} = 10 - 5 = 5$$

$$\text{Standard deviation (manual method is shown below)} = 1,87$$

Sample value	$(x - \bar{x})$	$(x - \bar{x})^2$
7	0	0
5	-2	4
10	3	9
7	0	0
6	-1	1

$$\sum (x - \bar{x})^2 = 14$$

and thus

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{14}{5 - 1}} = 1,87$$

Alternatively, the standard deviation may be obtained much more quickly and directly using a scientific calculator.

5.3 Presentation of data

5.3.1 Dot or line plot

Particularly when only a few observations are available, dot or line plots, such as those shown in Figure 1 and Figure 3, 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. The corresponding line plot for the dot plot of Figure 1 is shown in Figure 12.

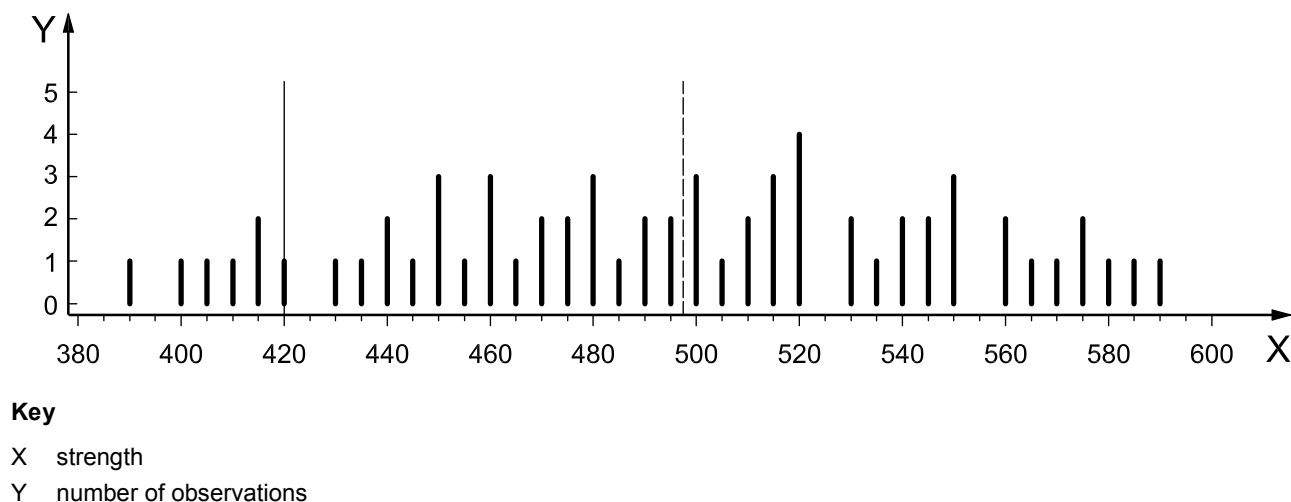


Figure 12 — Line plot of breaking strength of wire (Table 1 data)

5.3.2 Tally chart

A tally chart may be applied to both measured and classified data. It is used to visually represent the frequency of a particular value, or a specific type of event, in a series. The five-bar gate notation is used. Examples are shown in Figure 13 for both measured data and classified events.

21		open circuit	
22		short circuit	
23		dry joint	
24		solder splash	
25		wrong component	
26		broken lead	

a) Tally chart for measurements b) Tally charts for events/counts

Figure 13 — Typical tally charts

5.3.3 Stem and leaf plot

The stem and leaf plot displays the pattern of variation of measured data. It is an enhanced form of histogram or tally chart. In addition to showing the distribution of a set of data, it also shows individual values. Each value is split into two parts, the first part consisting of the leading digits, which are written downwards along the stem in ascending order, while the second part of the values, the remaining digits, are written in ascending order horizontally along the leaf on the line with their leading digits.

An example is shown for the following data in Figure 14.

Data: 29, 28, 41, 36, 36, 59, 50, 61, 44, 48, 35, 42, 53, 33, 31.

stem	leaf
2	8 9
3	1 3 5 6 6
4	1 2 4 8
5	0 3 9
6	1

Figure 14 — Stem and leaf plot for data

5.3.4 Box plot

The box plot is a very useful tool in exploratory data analysis. It is simple to construct and easy to interpret. Like the dot or line plot, it is used to depict the similarities within, or the differences between, different groupings of data. Tukey^[130] defines

$$\text{Lower limit (LL)} = Q_1 - 1,5 \times \text{IQR}$$

and

$$\text{Upper limit (UL)} = Q_3 + 1,5 \times \text{IQR},$$

where IQR is the inter-quartile range defined, i.e. $\text{IQR} = Q_3 - Q_1$. Observations outside these limits may be outliers, and should be checked. If the data are from a normal distribution, the area outside these limits is approximately 0,7 %.

The box plot represents a five-number summary of a data distribution consisting of

$\tilde{x}$ median (mid) value;

Q_1 first quartile (value below which $\frac{1}{4}$ of values lie);

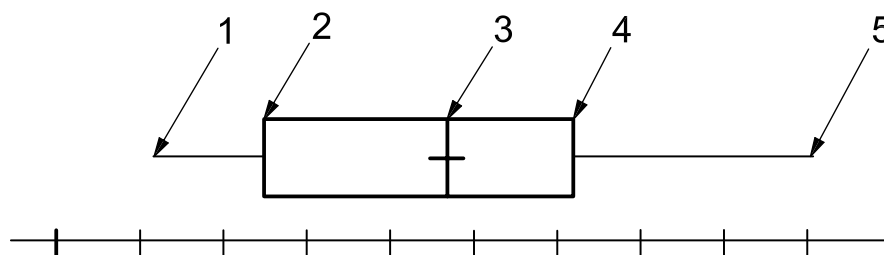
Q_3 third quartile (value above which $\frac{1}{4}$ of values lie);

A_1 lower adjacent value = smallest data value $\geq$ LL;

A_2 upper adjacent value = largest data value $\leq$ UL.

A basic box plot consists of a box, the length indicating the region where 50 % of the readings lie, a median line, and whiskers extending from the box to the lower and upper adjacent values. Each data value outside the whiskers is drawn as a point alongside the axis defined by the whiskers. A box plot is shown pictorially in Figure 15.

The box plot may be extended, for example, to include a display of statistical confidence limits around the median. An absence of overlap of these statistical bounds between groups would indicate statistically significant differences between the medians of these groups. Also, the width of the box may be varied to indicate changes in relative size of different groups. Outliers (apparent rogue values) may be shown by an asterisk.

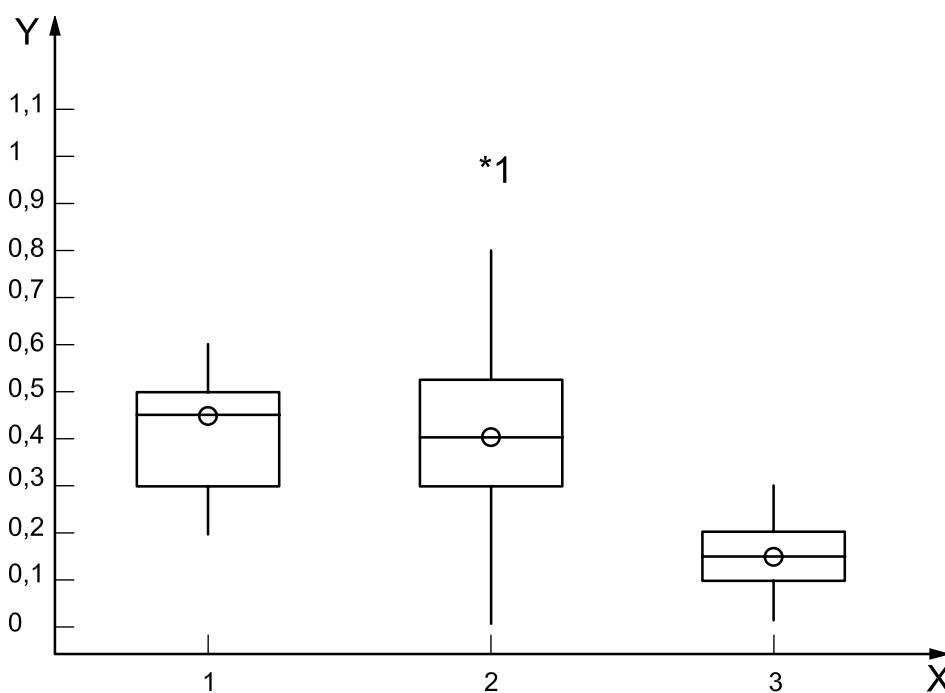
**Key**

- | | | | | | |
|---|----------------------------|---|----------------------|---|----------------------------|
| 1 | A_1 lower adjacent value | 3 | $\tilde{x}$ median, | 5 | A_2 upper adjacent value |
| 2 | Q_1 first quartile | 4 | Q_3 third quartile | | |

Figure 15 — Box plot

The box plot may be augmented by more formal statistical methods such as analysis of variance (ANOVA).

An example of the applicability and value of a box plot is shown in Figure 16. The shade variation of fabric of a particular colour was compared between adjacent panels on typical items of clothing sourced from three different suppliers. The results are shown in box plot form in Figure 16.

**Key**

- 1 two extreme values at 1,0 out of 30

X supplier

Y delta E value

NOTE 30 samples tested from each supplier.

Figure 16 — Box plot for Delta E panel shade variation between supply sources

The box plot indicates considerable variation in standards of performance between suppliers both in terms of process targeting (indicated by the relative positions of the median) and consistency about that target (indicated by the differences in lengths of the whiskers). The asterisk indicates two outlying values indicating lack of control of the dyeing process.

The box plot indicates that:

- a) supplier 1 has approximately the same median Delta E value as supplier 2;
- b) supplier 2 has a dyeing process with essentially the same variation as supplier 1; two very high values are also present (in a limited test sample of 30), which is likely to give rise to extreme customer dissatisfaction and a loss of quality reputation by the retailer; if typical of production, major recalls may be expected;
- c) supplier 3 has a dyeing process distributed around a low (good) Delta E value with much smaller variation about that value than the other two suppliers.

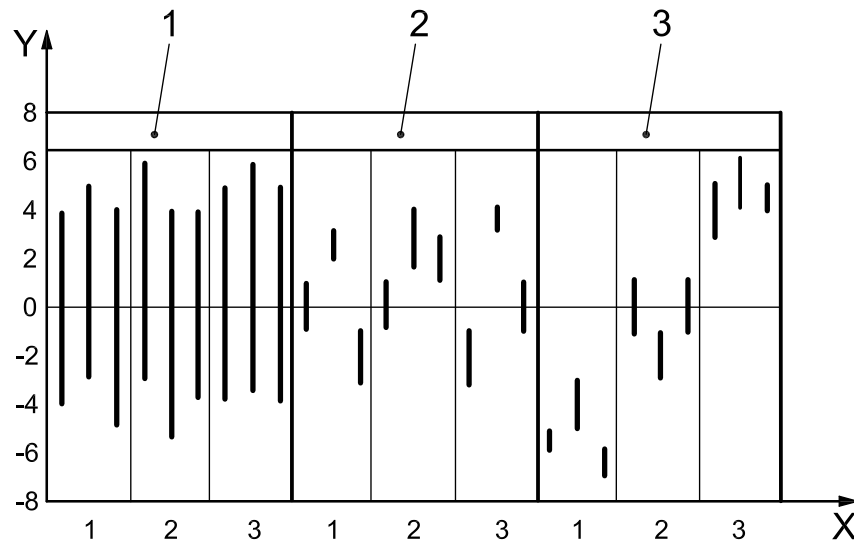
The Delta E results on supplier 3 merchandise indicate what can be, and is being, achieved in terms of shading performance. This “current state of the art” or “best practice” result then becomes the benchmark or reference standard for all supply sources.

5.3.5 Multi-vari chart

The multi-vari chart is a simple pictorial method of indicating and comparing the magnitude of different sources of variation. As such, it is very useful for diagnostic and investigation purposes rather than, and as a precursor to, ongoing process control. It consists essentially of vertical lines joining maximum and minimum values for a particular characteristic against a measurement scale: a max.-min. plot. A *dot* on nominal size represents an ideal value. The longer the line, the greater the variation.

Take a turned diameter where three consecutive components are taken from production and measured each hour. A clock gauge is used which records maximum and minimum values of the diameter of each component as it is rotated. The multi-vari chart in Figure 17 shows, for three quite different process performance scenarios, the dominant sources of variation prevailing, namely:

- a) within part (geometric form) variation;
- b) part to part variation;
- c) time to time variation.

**Key**

- | | | | |
|---|--|---|-------|
| 1 | scenario 1: Large within part variation | X | time |
| 2 | scenario 2: Large part to part variation | Y | value |
| 3 | scenario 3: Large time to time variation | | |

NOTE Maximum and minimum of three consecutive parts every hour for three hours showing three quite different process performance scenarios.

Figure 17 — Multi-vari chart as a tool for process variation analysis

5.3.6 Position-Dimension (P-D) diagram

A P-D diagram can be looked upon as an extension to the multi-vari chart to handle more than one feature. A P-D diagram representing ideal values in relation to, say, ovality and taper of a cylinder, is a horizontal straight line on a vertical dimension scale. If this line is coincident with the nominal or targeted value of the overall mean of the diameter, then this represents the ideal situation.

An example illustrates its usefulness. The variation in the outside diameter of a cylinder is being investigated for nominal size, ovality and taper. Measurements are taken at right angles to one another at each end of the cylinder as shown in Figure 18. A, B, C and D, as shown in Figure 18, identified these positional measurement values. One cylinder from production was measured every shift for 4 shifts. The machine tool was then overhauled and another set of readings taken.

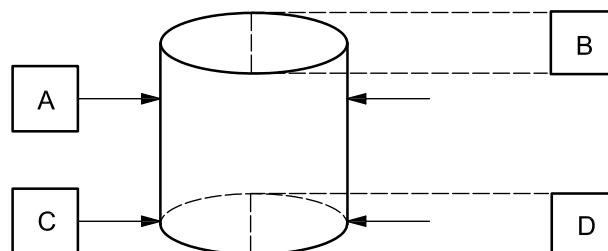
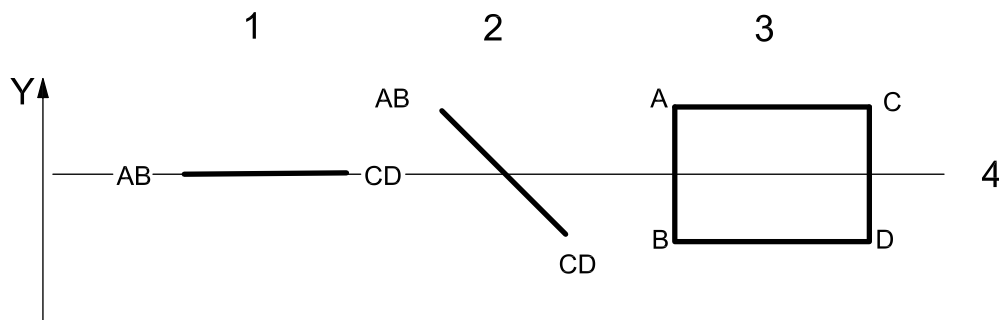


Figure 18 — Measurements on cylinder to determine nominal size, ovality and taper

Figure 19 provides the reference standard for this diameter and is used for judging the degree to which the diameter meets the preferred nominal value and the extent of geometric form variation present.



Key

1 ideal 2 pure taper 3 pure ovality 4 nominal value Y diameter

NOTE AB indicates A is coincident with B dimensionally and CD indicates C is coincident with D dimensionally.

Figure 19 — Measurement on cylinder — P-D diagrams showing ideal diameter values, pure taper and pure ovality

Regarding Figure 20 and the factors under investigation:

a) ovality:

- $A > B$ indicates ovality at the AB end and $C > D$ indicates ovality at the CD end;
- this progressively increases with time until overhaul;

b) taper:

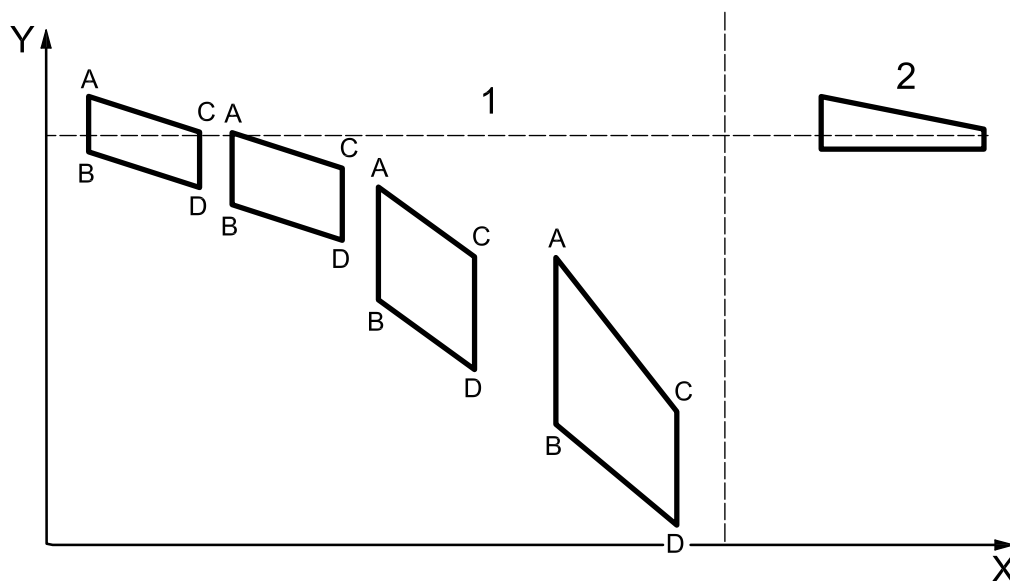
- the mean of A and B exceeding the mean of C and D indicates taper along the length of the cylinder;
- this, too, progressively increases with time until overhaul;

c) overall diameter size:

- the mean of A, B, C and D gives an estimate of the overall average diameter;
- this progressively decreases away from its nominal value until overhaul;

d) overhaul:

- improvements in overall diameter aim, and geometric form variation, with some ovality remaining at the AB end.



Key

1 before overhaul 2 after overhaul X time Y diameter

NOTE The dashed, vertical line indicates the time of the overhaul; on the left-hand side of the line, the situation before the overhaul is presented, and on the right-hand side, the situation after the overhaul. The dashed horizontal line represents the nominal value.

Figure 20 — Measurement on cylinder — P-D diagrams indicating progressive decrease of mean and increase in geometric form variation and the beneficial effects of overhaul

5.3.7 Graphical portrayal of frequency distributions

When only a few observations are available, *dot or line plots* as shown in Figures 3 and 12 or *stem and leaf plots* as in Figure 14 will often suffice. However, with a larger number of observations, and once the possibility of dependency has been discounted, it is generally found convenient firstly to arrange the data in numerical value order, the total observed range of variation in the measured characteristic is then divided into convenient equal intervals, and the number of observations falling into each interval is counted. This number is termed the frequency for that interval and the resulting tabulated series of numbers shows the frequency distribution. A simple method called Sturge's rule, i.e.

$$\text{Number of classes} = 1 + 3,3 \times \log_{10}(\text{number of observations})$$

provides guidance on the number of class intervals to select in terms of the total number of observations. This is shown, both in tabular and equation form, in Table 5. Sturge's rule should be taken purely as a starting point for experimenting with the number of classes to get the most informative view of the data. The number of class intervals actually chosen in a particular case should ultimately be chosen on the grounds of simplicity, clarity and ready understanding. This is illustrated in the example that follows.

Table 5 — Guidance on number of classes to select in terms of number of observations

Number of observations	20	46	91	191	371	751	1 501	3 001	6 001
	to 45	to 90	to 190	to 370	to 750	to 1 500	to 3 000	to 6 000	to 12 000
Number of classes	6	7	8	9	10	11	12	13	14

A grouped frequency distribution may be represented in a number of ways. Typical ones are:

- a) frequency table;
- b) frequency tally chart;
- c) histogram;
- d) cumulative frequency table;
- e) cumulative frequency plot.

The application of each of these methods is illustrated by example.

EXAMPLE Quality of zinc-coated item after galvanizing:

Selected test specimens typical of production are required to withstand a standard 4 min acid bath immersion test following galvanizing. Some 200 results that have accumulated over a period of time are used as the basis for this study. Measurements were taken to the nearest 0,1 min.

The results extended from 4,3 to 9,4. Sturge's rule suggests 9 class intervals for 200 results. This would give class intervals of $(9,4 \text{ min} - 4,3 \text{ min})/9 = 0,57 \text{ min}$. The results were arranged in ascending order and the actual class interval chosen was 0,5 min for greater simplicity and clarity. The resulting frequency and percentage frequency tabulation are shown in Table 6.

Table 6 — Frequency and percentage frequency table for immersion times withstood by test specimen

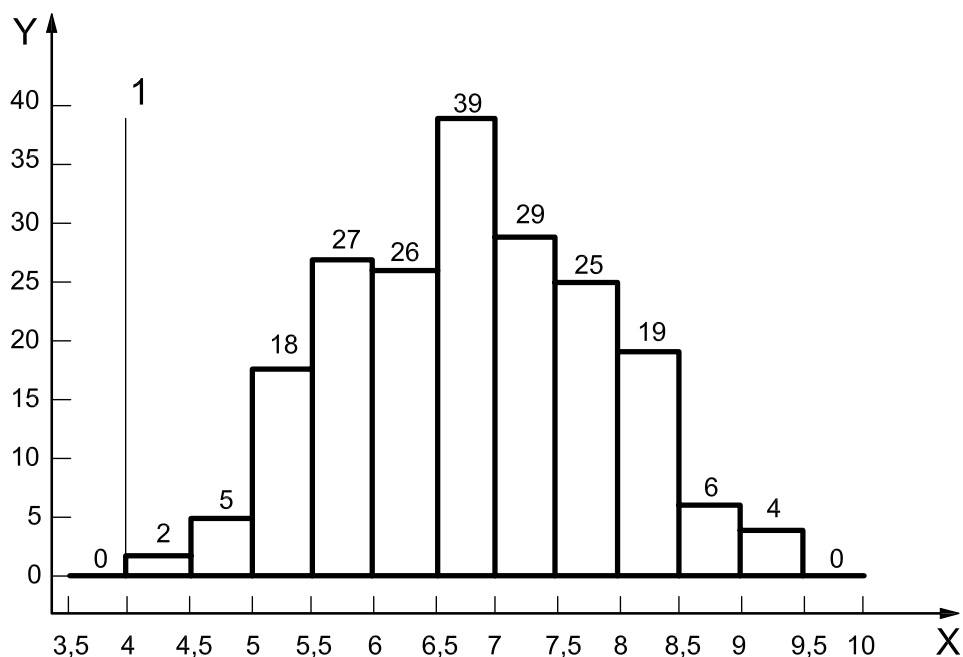
Immersion min	No. of observations (Frequency)	Frequency %
4,1 to 4,5	2	1
4,6 to 5,0	5	2,5
5,1 to 5,5	18	9
5,6 to 6,0	27	13,5
6,1 to 6,5	26	13
6,6 to 7,0	39	19,5
7,1 to 7,5	29	14,5
7,6 to 8,0	25	12,5
8,1 to 8,5	19	9,5
8,6 to 9,0	6	3
9,1 to 9,5	4	2

Table 6 illustrates that a frequency table summarizes a set of data by showing how often values within each class interval occur and that it may be enhanced by tabling the percentages that fall within each category.

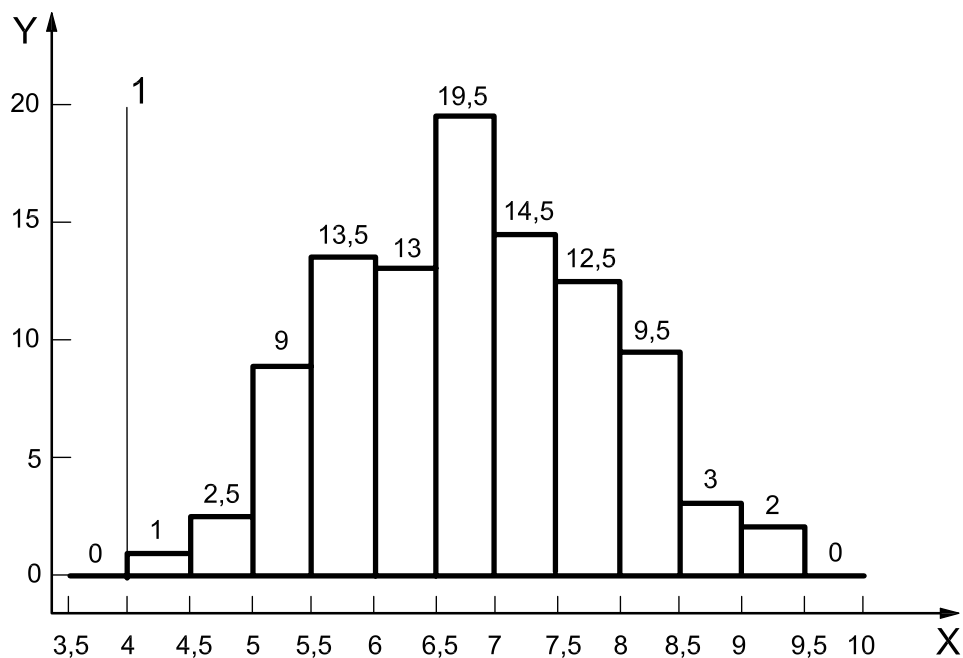
This permits some feel for how the data, as a whole, are distributed.

Various pictorial representations of the test data of Table 6 are shown in Figure 21 to Figure 26. They further enhance perception of the shape and pattern of the distribution of the data and its relation to the lower specification limit of 4 min.

The horizontal axis of the histogram corresponds with 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. Using this method, *frequency* in the table corresponds with *area* in the histogram.

**Key**

- 1 lower specification limit
- X time, in min
- Y number of observations (frequency)

Figure 21 — Frequency histogram for immersion times in Table 6**Key**

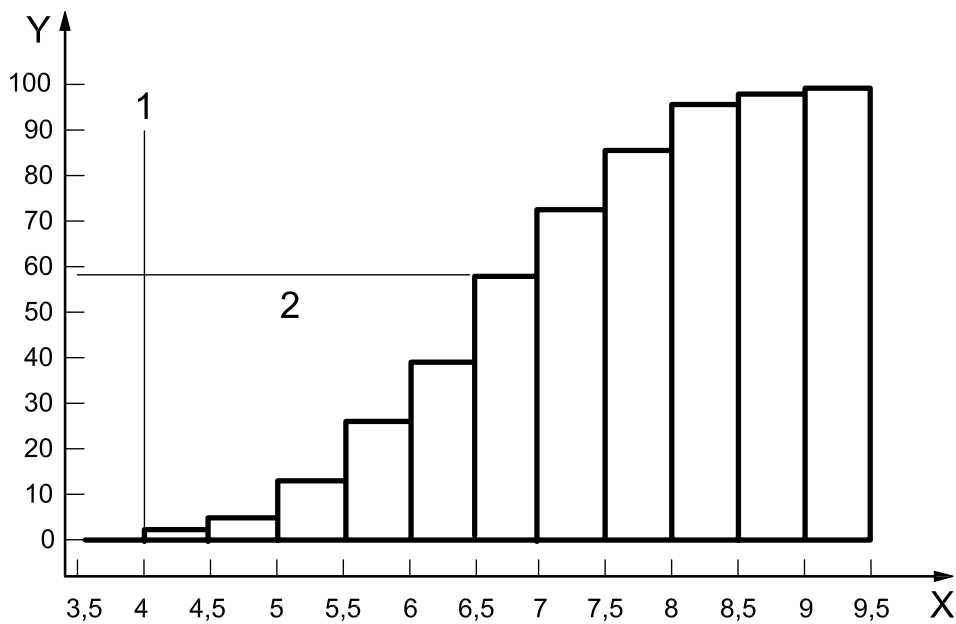
- 1 lower specification limit
- X time, in min
- Y frequency, in percent

Figure 22 — Percentage frequency histogram for immersion times in Table 6

Sometimes, it adds to understanding if relative (percentage) frequencies rather than actual counts of frequencies are used to construct the histogram. It also demonstrates the intermediate step to be taken in constructing a cumulative percentage frequency diagram.

The cumulative relative frequency diagram shows the percentage of observations falling below (or above) particular values. By way of illustration, in Figure 23 it is seen that 58,5 % fall below 7,0 min. Hence, this is a very useful diagram for determining the situation in relation to specification limits (lower and upper).

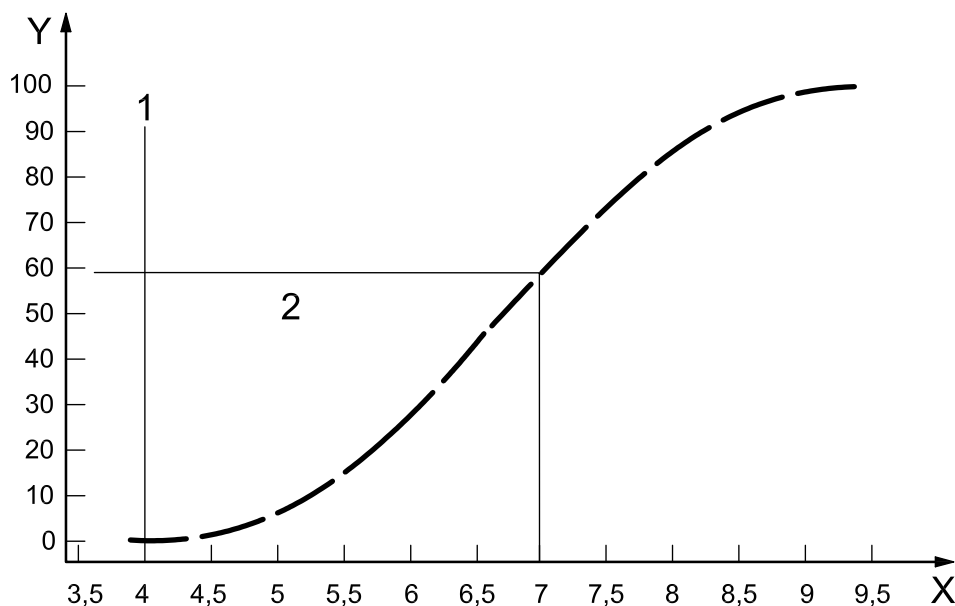
It should be borne in mind that the cumulative percentage values relate to the *upper limit* of the class interval and not to the mid-value.



Key

- 1 lower specification limit
- 2 58,5 % below 7,0 min
- X time, in min
- Y cumulative frequency, in percent

Figure 23 — Cumulative percentage frequency histogram for immersion times in Table 6

**Key**

1 lower specification limit

2 58,5 % below 7,0 min

X time, in min

Y cumulative frequency, in percent

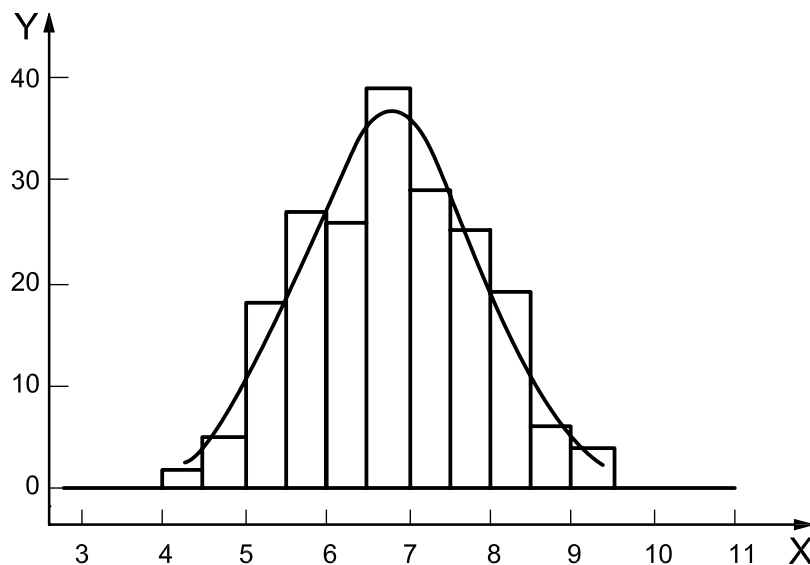
Figure 24 — Cumulative percentage frequency diagram for immersion times in Table 6

The cumulative histogram can alternatively be expressed in the form of a smooth curve as shown in Figure 24, or preferably as a straight line by transformation of the vertical scale. This latter concept will be developed later in this clause.

The diagrams of Figure 21 to Figure 26 portray the actual situation for a sample size of 200. What can be predicted about the galvanizing quality of production as a whole from this sample, assuming that the process is, and continues to be, stable about the present mean? This is where statistical modelling using the appropriate probability distribution can provide worthwhile quantitative information. A *best fit* probability distribution to match the actual frequency distribution is sought.

In the zinc plating case, the frequency histogram for immersion time (Figure 21) indicates a bell-shaped symmetrical pattern of variation about the mean. This is typical of the normal or Gaussian distribution. The normal curve has a definite mathematical equation that depends only on the values of the mean and standard deviation. Care should be taken not to interpret the word normal to mean that anything non-normal should be looked upon as being peculiar. Any characteristic that has a natural zero, for instance, such as taper, eccentricity and parallelism will naturally be skewed. Constant failure rates, looked upon as ideal in the reliability field, will naturally have a non-normal (negative exponential) frequency distribution. In fact, the existence of a normal failure frequency would indicate an undesirable increasing failure rate or wear-out regime.

However, a large number of the symmetrical frequency distributions met with in practice in the quality domain may be adequately represented by the normal curve. Does the normal distribution provide a reasonable fit to the immersion time data? The answer is given visually in Figure 25.



Key

X time, in min

Y number of observations (frequency)

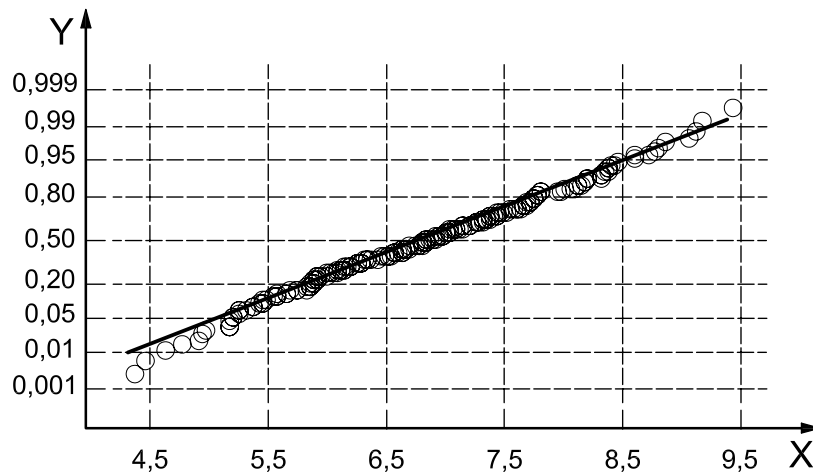
**Figure 25 — Normal curve overlaid on the immersion time histogram
(mean = 6,79; standard deviation = 1,08)**

Based on a calculated mean and standard deviation of the 200 results, the normal curve has been fitted to the data in histogram form as shown in Figure 25. These show, by eye, a good correspondence between the two indicating that a normal distribution with a mean of 6,79 min and standard deviation of 1,08 min is a reasonable representation, or model, of the actual data. There are a number of formal statistical tests for departure from normality. These include the Shapiro-Wilk and Epps-Pulley tests (for more information on these tests see ISO 5479 [18]).

A simple, practical, effective and graphical method involves the plotting of cumulative percentage frequencies on normal probability paper. If such a plot follows a straight line, then the sample can reasonably be regarded as having come from a normal distribution. If the plot indicates a systematic departure from a straight line, then the shape of the plot often suggests the type of distribution it represents. For example, a plot that forms a concave curve (curving downwards in the middle) would result from a log-normal distribution, whereas a plot curving upwards at its centre would result from an exponential or a gamma distribution. The plot on normal probability paper is often called a quantile-quantile diagram or a Q-Q plot and is done automatically as a part of the analysis in many statistical software packages.

In addition to checking for normality, this method is used extensively in statistical process control for capability and performance measurement (see 6.1.7).

An example of this test applied to the galvanized item immersion data is shown in Figure 26. It is seen that the normal cumulative probability scale shown on the vertical axis of Figure 26 transforms the bell-shaped normal distribution into a straight line when plotted against immersion times.

**Key**

X time, in min
Y probability

NOTE Average: 6,778 5; standard deviation: 1,056 31; number of data points: 200.

Figure 26 — Straight line plot on normal probability paper indicating normality of data in Table 6

Whatever the distribution, it is desirable to work in terms of a straight line reference standard for a number of reasons:

- 1) it permits a simple immediate visual test of fit against the underlying distribution (normal here);
- 2) it makes for ease of extrapolation and so facilitates numerical prediction of the likelihood of having values in a larger lot or consignment outside of those experienced in the sample;
- 3) it facilitates the correction of individual measurements and other errors;
- 4) it gives an immediate visual appraisal of the relationship of the data to any specification limits or reference standards in terms of both targeting and variability;
- 5) for an incapable process, it immediately provides an estimate of the proportion of values likely to be above and/or below specification limits;
- 6) it serves as a diagnostic tool to detect divergences from the model; for instance, a smooth concave or convex plot on a normal probability plot indicates skewness of the data.

Using probability paper based on the normal or other standard statistical distributions to represent data thus offers many practical advantages in the interpretation of results from samples. These other standard distributions include the fixed shape log-normal and extreme value distributions for moderately skewed data and the versatile multi-shaped Weibull distribution. The Weibull distribution is used extensively in the reliability field to model the various regimes of failure: infant mortality (decreasing failure rate), prime of life (constant failure rate) and wear-out (increasing failure rate). Confidence bands may readily be plotted around the best estimate straight line plots. These would be represented by curves.

5.3.8 The normal distribution

The previous example suggests the important descriptive part that the normal curve can play, always provided that its suitability to represent the type of variation in question has first been established. It would seem appropriate to take this opportunity to discourage a common belief. There is no magic about the normal curve that if a distribution follows this law then that is proof that the process giving rise to the product or service is, in fact, in control (i.e. stable). To arrive at reasonable judgements on past performance and to make rational predictions of future performance based on accumulated data, it is necessary to have prior knowledge that no special causes of variation were present over the period in which the data were gathered.

The name *normal* distribution originates from its use in pioneering statistical studies of human populations.

For a process not in control, the variation in its output is unpredictable. A primary role of statistical process control is to ensure, and assure, process stability. However, the normal distribution is the one most frequently encountered in many processes.

Moreover, the distribution of means of samples, or subgroups, will be very close to normal, even when the sample size is as low as 4 or 5, in cases where the distribution of individuals is distinctly non-normal.

The normal distribution is a two-parameter distribution uniquely described by its mean and standard deviation. Consequently, its characteristics can be made available in a convenient, practical format for users. This will initially be illustrated generally in a graphical manner and, secondly, in the form of a table (Table 7). Figure 27 shows a standardized symmetrical bell-shaped curve that characterizes this distribution. Additionally, some key percentages are included in relation to distances from the mean in terms of standard deviations. Whilst, in Figure 27 and Figure 28, the normal curve appears to end at some finite value about ± 3 to 4 standard deviations from the mean, mathematically it extends to infinity in both directions.

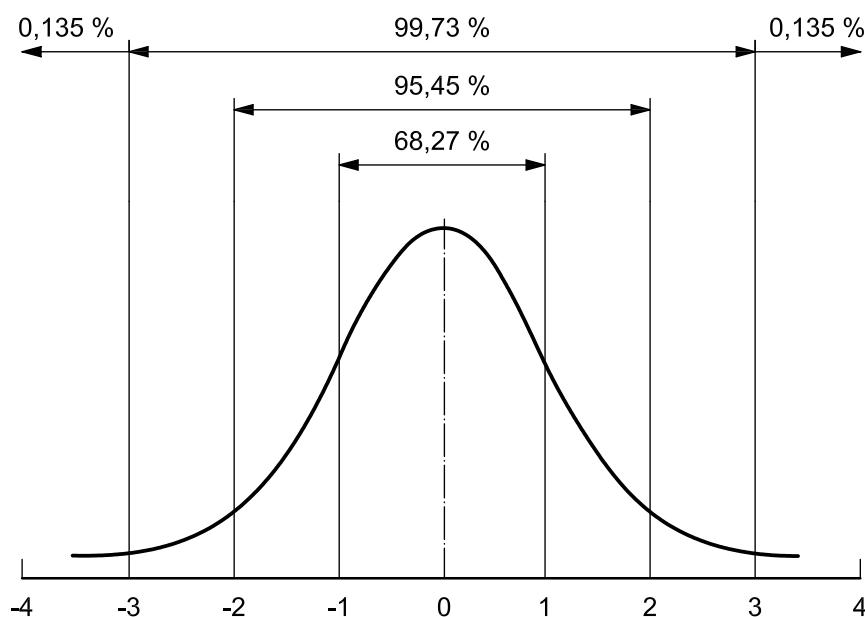


Figure 27 — Percentages of normal distribution in relation to distances from the mean in terms of standard deviations

Figure 27 illustrates that for a normal distribution

- a) 99,73 % of values lie within the limits: mean ± 3 standard deviations;
- b) of the remaining 0,27 %, 0,135 % lie below the mean -3 standard deviations and 0,135 % above the mean $+3$ standard deviations;
- c) 95,45 % of the values lie within the limits: mean ± 2 standard deviations;
- d) just over two thirds (68,27 %) of the values lie within the limits: mean ± 1 standard deviation.

This demonstrates a simple but useful property of the mean and standard deviation. Such a diagram shows the effectiveness of the normal distribution in predicting, from a sample, the proportion of the population lying within a specified range or above, or below particular limits. Whilst it is helpful in conveying certain principles, it is, however, not of sufficient resolution to be of real value in practice. Table 7 provides this by giving the probabilities of exceeding z in a standard normal distribution, i.e. a normal distribution with mean 0 and standard deviation 1. Because of the symmetry of the normal distribution, Table 7 also provides the probabilities of falling below $-z$ in a standard normal distribution. The probability of exceeding z in a standard normal distribution is the same as exceeding a point U that is z standard deviations above the mean in any normal distribution. Similarly the probability of falling below $-z$ in a standard normal distribution is the same as falling below a point L that is z standard deviations below the mean in any normal distribution. Those probabilities are illustrated in Figure 28 and they are also covered by Table 7.

NOTE 1 Example 2 following Table 7 shows how the percentage above or below a selected value can be determined when the mean and standard deviation are known. An alternative to the use of Table 7 is the application of the straight line probability plot shown in Figure 26 for making similar predictions.

EXAMPLE 1 Clothing size survey:

A size survey conducted on a random sample of a prospective customer base indicated that one particular characteristic, height, was normal with a mean equal to 175 cm, and a standard deviation equal to 7,5 cm.

By using the results of the survey, it was predicted, for instance, that:

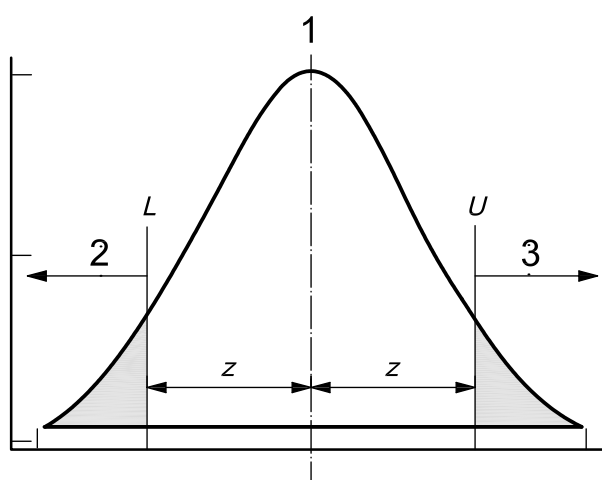
- 16 % (15,87 %) of the target customer population are taller than 182,5 cm (mean + 1 standard deviation);
- 25 % (25,14 %) of the target customer population are shorter than 170 cm (mean $-b \times$ standard deviation);

NOTE The b arises as 175 minus 170 expressed as a fraction of the standard deviation.

- 59 % (58,99 %) of the target customer population are between 170 cm and 182,5 cm.

Such estimates enable appropriate size ratios of garments to be ordered.

NOTE 2 Figure 27 and Figure 28 and Table 7 relate to a theoretical distribution representing a whole population. By convention, population parameters are symbolized by lower case italic Greek letters (e.g. population mean = μ and population standard deviation = σ).



Key

- | | | |
|------|-------------------|-------------------|
| 1 | 2 | 3 |
| mean | percent below L | percent above U |

Figure 28 — Standard normal probability density with indications of percentage expected beyond a value, U or L , that is z standard deviation units from the mean

In the real life examples shown, sample statistics are used to provide estimates. By convention, sample statistics are distinguished from population parameters by italic Roman letters (e.g. sample mean = $\bar{x}$ or $\bar{X}$ and sample standard deviation = s or S). Sometimes such sample statistics are limited to upper case to distinguish from their actual realization, which are then shown in lower case. In this standard, this latter distinction is not used because of common usage considerations in the application areas concerned.

Clause 8 deals with the statistical relationship between sample and population.

Table 7 — Probabilities (in percent for $z < 4$ and in parts per million for $z \geq 4,0$) of exceeding z in a standard normal distribution

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
6,0	0,001	0,000 9	0,000 9	0,000 8	0,000 8	0,000 7	0,000 7	0,000 6	0,000 6	0,000 6
5,0	0,286 7	0,272 2	0,258 4	0,245 2	0,232 8	0,220 9	0,209 6	0,198 9	0,188 7	0,179
4,0	31,67	30,36	29,10	27,89	26,73	25,61	24,54	23,51	22,52	21,57
NOTE For $z \geq 4,0$, values are given in parts per million.										
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
3,5	0,023 3	0,022 4	0,021 6	0,020 8	0,020 0	0,019 3	0,018 5	0,017 8	0,017 2	0,016 5
3,4	0,033 7	0,032 5	0,031 3	0,030 2	0,029 1	0,028 0	0,027 0	0,026 0	0,025 1	0,024 2
3,3	0,048 3	0,046 6	0,045 0	0,043 4	0,041 9	0,040 4	0,039 0	0,037 6	0,036 2	0,034 9
3,2	0,068 7	0,066 4	0,064 1	0,061 9	0,059 8	0,057 7	0,055 7	0,053 8	0,051 9	0,050 1
3,1	0,096 8	0,093 5	0,090 4	0,087 4	0,084 5	0,081 6	0,078 9	0,076 2	0,073 6	0,071 1
3,0	0,135 0	0,130 6	0,126 4	0,122 3	0,118 3	0,114 4	0,110 7	0,107 0	0,103 5	0,100 1
2,9	0,186 6	0,180 7	0,175 0	0,169 5	0,164 1	0,158 9	0,153 8	0,148 9	0,144 1	0,139 5
2,8	0,255 5	0,247 7	0,240 1	0,232 7	0,225 6	0,218 6	0,211 8	0,205 2	0,198 8	0,192 6
2,7	0,346 7	0,336 4	0,326 4	0,316 7	0,307 2	0,298 0	0,289 0	0,280 3	0,271 8	0,263 5
2,6	0,466 1	0,452 7	0,439 6	0,426 9	0,414 5	0,402 5	0,390 7	0,379 3	0,368 1	0,357 3
2,5	0,621 0	0,603 7	0,586 8	0,570 3	0,554 3	0,538 6	0,523 4	0,508 5	0,494 0	0,479 9
2,4	0,819 8	0,797 6	0,776 0	0,754 9	0,734 4	0,714 3	0,694 7	0,675 6	0,656 9	0,638 7
2,3	1,072	1,044	1,017	0,990 3	0,964 2	0,938 7	0,913 7	0,889 4	0,865 6	0,842 4
2,2	1,390	1,355	1,321	1,287	1,255	1,222	1,191	1,160	1,130	1,101
2,1	1,786	1,743	1,700	1,659	1,618	1,578	1,539	1,500	1,463	1,426
2,0	2,275	2,222	2,169	2,118	2,068	2,018	1,970	1,923	1,876	1,831
1,9	2,872	2,807	2,743	2,680	2,619	2,559	2,500	2,442	2,385	2,330
1,8	3,593	3,515	3,438	3,363	3,288	3,216	3,144	3,074	3,005	2,938
1,7	4,457	4,363	4,272	4,182	4,093	4,006	3,920	3,836	3,754	3,673
1,6	5,480	5,370	5,262	5,155	5,050	4,947	4,846	4,746	4,648	4,551
1,5	6,681	6,552	6,426	6,301	6,178	6,057	5,938	5,821	5,705	5,592
1,4	8,076	7,927	7,780	7,636	7,493	7,353	7,215	7,078	6,944	6,811
1,3	9,680	9,510	9,342	9,176	9,012	8,851	8,692	8,534	8,379	8,226
1,2	11,51	11,31	11,12	10,93	10,75	10,57	10,38	10,20	10,03	9,853
1,1	13,57	13,35	13,14	12,92	12,71	12,51	12,30	12,10	11,90	11,70
1,0	15,87	15,62	15,39	15,15	14,92	14,69	14,46	14,23	14,01	13,79
0,9	18,41	18,14	17,88	17,62	17,36	17,11	16,85	16,60	16,35	16,11
0,8	21,19	20,90	20,61	20,33	20,05	19,77	19,49	19,22	18,94	18,67
0,7	24,20	23,89	23,58	23,27	22,97	22,66	22,36	22,07	21,77	21,48
0,6	27,43	27,09	26,76	26,43	26,11	25,78	25,46	25,14	24,83	24,51
0,5	30,85	30,50	30,15	29,81	29,46	29,12	28,77	28,43	28,10	27,76
0,4	34,46	34,09	33,72	33,36	33,00	32,64	32,28	31,92	31,56	31,21
0,3	38,21	37,83	37,45	37,07	36,69	36,32	35,94	35,57	35,20	34,83
0,2	42,07	41,68	41,29	40,90	40,52	40,13	39,74	39,36	38,97	38,59
0,1	46,02	45,62	45,22	44,83	44,43	44,04	43,64	43,25	42,86	42,47
0,0	50,00	49,60	49,20	48,80	48,40	48,01	47,61	47,21	46,81	46,41

EXAMPLE 2 Use of Table 7:Specified tolerance = 42 ± 4

Mean = 40

Standard deviation = 2,2

Process is in statistical control with an output that is normal.

What percentage is expected outside the specification limits?

To find the percentage above the upper specification limit:

$$Z_{\text{upper}} = \frac{\text{upper specification limit (U)} - \text{mean}}{\text{standard deviation}} = \frac{46 - 40}{2,2} = 2,73$$

Enter Table 7 at 2,73 (2,7 from the left and 0,03 from the top as indicated by the arrows) to give 0,32 % above the upper specification limit.

To find the percentage below the lower specification limit:

$$Z_{\text{lower}} = \frac{\text{mean} - \text{lower specification limit (L)}}{\text{standard deviation}} = \frac{40 - 38}{2,2} = 0,91$$

Enter Table 7 at 0,91 (0,9 from left and 0,01 from the top) to give 18,41 % below the lower specification limit.

Hence, the expected total fraction nonconforming is 0,32 % + 18,41 % = 18,73 %.

5.3.9 The Weibull distribution**5.3.9.1 General**

Most of the earlier discussion has been based on the assumption that the population or populations under consideration are normally distributed, at least approximately. This assumption is found in practice to be valid for a very wide range of situations. However, it is not appropriate for distributions that are typically skewed, and the Weibull distribution provides a better approximation to the kind of skewed distributions often arising in time-to-failure or breaking-strength data. For the purposes of discussion, we shall suppose that the characteristic in question is a failure time, and denote it by t . Here we shall briefly consider the simplest form of the Weibull distribution, with two parameters α and β where α controls the scale and β the shape. In the three-parameter Weibull distribution, the third parameter is the threshold which is the smallest observation possible. The smallest observation possible is 0 for the two-parameter Weibull distribution. The three-parameter Weibull distribution is briefly mentioned in 5.3.9.5.

The probability density of the two-parameter Weibull distribution is as follows:

$$f(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha} \right)^{\beta-1} e^{-\left(\frac{t}{\alpha} \right)^{\beta}}, \text{ for } t > 0 \quad (1)$$

For $\beta = 1$, the expression [Equation (1)] simplifies to

$$f(t) = \frac{1}{\alpha} e^{-\frac{t}{\alpha}} \text{ for } t > 0 \quad (2)$$

which is the density of the exponential distribution with mean equal to α . Figure 29 shows the way the Weibull density function changes shape for the case $\alpha = 1$ as β increases from $\frac{1}{2}$ to 4. Four different situations can be distinguished and they are described as follows:

- 1) $\beta < 1$ represents a range of hyper-exponential distributions;
- 2) $\beta = 1$ represents an exponential distribution;
- 3) $1 < \beta < 3,5$ represents a range of skew distributions with the skewness decreasing as beta increases until at about 3,5 the distribution looks roughly symmetrically normal;
- 4) $\beta > 3,5$ represents a distribution that stays largely symmetrical (slight skewness) and becomes progressively more peaky as beta increases.

Increasing or decreasing α has the effect of simply stretching or compressing the horizontal scale.

The probability that the failure time is less than t is given by Equation (3):

$$F(t) = 1 - e^{-\left(\frac{t}{\alpha}\right)^\beta} \quad \text{for } t > 0 \quad (3)$$

The *reliability function*, sometimes called the *survival function*, $R(t)$, is the probability that an item is still functioning at time t , so it is the complement of $F(t)$, i.e.:

$$R(t) = 1 - F(t) = e^{-\left(\frac{t}{\alpha}\right)^\beta} \quad \text{for } t > 0 \quad (4)$$

It is often impracticable to continue a trial until all the members of the sample reach the end of their lives. For example, twenty light bulbs may be switched on and left burning in order to provide information about their lifetime distribution. To prevent the trial going on indefinitely, a time limit may be set, say at 1 500 h, at which time the trial will be stopped. Alternatively, it may be decided in advance that the trial will be stopped when a specified number, say 15, of the light bulbs have burnt out. Both lead to what is called censored data, the former with respect to time and the latter with respect to numbers of failures. For small samples, it can be important when estimating parameters and calculating confidence intervals to take account of which type of censoring was used.

5.3.9.2 The failure rates of the Weibull distributions

Because of the application of the Weibull distribution to reliability or survival data, it is convenient to consider the (*instantaneous*) *failure rate* (also known as the hazard rate) of the Weibull distribution. It is the ratio of the density function [Equation (1)] to the reliability function [Equation (4)], so it is

$$r(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} \quad \text{for } t > 0 \quad (5)$$

The probability that an item will fail in the small interval between time t and $t + dt$ given that it has been functioning up to time t is given as $r(t)dt$, where dt is the length of the short interval.

The three possible failure regimes are often represented by the generic *bathtub* curve, that of infant mortality (decreasing failure rate), prime of life (constant failure rate) and wear-out (increasing failure rate).

The Weibull β parameter distinguishes between these regimes, thus:

- a) $\beta < 1$ represents a decreasing failure rate regime (colloquially called infant mortality);
- b) $\beta = 1$ represents a constant failure rate regime (often called prime of life);
- c) $\beta > 1$ represents an increasing failure rate regime (frequently termed wear-out).

So the failure rate of a Weibull distribution does not look like a bathtub curve. It is either decreasing, constant, or increasing.

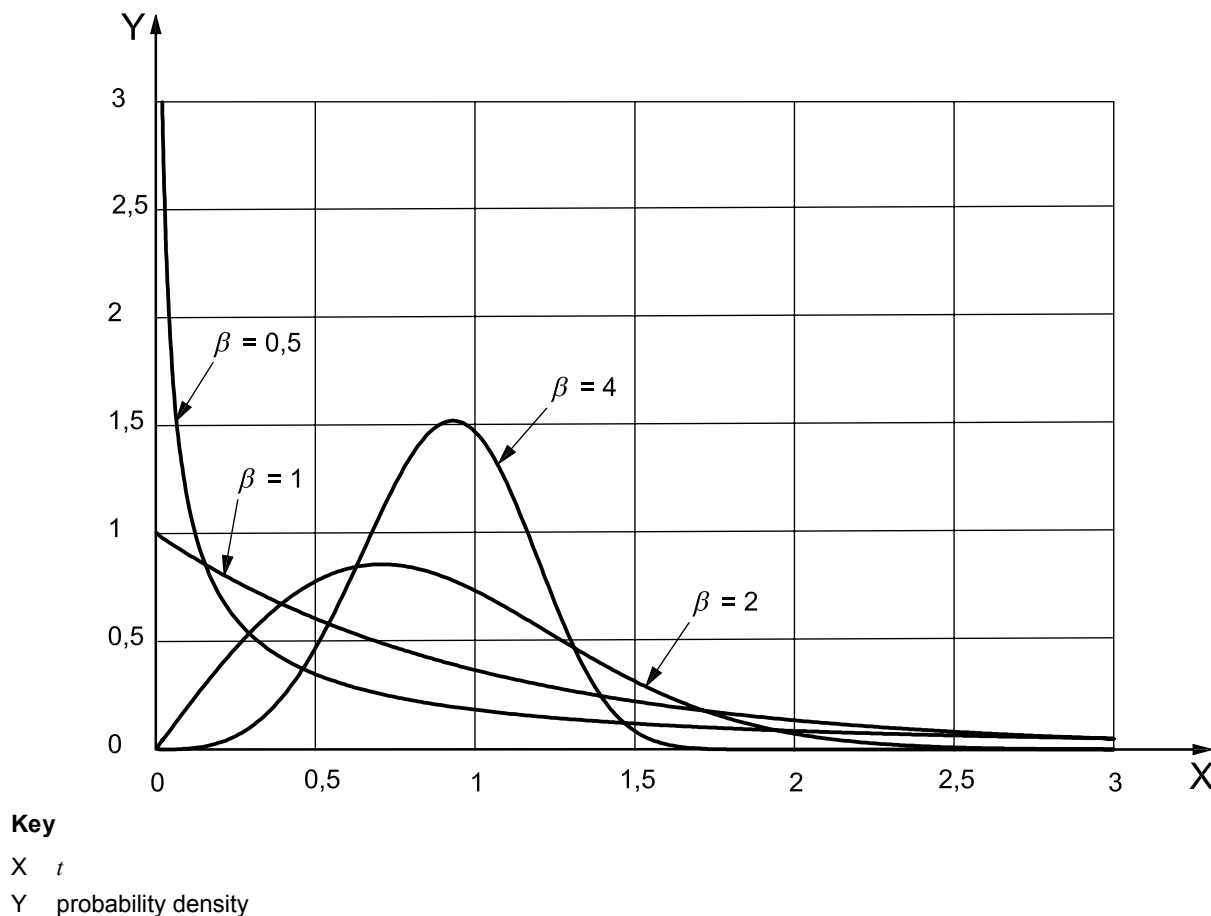


Figure 29 — Comparison with Weibull distributions, all with $\alpha = 1$

5.3.9.3 Does the Weibull distribution fit the data?

Although it is somewhat subjective, the easiest way to check if the two-parameter Weibull distribution will provide a reasonable fit to a given set of data is to use a graphical method. It is based on the fact, which follows from the expression for the reliability function [Equation (4)], that:

$$\ln[-\ln(1 - F(t))] = \ln\left[\left(\frac{t}{\alpha}\right)^\beta\right] = \beta \ln(t) - \beta \ln(\alpha) \quad (6)$$

which is a straight line relationship between $\ln[-\ln(1 - F(t))]$ and $\ln(t)$.

The plotting procedure is as follows. The n sample values are first arranged in ascending order to give the *order statistics* $t_{[1]}, t_{[2]}, \dots, t_{[n]}$, i.e. such that $t_{[1]} \leq t_{[2]} \leq \dots \leq t_{[n]}$. For each $t_{[i]}$, the value of $F(t_{[i]})$ is estimated by $(i - 0,5)/n$, and the point with co-ordinates $(\ln(t_{[i]}), \ln\{-\ln[1 - (i - 0,5)/n]\})$ is plotted. If the data is a sample from a Weibull distribution, the plotted points will lie approximately on a straight line. The reason that the plotted points will only approximately lie on a straight line even if the data is a sample from the Weibull distribution is that $(i - 0,5)/n$ is just an estimate of $F(t_{[i]})$, and all estimates are subject to random variation. So this graphical method consists of making the plot as described and deciding whether or not the points lie approximately on a straight line. It is the last part that is described as somewhat subjective in the beginning of this clause.

The graphical method is illustrated with the data in the example in 5.3.9.4, and the plot is shown in Figure 30.

5.3.9.4 Example: Days between accidents in a company

The following data are the times in days between accidents in a large industrial plant. The data are recorded in days except for one case, where two accidents occurred within a few hours on the same day. The data are given ascending order in column 1 of Table 8. The calculations needed to make the plot to check if the data can be described by a Weibull distribution are given in columns 2 to 5. Column 2 contains the natural logarithm of the observations, which will be plotted along the first axis. The remaining three columns go through the calculations needed to find the quantity that will be plotted along the second axis. First, the observations are numbered from 1 to n in column 3. This number is used to calculate $(i - 0,5)/n$ which is shown in column 4. Finally, the logarithm of minus the logarithm of the values of column 4 is shown in the fifth column. The Weibull plot of Figure 30 is obtained by plotting the second column against the fifth column.

The points lie sufficiently close to a straight line, so the plot confirms that the data can be described by a Weibull distribution. The full line on the plot is a line fitted by eye, or rather by least squares. The dashed line corresponds to the Weibull distribution fitted by the maximum likelihood method.

Table 8 — Days between accidents in the first column sorted in ascending order

Days between accidents in ascending order	Logarithm (ln) of days between accidents	Number of observation after ordering i	$(i - 0,5)/n$	$\ln\{-\ln[1-(i-0,5)/n]\}$
0,2	-1,61	1	0,022	-3,81
1	0,00	2	0,065	-2,70
2	0,69	3	0,109	-2,16
4	1,39	4	0,152	-1,80
5	1,61	5	0,196	-1,52
5	1,61	6	0,239	-1,30
9	2,20	7	0,283	-1,10
10	2,30	8	0,326	-0,93
10	2,30	9	0,370	-0,77
10	2,30	10	0,413	-0,63
16	2,77	11	0,457	-0,49
22	3,09	12	0,500	-0,37
24	3,18	13	0,543	-0,24
26	3,26	14	0,587	-0,12
26	3,26	15	0,630	-0,01
28	3,22	16	0,674	0,11
30	3,40	17	0,717	0,23
32	3,47	18	0,761	0,36
50	3,91	19	0,804	0,49
64	4,16	20	0,848	0,63
91	4,51	21	0,891	0,80
94	4,54	22	0,935	1,01
150	5,01	23	0,978	1,34

NOTE The second column is plotted against the fifth column to produce the Weibull plot of Figure 30.