BRITISH STANDARD

BS 600:2000

A guide to the application of statistical methods to quality and standardization

ICS 03.120.30

Committees responsible for this British Standard

The preparation of this British Standard was entrusted by Technical Committee SS/3, Application of statistical methods in standardization, upon which the following bodies were represented:

Institute of Quality Assurance
Laboratory of the Government Chemist
National Physical Laboratory
Royal Statistical Society

This British Standard, having been prepared under the direction of the Management Systems Sector Committee, was published under the authority of the Standards Committee and comes into effect on 15 October 2000

Amendments issued since publication

Amd. No.	Date	Comments

The following BSI references relate to the work on this standard:
Committee reference SS/3
Draft for comment 99/404263 DC

ISBN 0 580 33227 6

Contents

Page

Foreword

This British Standard is published under the authority of BSI Technical Committee SS/3, Application of statistical methods in standardization. It supersedes BS 600:1935 which is withdrawn.

BS 600 is intended primarily to bring about a greater acceptance by non-statisticians of the value of statistical methods in business. It is in order to gain the interest of this wide field that the form of presentation in BS 600 has been chosen. No attempt has been made to provide a comprehensive introduction to the methods of statistical analysis; rather the aim is to demonstrate the necessity and usefulness of such methods not only in standardization but also in the continuous improvement of business processes.

BS 600 was written by Dr. Egon S. Pearson and first published in 1935. Amendments were incorporated in August 1957 and October 1960. This revision has been prepared to reflect the significant developments that have taken place since then in the application of statistical methods in standardization and in the achievement, control and improvement of quality in design, development, planning, manufacture and service processes.

The original BS 600 focused on repetition and routine production work in manufacturing. This BS 600 recognizes that the present day focus is on the use of statistical methods as far "up-stream" as possible. Application of statistical methods, at the marketing, design, development and advanced planning stage, before going into production or delivering the service, is often considered more rewarding from a business viewpoint, by exploiting the so called "quality lever". The earlier a concern is dealt with in the marketing and development cycle the greater the potential reward.

Hence the statistical methods covered in this standard extend beyond the purely manufacturing area to product, system, material and process design quality, robustness and proving; and business process control, performance measurement and continuous improvement. These methodologies are also applicable to administrative areas and to all sectors including commerce and public service (e.g. transport and health care).

It is the belief of the committee that the simple treatment of the elementary techniques of statistical analysis provided cannot fail to change the mindset and convince those concerned, technically and managerially, with product, process and system design and development, with the procurement of materials, with manufacture and with the provision of services, that a valuable set of tools is within the reach of all.

A British Standard does not purport to include all the necessary provisions of a contract. Users of British Standards are responsible for their correct application.

Compliance with a British Standard does not of itself confer immunity from legal obligations.

Summary of pages

This document comprises a front cover, an inside front cover, pages i to x, pages 1 to 162, an inside back cover and a back cover.

The BSI copyright notice displayed in this document indicates when the document was last issued.

Introduction

BS 600 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 sub-groups" coupled with the use of cause and effect diagrams and line plots. It also shows how to exploit inter-relationships between process parameters to achieve "robust" designs. The need to treat numerical data, not just as a set of figures but as potentially meaningful information on a process, is emphasised. 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 possible to any organization.

Example 2, on weight of fabric, 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 which are turned to practical use.

Example 3 concerns the percentage ash content of 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 which can most readily be communicated to, and be readily 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 which 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: one, is the process in control?; two, what is the performance of the process?; and three, 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, Cp, Cpk_U and Cpk_L. It also discusses the business implications, in terms of aiming at preferred value and minimizing variation, with the quotation of minimum Cp_m 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.

1 Scope

BS 600 describes a broad range of statistical methods applicable to the management, control and improvement of processes.

2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this British Standard. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. For undated references, the latest edition of the publication referred to applies.

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms.*

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

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

ISO 9000, *Quality management systems — Fundamentals and vocabulary.*

3 Terms and definitions

For the purposes of this British Standard the terms and definitions given in ISO 3534-1, ISO 3534-2, ISO 3534-3 and ISO 9000 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 is this “statistical method” that may gainfully be applied in the field of quality and standardization. It is important to give some preliminary answer 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 *line 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 the 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 minimum specified value*

Suppose 64 test results were obtained on the breaking strength of wire where the minimum specified value is 420 units. The results are shown in ascending order in Table 1 and as a line 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	435	460	480	500	515	540	560
400	440	460	480	500	520	540	565
405	440	460	480	500	520	545	570
410	445	465	485	505	520	545	575
415	450	470	490	510	520	550	575
415	450	470	490	510	530	550	580
420	450	475	495	515	530	550	585
430	455	475	495	515	535	560	590

Mean = 495.

Minimum = 390.

Maximum = 590.

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 and median values are 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 *line plot* of Figure 1. (For the corresponding *dot plot* see Figure 12.)

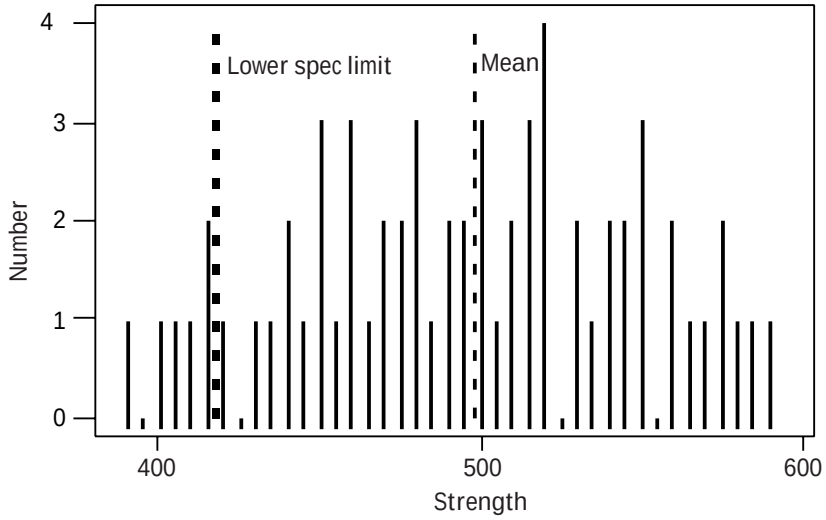


Figure 1 — Line plot of breaking strength of 64 test specimens

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 always to 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 to both 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* as shown in Figure 2 which suggests a dependence of the wire strength on material composition and levels of steel and oil quench temperatures.

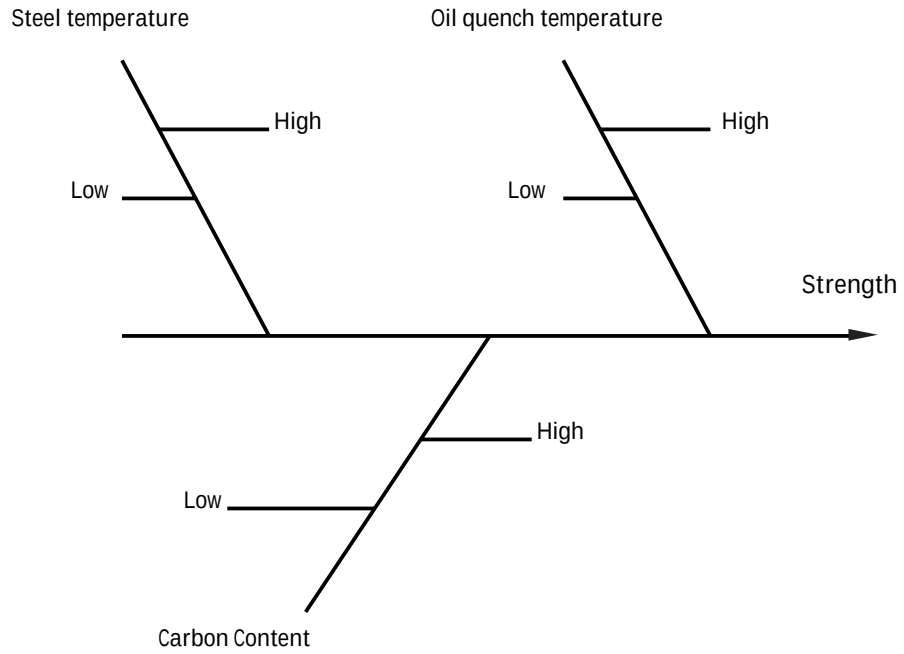


Figure 2 — Basic cause and effect diagram for variation in wire strength (due to possible changes of material and process parameters within specified tolerances)

The next stage involved the division of the test records into a number of groups, within each of which all or some of these possible factors were roughly constant. This grouping, which is essential in any process of analysis, is described as division into *rational subgroups*. 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 *line plots* of Figure 3.

Figure 3a)
Low oil quench temperature.
High steel temperature.

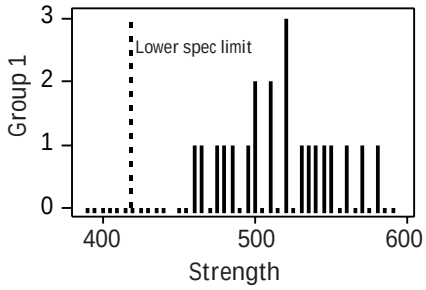


Figure 3b)
High oil quench temperature.*
High steel temperature.
* with inadequate control

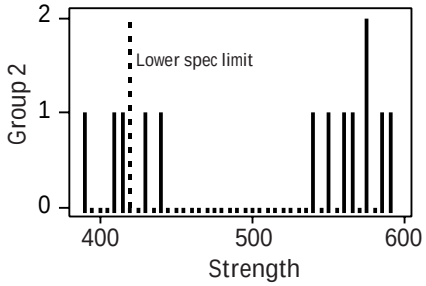


Figure 3c)
Low oil quench temperature.
Low steel temperature.

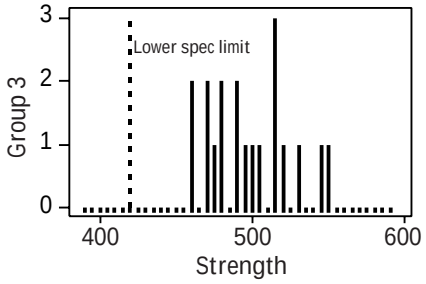


Figure 3d)
High oil quench temperature.
Low steel temperature.

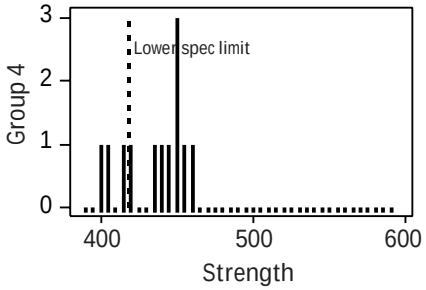


Figure 3 — Line plots showing pattern of results after division into rational subgroups

The results indicate that:

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 pre-set 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 is shown diagrammatically in Figure 4.

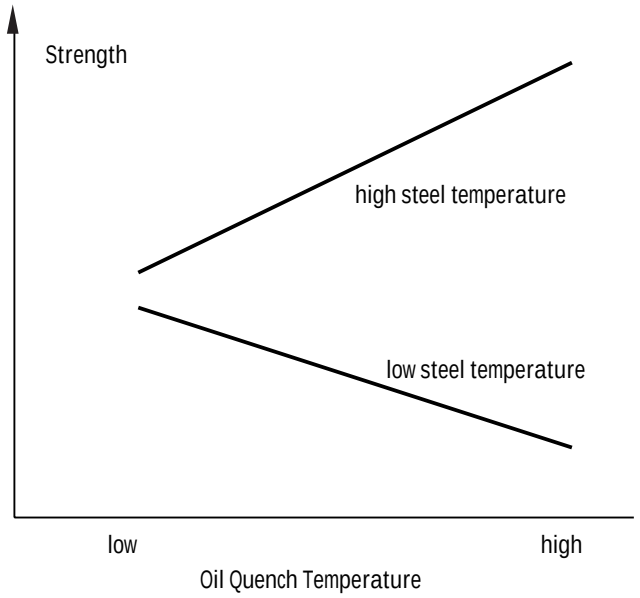


Figure 4 — Diagram indicating the effect of the inter-relationship between oil quench temperature and steel temperature on wire strength

Now is decision time. How should this process be run to ensure uniform strengths of wire which do not contravene the lower specification limit?

There is a clear choice. This 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 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.

4.2.6 *Explanation of statistical terms and tools used in this example*

Rational subgroup: is one 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 more readily discriminate 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 to 0.2 mm per metre and a sand casting to 0.8 mm per metre. Hence common cause variation can often be reduced only by people responsible *for* the system. 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 their number.

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).

Cause and effect diagram: frequently called a fishbone diagram (because of its shape) or an Ishikawa diagram (after its creator). 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. Sub-categories are indicated on sub-branches.

4.3 Example 2: Weight 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 which is of importance, but also the uniformity as measured by the variation about that average.

It follows that, in checking for conformity to a specification that is based on a series of sets of measurements, it is essential 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 on the data given in Table 2. The figures represent the weights of standard specimens taken from a roll of fabric. They have been grouped for purposes of illustration into 32 samples each of 4 specimens.

Table 2 — Weights of 128 standard specimens from a roll of fabric — Minimum specification limit = 98

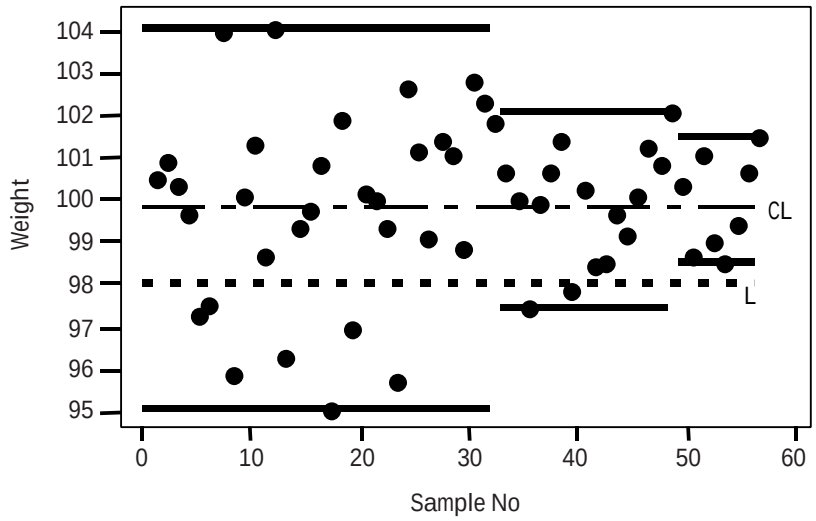
Unit of measurement = 0.1 g															
No.	**Weight**	**No.**	**Weight**	**No.**	**Weight**	**No.**	**Weight**	**No.**	**Weight**	**No.**	**Weight**	**No.**	**Weight**	**No.**	**Weight**
1	101	5	96	9	104	13	95	17	100	21	100	25	102	29	100
	99		97		102		94		97		101		100		97
	100		100		95		97		91		95		105		100
	102		96		100		100		92		103		98		98
2	106	6	101	10	98	14	102	18	106	22	101	26	99	30	104
	98		96		101		100		100		99		98		103
	101		97		99		100		102		99		103		104
	99		97		107		95		100		99		97		100
3	98	7	109	11	99	15	97	19	97	23	94	27	97	31	105
	101		100		98		101		97		96		98		99
	102		106		99		102		94		94		106		103
	100		101		99		98		99		98		104		103
4	103	8	92	12	109	16	103	20	99	24	99	28	97	32	98
	104		97		101		101		101		100		101		104
	95		100		105		99		100		104		108		102
	96		95		102		100		101		108		99		103

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

a) the first 32 samples each have 4 weights (as shown in Table 2);

b) samples 33 to 48 relate to the same data in Table 2 which has now been divided into 16 samples each of 8 weights; and

c) samples 49 to 56 relate to the same data in Table 2 which has now been divided into 8 samples each of 16 weights.



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

NOTE 2 L = lower specification limit.

NOTE 3 CL = centreline.

Figure 5 — Plot of means of weights v sample number (illustrating decreasing variation in the mean with sample size increase)

Figure 6 illustrates the data of Table 2, in 3 ways, in relation to sample ranges. As in Figure 5:

1) the first 32 samples each have 4 weights (as shown in Table 2);

2) samples 33 to 48 relate to the same data in Table 2 which has now been divided into 16 samples each of 8 weights; and

3) samples 49 to 56 relate to the same data in Table 2 which has now been divided into 8 samples each of 16 weights.

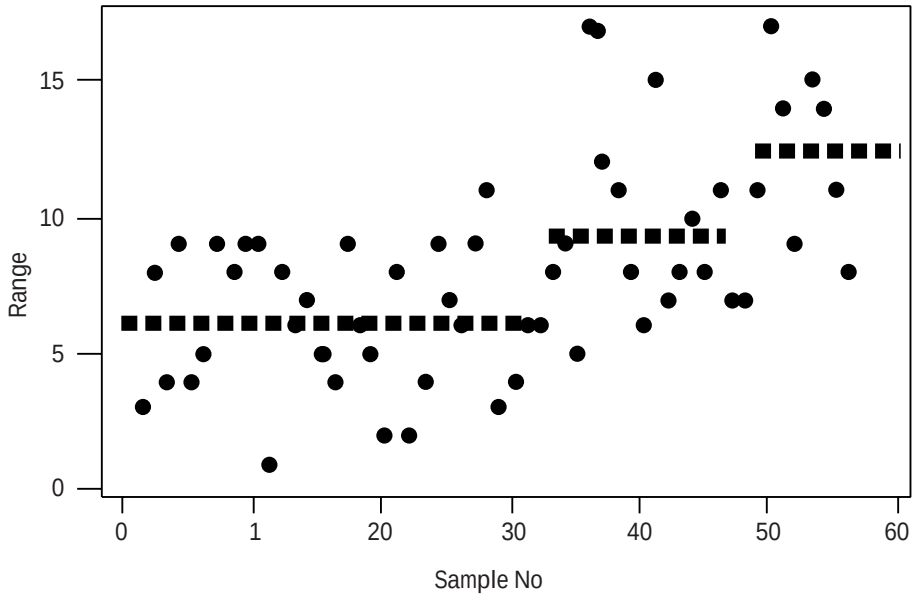


Figure 6 — Plot of ranges of weights within each sample v sample number [illustrating increasing (range) variation within a sample with sample size increase]

4.3.3 *Discussion of specific results*

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

a) The mean (Figure 5).

Sample size	Range of means
4	95 to 104 = 9
8	97 to 102 = 7
16	99 to 101 = 2

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

b) The range (Figure 6).

Sample size	Average range
4	6
8	10
16	12

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

c) Conformance to specification.

Suppose that the minimum weight, 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) 6 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.

4.3.4 *Discussion on general findings*

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

a) The larger the number of observations or tests, the smaller the variation between the mean of one set of tests and that of another.

b) The larger the number of observations or tests, the greater the range of variation to be expected among them.

c) A statement by way of specification that "the lower specification limit = 98", say, is inadequate unless supplemented by:

1) information as to the number of specimens to be tested;

2) whether the limit refers to the mean or to the minimum value of an individual specimen.

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

4.4 Example 3: Percentage ash content of cargo of coal

4.4.1 *General*

This example illustrates four principal concepts:

— the handling of apparent fluctuation of quality 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 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.);

b) sampling to make an estimation of the average quality of a particular characteristic of the bulk commodity. (See ISO 11648.)

Illustrations of the application of **4.4.1**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 percentage ash content 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 of the bulk of the fuel as a basis for determining the price to pay for the consignment.

4.4.2 *Test results (ref. ISO 11648: Sampling from bulk materials)*

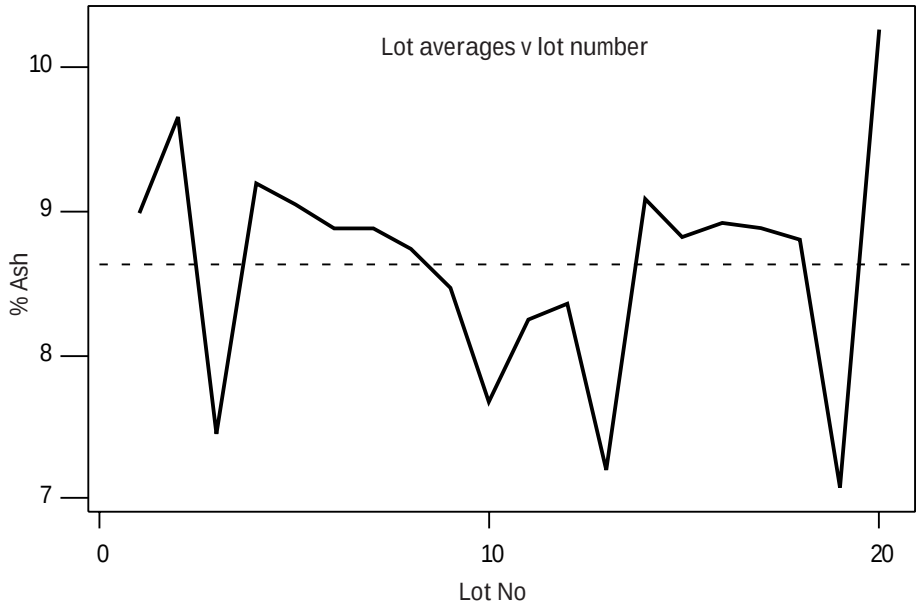
Table 3 shows the test results from a series of 20 lots of coal being unloaded from a ship. For each lot, eight samples of coal were drawn and the percentage ash content measured.

Table 3 — Percentage ash content 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*

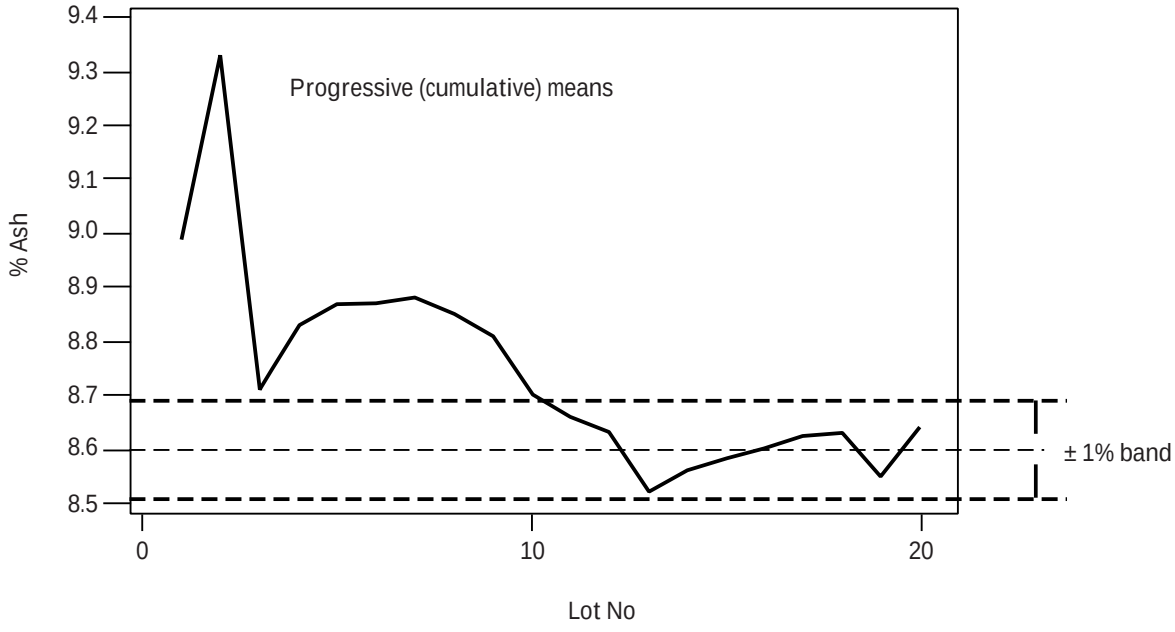
A plot of the averages of percentage ash content of the coal by lot is shown in Figure 7.



NOTE The value of percentage ash content in the figure for each lot represents the average of 8 results for that 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.

Figure 7 — Plot of averages of percentage ash content of coal by lot from cargo

Considerable fluctuation is observed about the overall average of 8.63 % ash content. Suppose 8.6 % is taken to represent the true measure of 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 is best answered by reference to the progressive average plot in Figure 8.



NOTE The value of percentage ash content 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.

Figure 8 — Plot of progressive averages of percentage ash content in terms of lot

Figure 8 shows that, as the number of sampling lots increases, the progressive average approaches the limiting value of 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 true 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 weight 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 called for.

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 considerable variation within each column of results;

b) the rows of results for the main peaks and troughs of Figure 7, lots 2 (high), 3, 13 and 19, (low), and 20 (high), are quite consistent within each lot;

c) there are adjacent column pairs of very low values in lot 9 (7.10 and 7.22 in columns 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 in columns 5 and 6 respectively).

What does this really mean?

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.

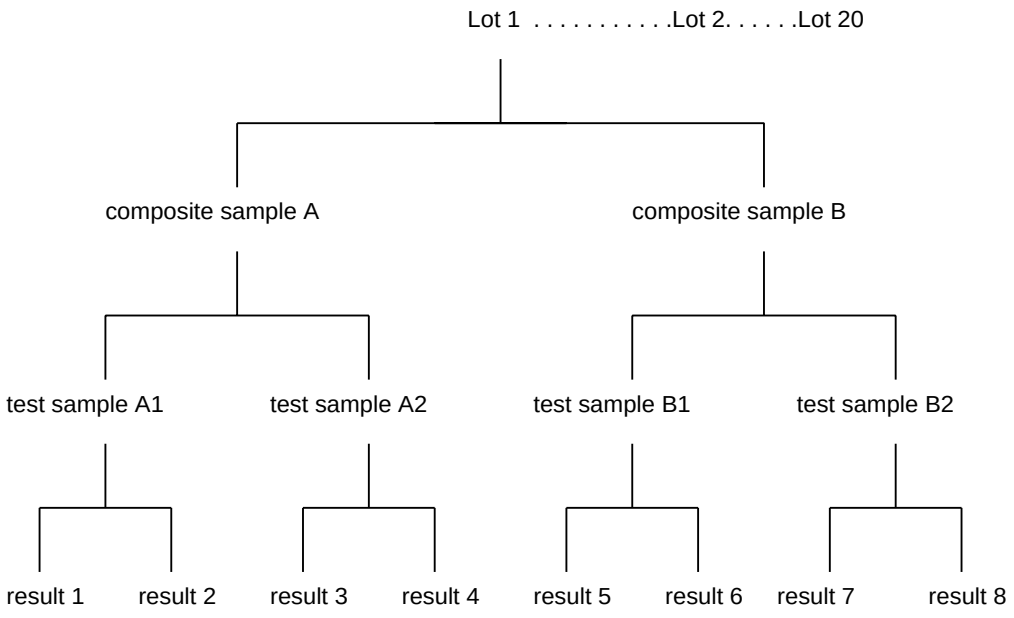


Figure 9 — Schematic diagram showing plan for sampling percentage ash from cargo of ship

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 (ref. ISO 11648) when there is no, or little, prior knowledge about the sampling situation.

In this particular case, the belt conveyor unloading coal from the ship was stopped at uniform time intervals. A pre-specified weight increment of coal was taken from the conveyor belt, using a shovel. Individual consecutive increments were placed alternatively 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. This gives rise to the 8 results shown for each lot. 20 lots were specified here.

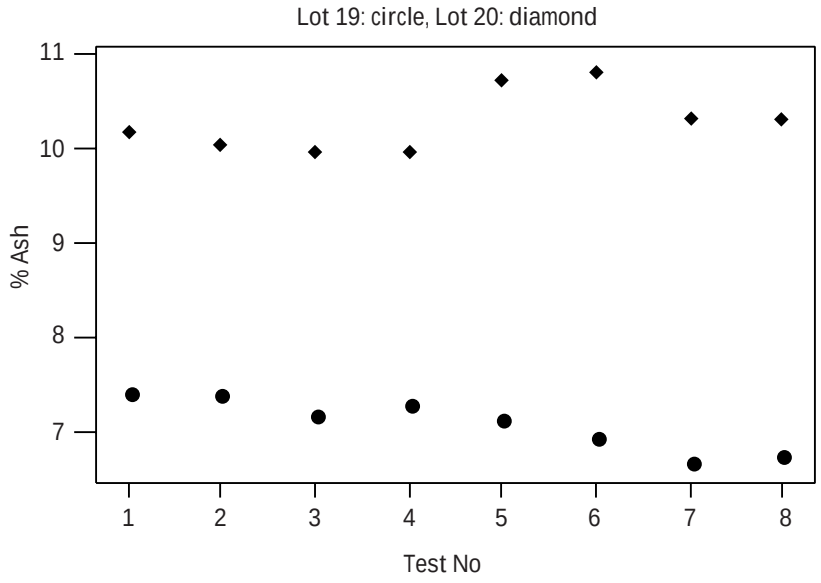


Figure 10 — Plot of percentage ash v test number for lots 19 and 20 (illustrating relative consistency of percentage ash within each of these lots)

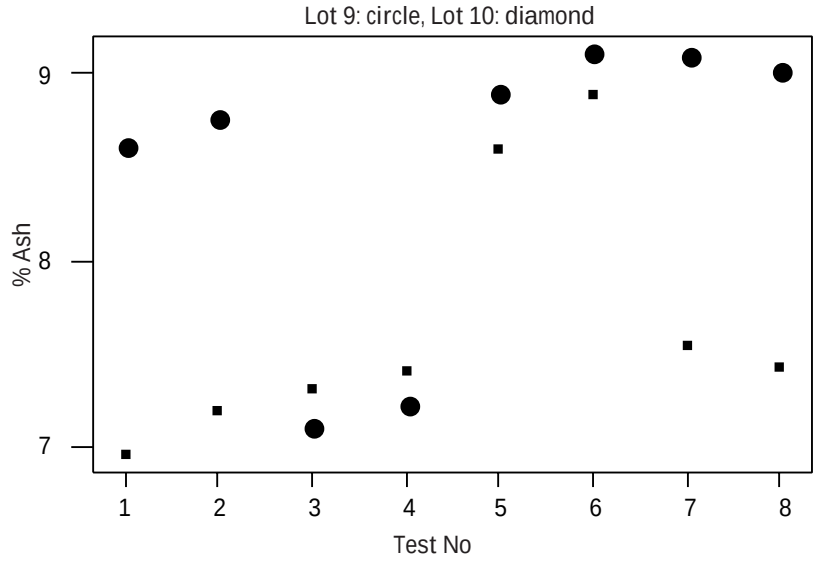


Figure 11 — Plot of percentage ash v test number for lots 9 and 10 (illustrating rogue pairs in both lots)

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

a) in general, there is far more lot to lot variation (row to row variation in Table 3) than within lot variation (within row variation);

b) Figure 10 confirms, with respect to lots 19 and 20, the relative consistency of test results within these lots;

c) Figure 11 indicates two pairs of rogue values in lots 9 and 10. Reference to the sampling plan indicates that the two rogue pairs are associated with test sample A_2 in lot 9 and B_1 in lot 10. This 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 specifically assign this special cause. 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 assigned, at the time and place of the sampling activity. Action could then be taken, by operational or technical people, 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 it 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 persons who are technically and operationally familiar with the activity under scrutiny are involved in the progressive statistical analysis. This will facilitate the early assignment 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 percentage ash content (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 the simpler 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 rather more fully some of the simpler 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; times to answer a telephone; times to failure; noise levels; emission levels of engines. This partial listing gives some impression of the universal 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*. 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 sections that follow it will be necessary to discuss various aspects of the relationship between the sample and the population. But it is first necessary to introduce certain statistical measures which, from the descriptive 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, which is frequently, and unfortunately, not taken into consideration in the application of these measures, is the *pattern of variation*. For example, too often normality "symmetrical bell shaped distributions" are 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 is ranked in order of size;

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 this variation the smaller the value. When derived from a sample, the value is given by the expressions:

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

where

$\sum$ is the sum of;

X is the individual value;

$\bar{X}$ is the arithmetic mean;

n is the number of values;

s is the 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	Affected by very high or low values
	Commonly used	Need all the data to calculate
Median	Unchanged by very high or low values	Slow and tedious to calculate
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.

Arithmetic mean = (7 + 5 + 10 + 7 + 6)/5 = 7

Median = central value of ordered set, 5, 6, 7, 7, 10 = 7

Mode = most frequent value = 7

Range = maximum value − minimum value = 9 − 5 = 4

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 that shown in Figures 1 and 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 dot plot for the line plot of Figure 1 is shown in Figure 12.

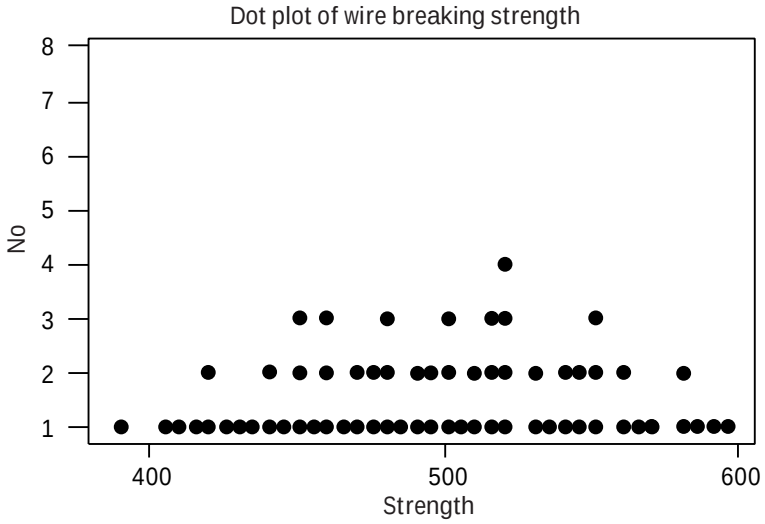


Figure 12 — Dot plot of breaking strength of wire (Table 1 data)

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.

Figure 13a) Tally chart for measurements

21	𝍸
22	𝍸
23	𝍸 ||
24	𝍸 𝍸
25	𝍸 ||
26	𝍸

Figure 13b) Tally chart for events/counts

open circuit	||
short circuit	𝍸
dry joint	𝍸 𝍸
solder splash	𝍸
wrong component	𝍸 𝍸
broken lead	𝍸

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. It consists essentially of 2 parts:

a) the first column holds the stem or leading digit(s);

b) the second column holds the leaf or following digit(s).

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 (also called box and whisker 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. 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 maximum and minimum values. It shows a number of key statistical measures in graphical scaled format, such as:

M: median (mid) value;

Q1: first quartile (value below which ¼ of values lie);

Q3: third quartile (value above which ¼ of values lie);

min: minimum value in the data;

max: maximum value in the data.

It is shown pictorially as follows.

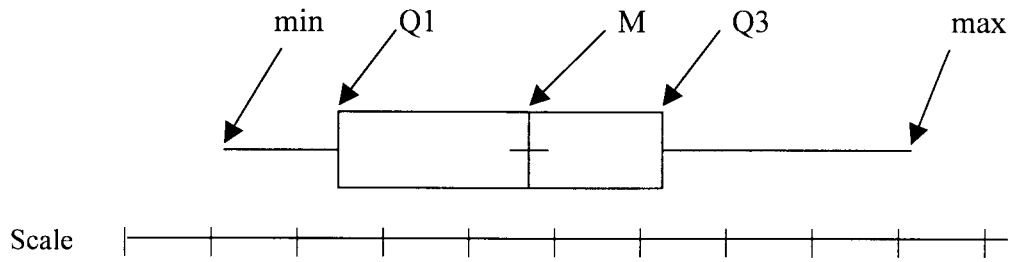


Figure 15 — Basic box plot

The box plot may be extended, for example, to include a display of statistical confidence limits around the median. An absence of overlapping 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.

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. Shade variation of fabric of a particular colour way was compared between adjacent panels on representative items of clothing sourced from 3 different suppliers. The results are shown in box plot form in Figure 16.

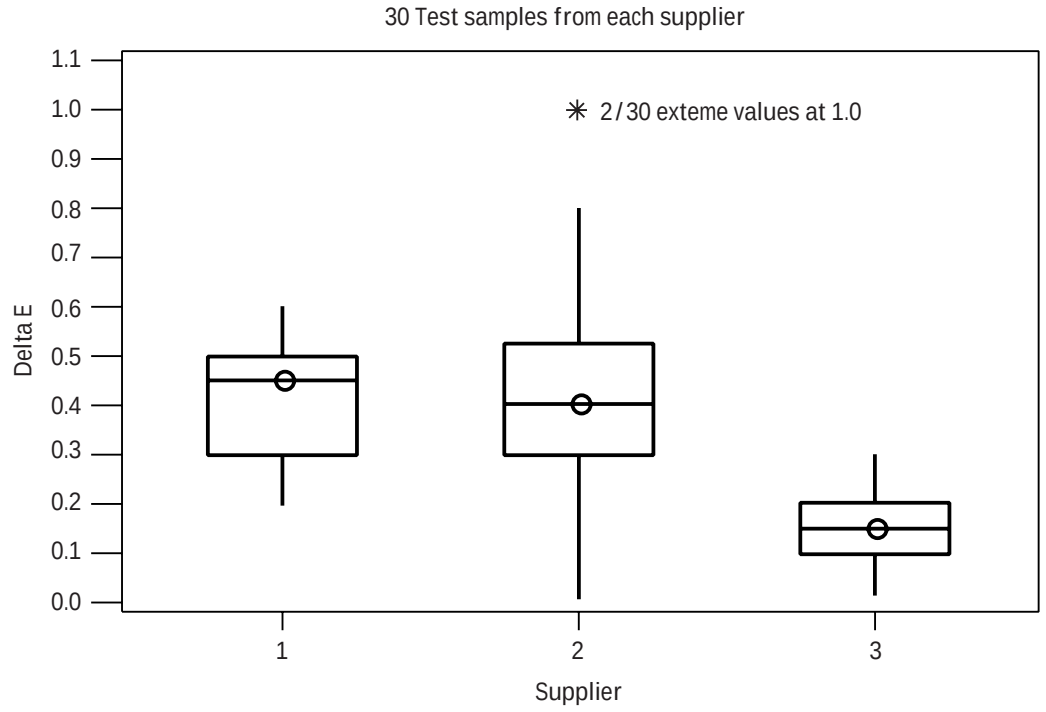


Figure 16 — Box plot for Delta E panel shade variation between supply sources

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 "*" shows two outlying values indicating lack of control of the dyeing process.

The dot plot indicates that:

a) supplier 3 has a dyeing process targeted on a low (good) Delta E value with small variation about that value;

b) supplier 2 has a dyeing process targeted on a higher (worse) Delta E value with greater variation around it. 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 representative of production major recalls may be expected;

c) supplier 1 is targeted slightly higher than supplier 2 but less variation is evident.

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 3 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.

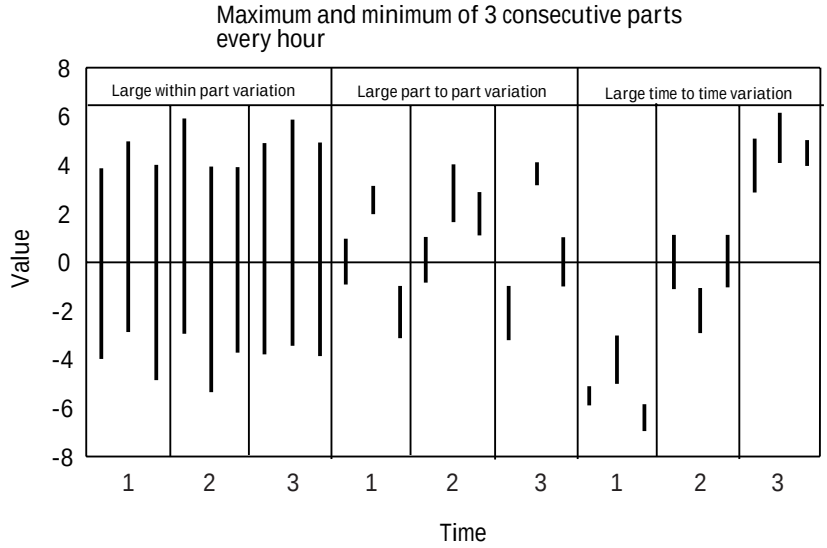


Figure 17 — Multi-vari chart as a tool for process variation analysis

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 18a). These positional measurement values were identified by A, B, C and D as shown in Figure 18a). One cylinder from production was measured every shift for 4 shifts. The machine tool was then overhauled and another set of readings taken.

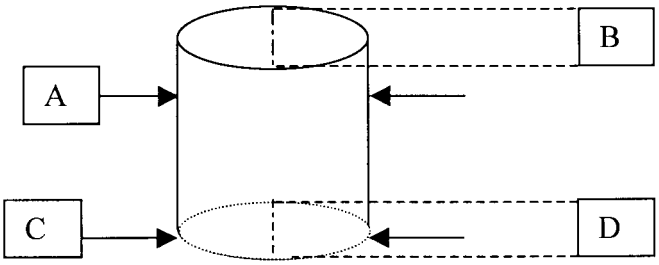
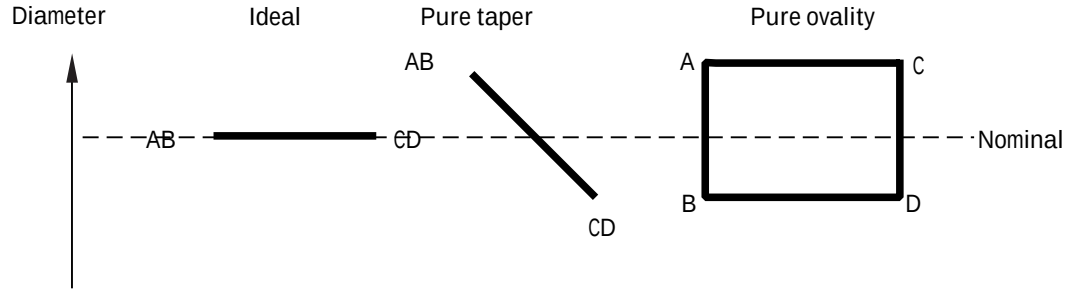


Figure 18a) — Measurements on cylinder to determine nominal size, ovality and taper

Figure 18b) 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.



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

Figure 18b) — P-D diagrams showing ideal diameter values, pure taper and pure ovality

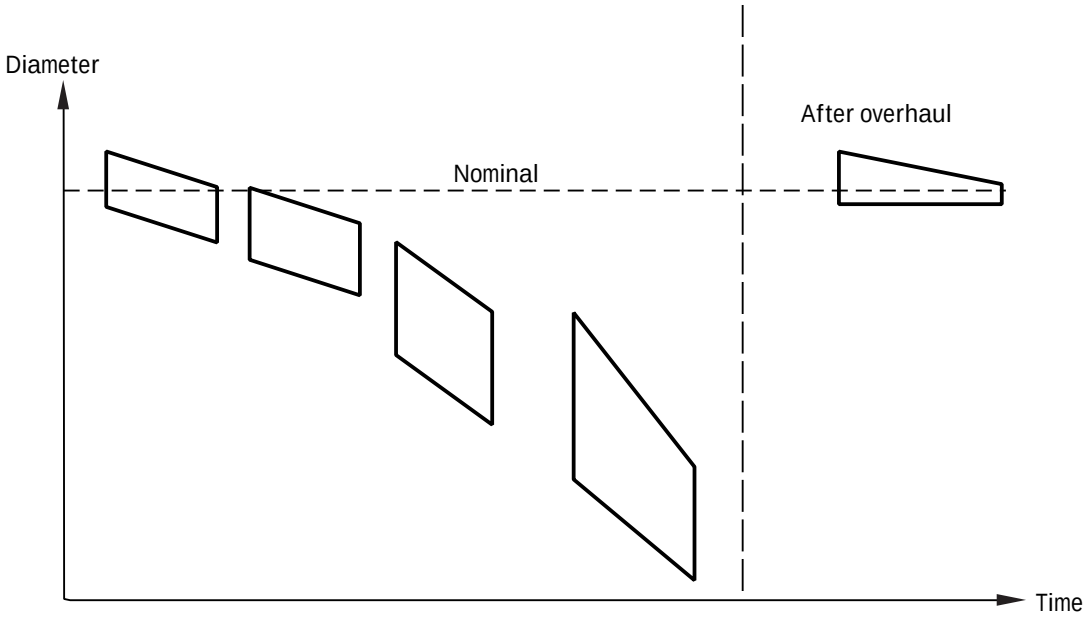


Figure 18c) — P-D diagrams indicating progressive decrease of mean and increase in geometric form variation and the beneficial effects of overhaul

Regarding Figure 18c) 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.

5.3.7 *Graphical portrayal of frequency distributions*

When only a few observations are available, *line or dot plots* as shown in Figures 3 and 12 or *stem and leaf plots* as in Figure 15 will often suffice. However, with a larger number of observations 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 gives some technical 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 rough guide. 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
[*Sturge's rule: no. of classes* = 1 + 3.3.log_{10} *(no. of observations)*]

Number of observations	30 to 40	50 to 90	100	200 to 300	400 to 700	800 to 1 000	2 000 to 3 000	4 000 to 6 000	7 000 to 10 000
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 representative of production are required to withstand a standard 4 minute 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 minute.

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 − 4.3)/9 = 0.57 minutes. The results were arranged in ascending order and the actual class interval chosen was 0.5 minutes for greater simplicity and clarity. The resulting frequency and percentage frequency tabulation is 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 %
below 4.1	0	0
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	19.8
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
above 9.5	0	0

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, is distributed.

Various pictorial representations of the test data of Table 6 are shown in Figure 19. 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 minutes.

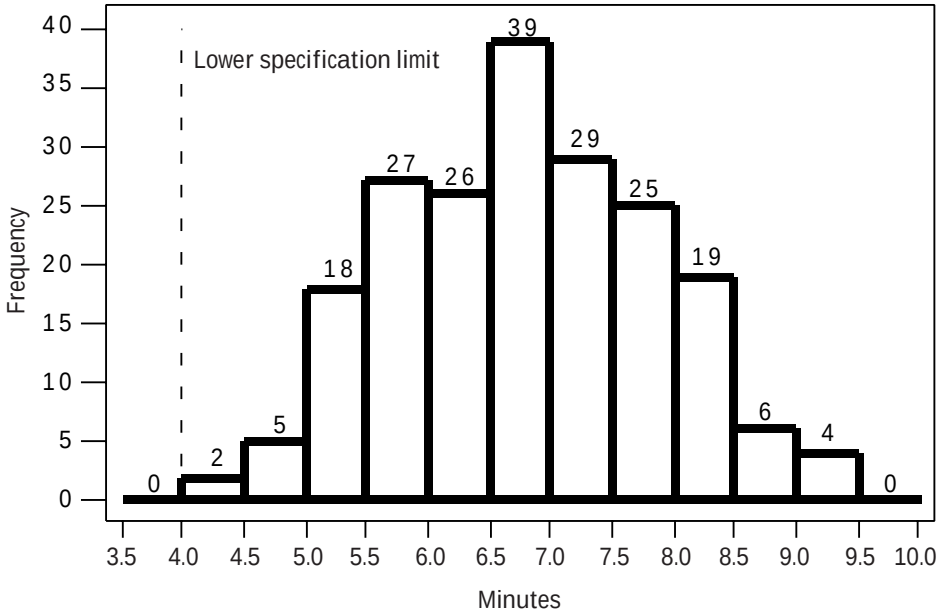


Figure 19a) — Frequency histogram for immersion times

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.

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.

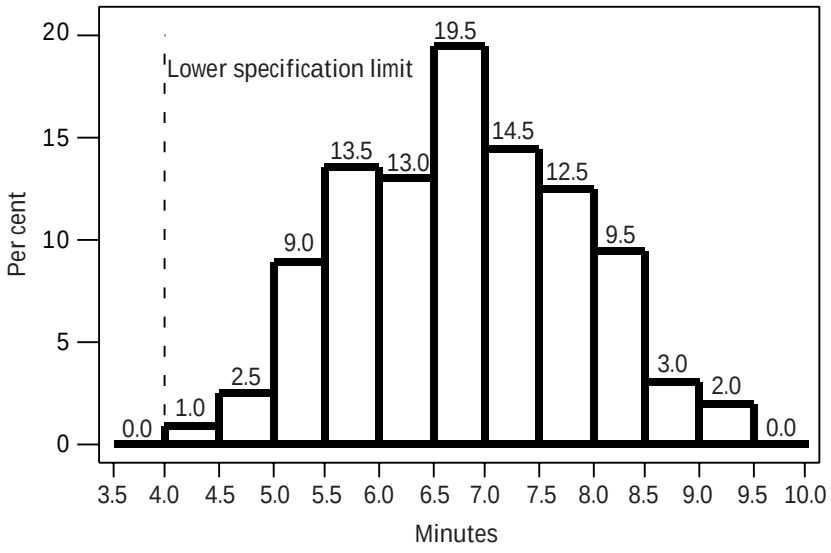


Figure 19b) — Percentage frequency histogram for immersion times

The cumulative relative frequency diagram shows the percentage of observations falling below (or above) particular values. By way of illustration, in Figure 19c) it is seen that 58.5 % fall below 7.0 minutes. 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.

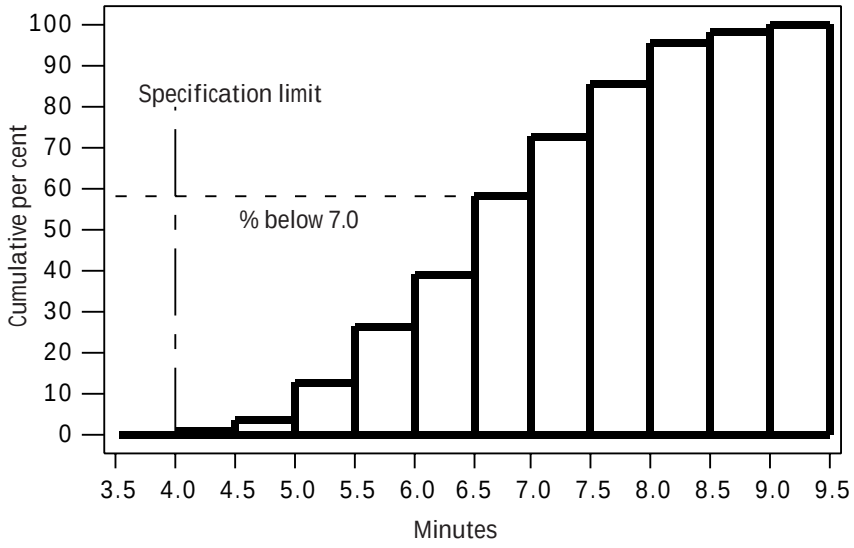


Figure 19c) — Cumulative percentage frequency histogram for immersion times

The cumulative histogram can alternatively be expressed in the form of a smooth curve as shown in Figure 19d), or preferably as a straight line by transformation of the vertical scale. This latter concept will be developed later in this clause.

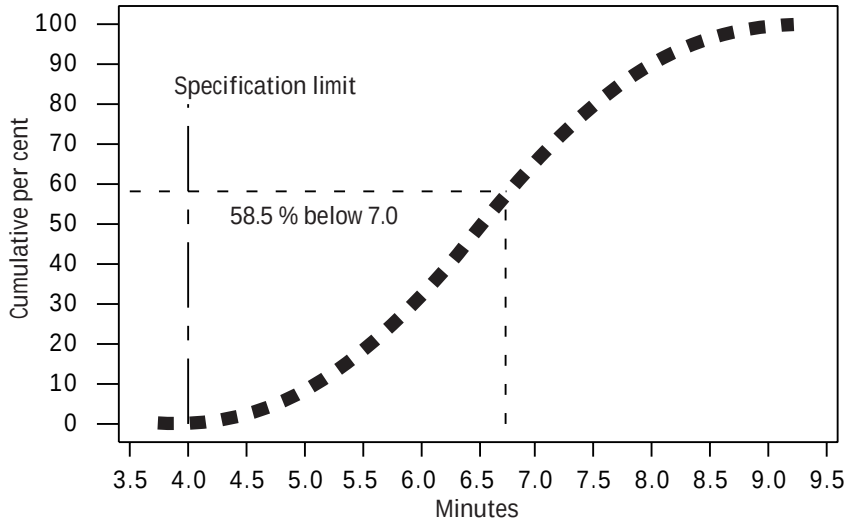


Figure 19d) — Cumulative percentage frequency diagram for immersion times

The diagrams of Figure 19 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 case of the zinc plating case the frequency histogram for immersion time [Figure 19a)] indicates a bell shaped symmetrical pattern of variation about the mean. This is typical of the standard 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 19e).

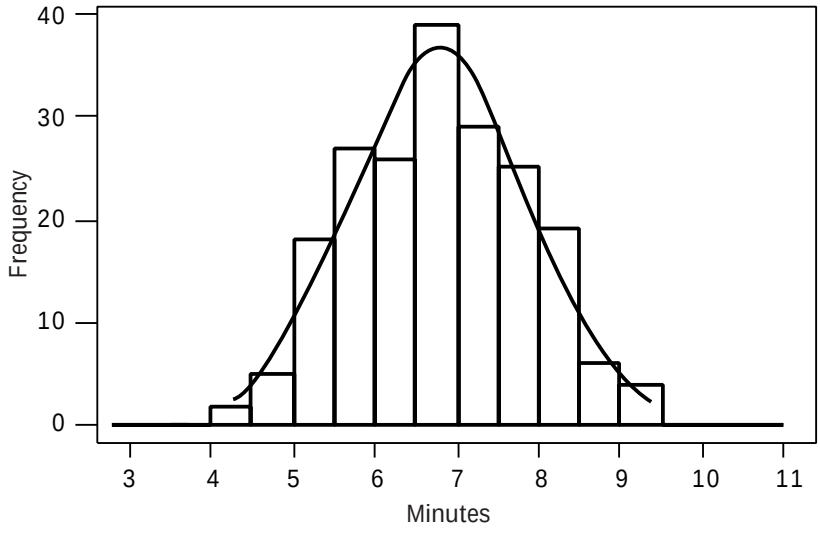


Figure 19e) — Normal curve overlaid on the immersion time histogram (mean = 6.79: standard deviation = 1.08)

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 19e). These show, by eye, a good correspondence between the two indicating that a normal distribution with a mean of 6.79 minutes and standard deviation of 1.08 minutes 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 (ref. ISO 16269).

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

An example of this test applied to the galvanized item immersion data is shown in Figure 19f). It is seen that the normal cumulative probability scale shown on the vertical axis of Figure 19f) transforms the bell shaped normal distribution into a straight line when plotted against immersion times.

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.

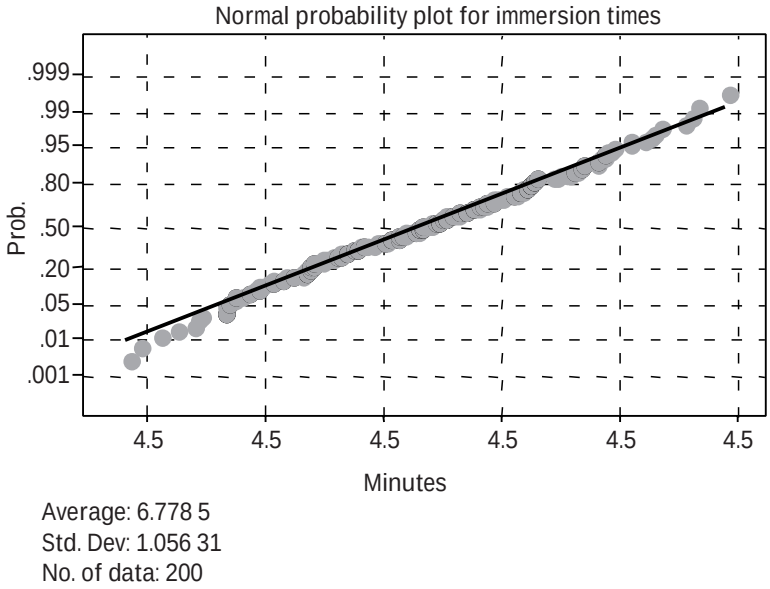


Figure 19f) — Straight line plot on normal probability paper indicating normality of data

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.

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.

Having said this, the normal distribution is the one most frequently encountered in many processes. Moreover, the distribution of means of samples, or subgroups, will be very closely 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 20 shows a standardized symmetrical bell shaped curve which characterizes this distribution. Additionally, some key percentages are included in relation to distances from the mean in terms of standard deviations. Whilst, in Figure 20, 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.

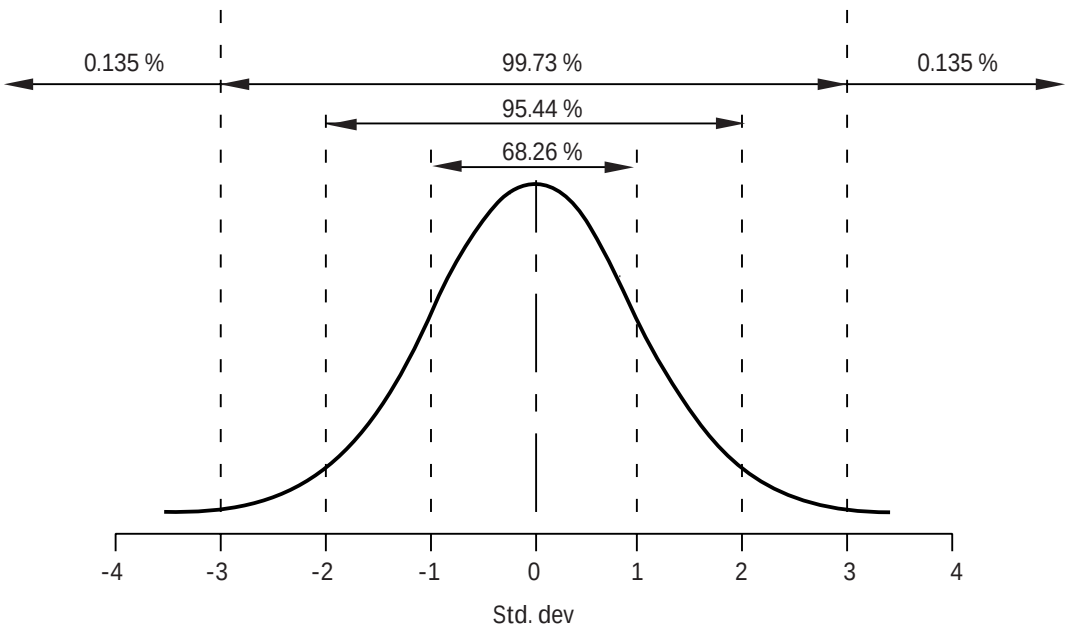


Figure 20 — Percentages of normal distribution (in terms of standard deviations)

Figure 20 illustrates that for a normal distribution with a stable mean and standard deviation:

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) 95.44 % of the values lie within the limits: mean ± 2 standard deviations;

d) just over two thirds (68.26 %) 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.

The example at the base of 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 19f) for making similar predictions.

Example: Clothing size survey

A size survey conducted on a representative sample of a prospective customer base indicated that one particular characteristic, height, was normal with a mean equal to 69″, and a standard deviation equal to 3″. By using the results of the survey it was predicted, for instance that:

— 16 % (15.87 %) of the target customer population are taller than 72″ (mean + 1 standard deviation);

— 25 % (25.14 %) of the target customer population are shorter than 67″ (mean − ⅔ standard deviation);

NOTE The ⅔ arises as 69 minus 67 expressed as a fraction of the standard deviation.

— 59 % (58.99 %) of the target customer population are between 67″ and 72″.

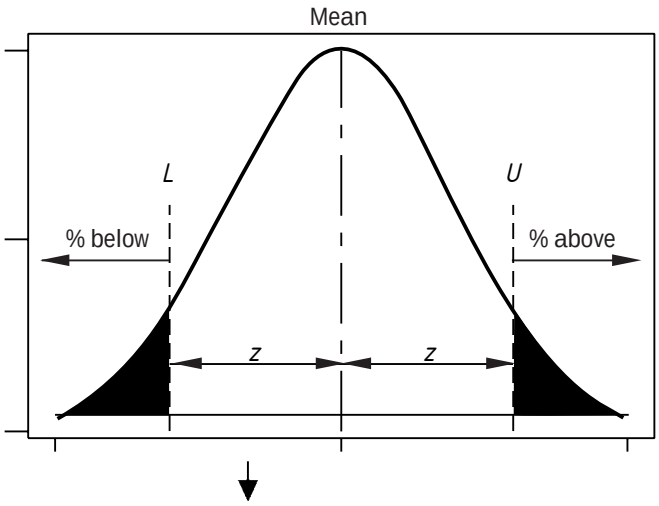
Such estimates enable appropriate size ratios of garments to be ordered.

NOTE Figure 20 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).

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 (English) 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 — Standard normal distribution — Percentage expected beyond a value, *U* or *L*, that is *z* standard deviation units from the mean



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	ppm								
5.0	0.3	ppm								
4.0	32	ppm								
3.5	.023	.022	.022	.021	.020	.019	.019	.018	.017	.017
3.4	.034	.032	.031	.030	.029	.028	.027	.026	.025	.024
3.3	.048	.047	.045	.043	.042	.040	.039	.038	.036	.035
3.2	.069	.066	.064	.062	.060	.058	.054	.054	.052	.050
3.1	.097	.094	.090	.087	.084	.082	.079	.076	.074	.071
3.0	.135	.131	.126	.122	.118	.114	.111	.107	.104	.100
2.9	.19	.18	.18	.17	.16	.16	.15	.15	.14	.14
2.8	.26	.25	.24	.23	.23	.22	.21	.21	.20	.19
→ 2.7	.35	.34	.33	.32	.31	.30	.29	.28	.27	.26
2.6	.47	.45	.44	.43	.41	.40	.39	.38	.37	.36
2.5	.62	.60	.59	.57	.55	.54	.52	.51	.49	.48
2.4	.82	.80	.78	.75	.73	.71	.69	.68	.66	.64
2.3	1.07	1.04	1.02	.99	.96	.94	.91	.89	.87	.84
2.2	1.39	1.36	1.32	1.29	1.25	1.22	1.19	1.16	1.13	1.10
2.1	1.79	1.74	1.70	1.66	1.62	1.58	1.54	1.50	1.46	1.43
2.0	2.28	2.22	2.17	2.12	2.07	2.02	1.97	1.92	1.88	1.83
1.9	2.87	2.81	2.74	2.68	2.62	2.56	2.50	2.44	2.39	2.33
1.8	3.59	3.51	3.44	3.36	3.29	3.22	3.14	3.07	3.01	2.94
1.7	4.46	4.36	4.27	4.18	4.09	4.01	3.92	3.84	3.75	3.67
1.6	5.48	5.37	5.26	5.16	5.05	4.95	4.85	4.75	4.65	4.55
1.5	6.68	6.55	6.43	6.30	6.18	6.06	5.94	5.82	5.71	5.59
1.4	8.08	7.93	7.78	7.64	7.49	7.35	7.21	7.08	6.94	6.81
1.3	9.68	9.51	9.34	9.18	9.01	8.85	8.69	8.53	8.38	8.23
1.2	11.51	11.31	11.12	10.93	10.75	10.56	10.38	10.20	10.03	9.85
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	22.06	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	24.83	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

NOTE For z = 4.0, 5.0 and 6.0 values are given in parts per million (ppm).

Example of 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 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 at 2.73 (2.7 from the left and 0.03 from the top as indicated by the arrows) to give 0.32 % above upper specification limit.

To find percentage below 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 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 percentage nonconforming is 0.32 + 18.41 = 18.73 %.

5.3.9 ***The Weibull distribution***

Unlike the fixed shape two-parameter normal pattern of variation, the Weibull distribution is a flexible versatile three-parameter (*shape*, *scale and location*), multi-shaped one. This is extremely useful for failure diagnosis in design and development and supplier/customer interface activity in the reliability field and for curve fitting, process diagnosis, and separating out elements of variation in the quality field. It can deal with parameters that affect both reliability and quality, namely:

a) total instantaneous breakdown; and

b) variation in performance.

As with the normal distribution, a simple graphical approach may be adopted using Weibull probability paper. The principal benefit is that it reduces a complex mathematical formula to linear form in a simple graph. The basis of the scales on the graph need not concern us at this stage other than that the horizontal scale is 2 cycle logarithmic and therefore the values appropriate to the size and range of measurements need to be inserted on the scale prior to inserting data plotting points. Alternatively, computer based routines are readily available. An example is shown in Figure 21.

The Weibull shape parameter, β, can take the value of any positive real number. This is a very powerful parameter that gives the Weibull distribution its versatility in representing any of a number of distribution forms. Examples are now given:

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 becomes roughly symmetrically normal;

4) $\beta > 3.5$ the distribution stays largely symmetrical (slight skewness) and becomes progressively more peaky as beta increases.

A mental referencing of these four cases thus enables one to visualize the underlying form of frequency distribution once the shape parameter has been determined. Graphic examples of the effect of different shape parameters on distribution shape are shown in Figure 30.

Whatever the value of β the distribution may be represented by straight lines with differing slopes on Weibull probability paper.

Quite apart from this straight line representation of frequency distributions, Weibull analysis plays a major role in reliability analysis in its power of discrimination between different failure regimes and failure rates. The three possible failure regimes are often represented by the generic *bath tub* curve, that of infant mortality (decreasing failure rate), prime of life (constant failure rate) and wear-out (increasing failure rate). It should be noted that departures from this sequence occur in practice. The Weibull β parameter distinguishes between these regimes thus:

i) $\beta < 1$ represents a decreasing failure rate regime (colloquially called infant mortality);

ii) $\beta = 1$ represents a constant failure rate regime (often called prime of life);

iii) $\beta > 1$ represents an increasing failure rate regime (frequently termed wear-out).

The formula for instantaneous failure rate is $\frac{\beta}{\alpha} t^{\beta-1}$ where α is the scale parameter and t is the operating time.

For example, when β equals 1, giving a constant failure rate, α becomes the customary mean time between failures, namely:

$$\text{Mean time to failure} = \frac{1}{\text{Instantaneous failure rate}}$$

When β equals 2, the failure rate increases linearly with respect to the operating time, t; when β equals 3, it increases according to the square of operating time.

On those occasions when the possibility of failure (or the origin of the distribution) does not start at time zero, the third Weibull parameter, γ, comes into play. γ is the location parameter which is often zero in reliability situations. Reliability illustrations are when it is feasible for an entity to fail before operation, for example, it "fails on the shelf" or is found "dead on delivery". When γ is not zero, such a situation is recognized in a Weibull probability plot by the data points lying on a smooth curve rather than on a straight line. In such a case the estimated value of γ is subtracted from each of the plotted points and the new values re-plotted. It may take a few iterations to arrive at a best estimate straight line and hence at a reliable value for γ.

Example

Times to failure, in hours, of hybrid units, tested under similar conditions, are as follows:

179, 507, 949, 1 454, 2 317, 3 345, 4 302, 5 687, 7 674, 12 315.

It is wished to provide a best estimate of the following:

I) the regime of failure;

II) the probability of survival to 1 000 hours.

From Figure 21 it is seen that the shape parameter is 1.0. Hence the failure regime is one of constant failure rate. It can also be seen that the cumulative percent failing by 1 000 hours is 23 %. Hence the probability of survival (or reliability) at 1 000 hours is 77 %.

Alternatively, reliability is given by the formula:

$$\text{Reliability} = e^{\frac{t^{\beta}}{\alpha}}$$

$$\text{Thence, reliability} = e^{\frac{1\,000^{1.0}}{3\,875}} = 77\ \%.$$

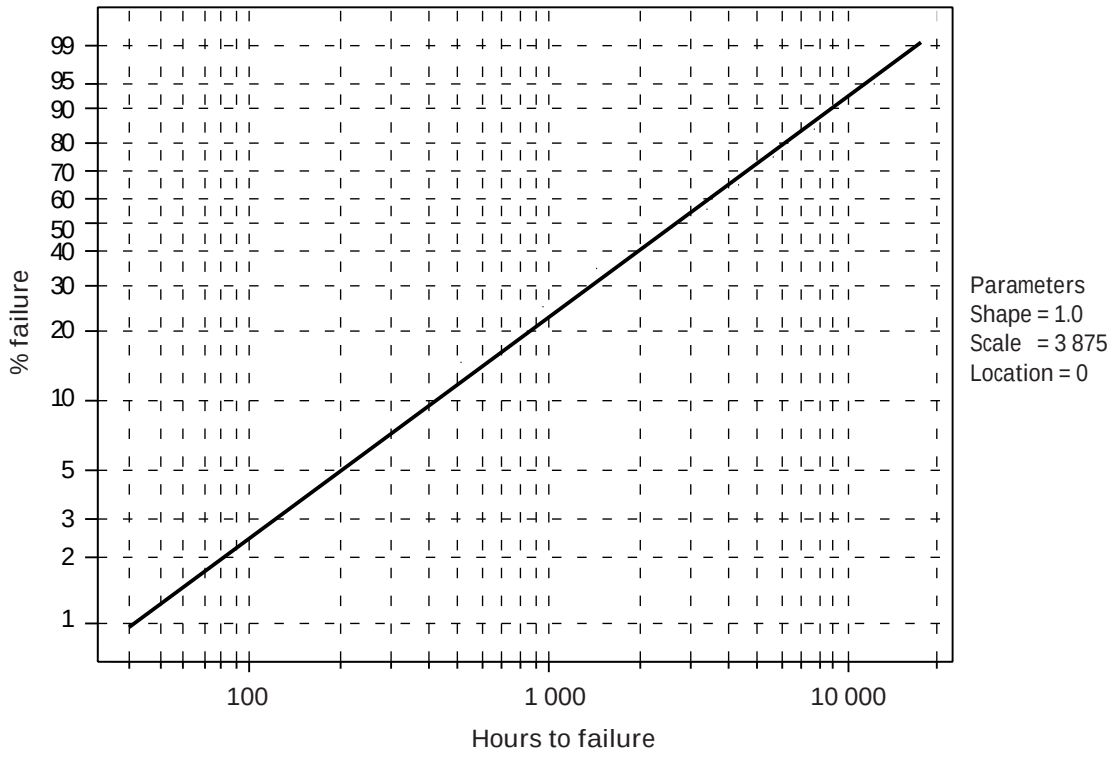


Figure 21 — Weibull probability plot for hybrid unit on life test

5.3.10 ***Graphs***

A graph is essentially a representation of data by a continuous curve (a line on a graph may be referred to as a curve even though it may be straight). Graphs are constructed to provide visual communication of information with clarity and precision. There are various forms of graphs, such as the following:

a) arithmetic (linear): these are the most familiar and are easily identified by the fact that both horizontal and vertical scales are arithmetic (linear) (see Figure 22);

b) log-linear: semi-logarithmic graphs have a linear horizontal scale and a logarithmic vertical scale and are used to display rates of change; a constant rate of change will appear as a straight line;

c) log-log: these have both scales logarithmic and are used to express learning curves in straight line form;

d) probability plot: these transform a regular pattern of variation into a straight line [see Figure 19f)];

e) nomograph: these provide graphical solutions to formulae.

5.3.11 ***Scatter diagram and regression***

A scatter diagram is used to test and display possible relationships between one variable and another. The nature of the relationship between the two variables is represented by the direction and shape of the line (or curve) of best fit. In determining the degree of correlation between variables the distinction between correlation and causation should be borne in mind.

An example of a scatter diagram is shown in Figure 22.

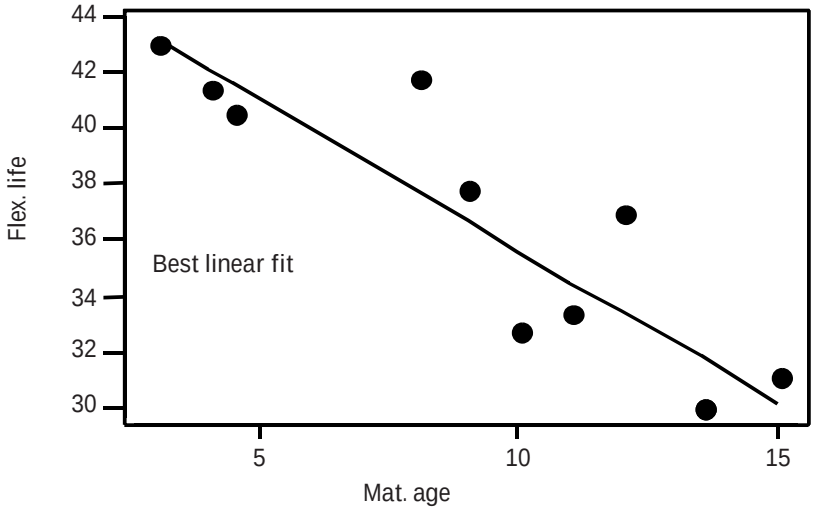


Figure 22 — Scatter diagram for flexing life of rubber in terms of material age

The degree of linear fit is indicated by the correlation coefficient, r. The closer r is to ± 1 the better the fit. Here, r is -0.9. This indicates that some 81 % (r^2) of the variation in flexing life is explained by the simple linear regression model:

flexing life = 46.4 − 1.08 (material age).

5.3.12 ***"Pareto" (or Lorenz) diagram***

A " Pareto" diagram is a simple graphical technique for displaying the relative importance of features, problems or causes of problems as a basis for establishing priorities. It distinguishes between the "vital few" and the "trivial many" and hence focuses attention on issues where maximum quality may be secured most quickly.

It displays, in decreasing order, the relative contribution of each element (or cause) to the total situation (problem). Relative contribution may be based on relative frequency, relative cost or some other measure of impact. Contributions are shown in bar chart form. Sometimes a cumulative line may be added to show the accumulated contribution. An example is shown in Figure 23.

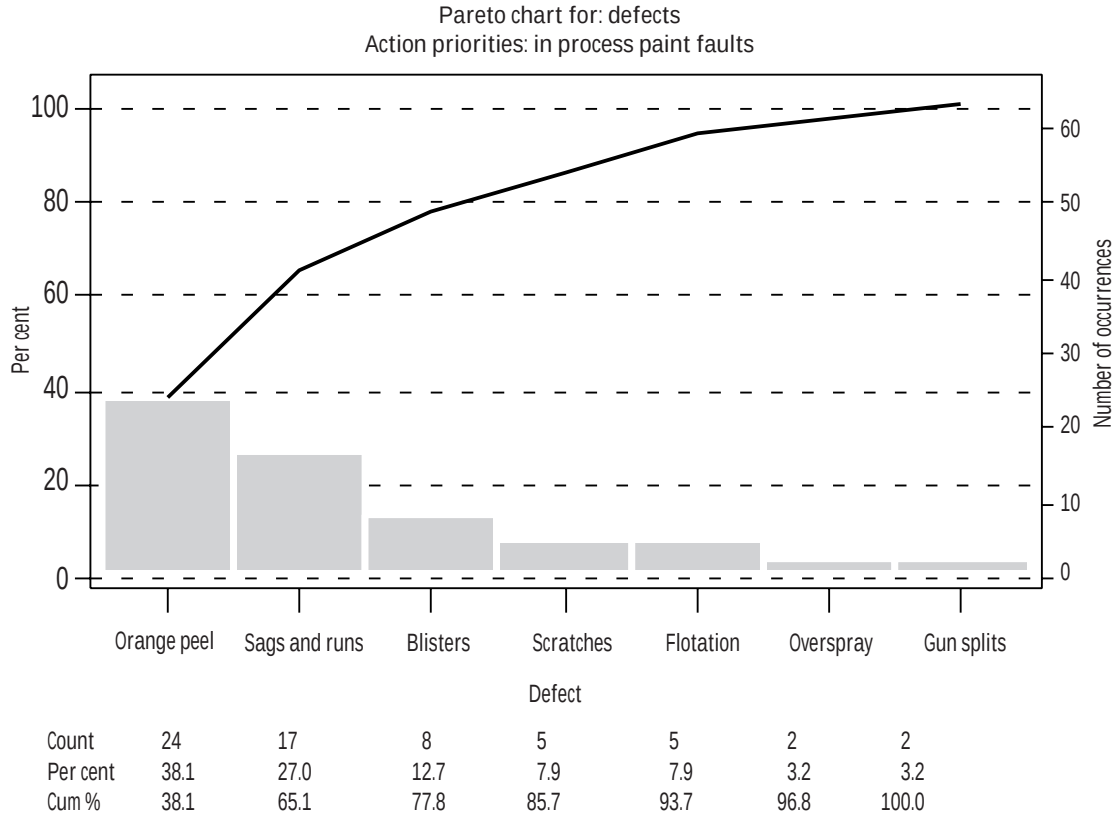


Figure 23 — Relative contribution of different types of in-process paint faults

Figure 23 shows that orange peel and sags and runs make up some 65 % of total paint faults in a particular paint shop. These were selected for priority attention in a quality improvement drive.

5.3.13 ***Cause and effect diagram***

A cause and effect diagram is frequently referred to as a fishbone diagram (because of its shape) or an Ishikawa diagram (after its creator). It applies where it is required to show pictorially cause and effect relationships. There are several types, based on the formation of the main branches (categories), including:

a) general 4M (manpower, machines, materials, methods);

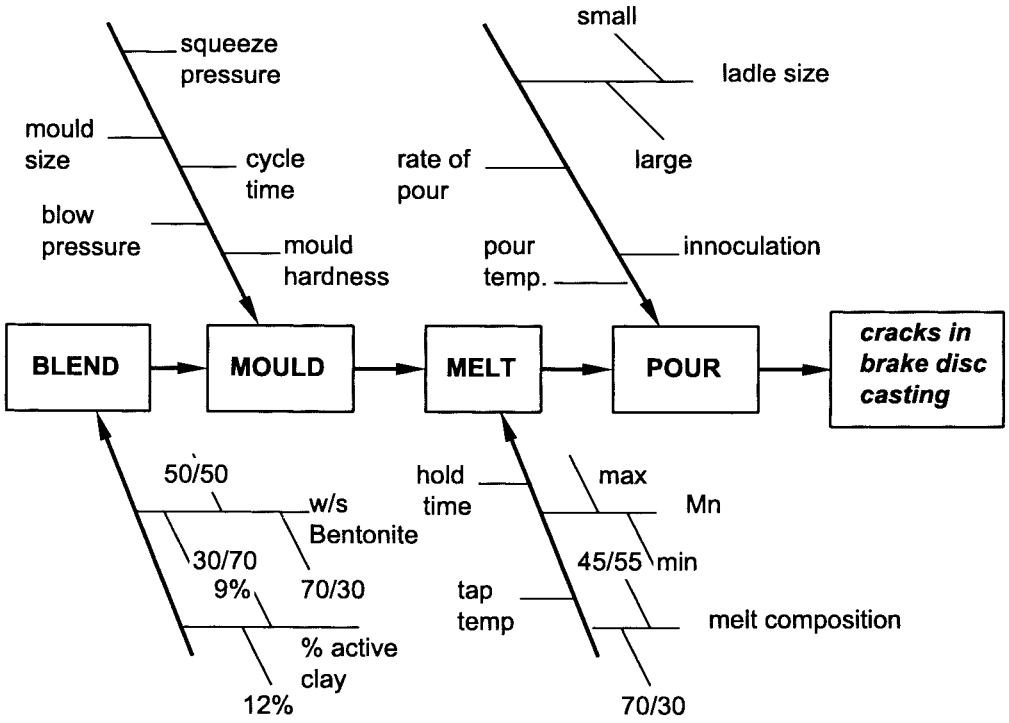
b) general 4P (people, procedures, plant, process);

c) process (by process steps and sequence);

d) assembly (by sub-assemblies);

e) specific (by technical consideration).

A process cause and effect diagram for a foundry process is shown in Figure 24.



NOTE w/s identified the origin of the bentonite, namely western/southern.

Figure 24 — Process cause and effect diagram for cracks in a casting

6 Variation and sampling considerations

6.1 Statistical control and process capability

6.1.1 *Statistical control*

Post process 100 % inspection is often neither practicable, relevant nor timely enough to meet today's needs. Monitoring in real time is required to enable processes to be steered and managed in an effective manner. From economic considerations, amongst others, monitoring usually involves the assessment of process parameters and resulting product characteristics from a limited number of observations or items, which are defined in statistical terms as a sample.

It is essential to take the variation between similar items or observations into account when considering the relation of the sample to the totality of objects under consideration. This is true whether the object is a process parameter such as teeming temperature, a constituent of a material such as % silicon, or a characteristic of a product such as a the diameter of a rod. In all cases where sampling is undertaken, estimates for the totality of objects under consideration can only be answered satisfactorily with the aid of statistical treatment.

Two statistical definitions, relating to population and lot, are relevant to an understanding of the following text. A population is defined as the totality of objects under consideration. A lot is defined as a definite part of a population constituted under essentially the same conditions as the population with respect to the sampling purpose.

The relationship between sample and lot is the kernel of the problem. In discussing it, it is necessary to introduce certain ideas that may appear difficult because they are unfamiliar. The following paragraphs, if studied in conjunction with Figure 25, should convey the essential features of the statistician's method of approach.

Suppose that Figure 25 represents the results of tests made at a supplier on similar articles or component parts sampled five at a time, at regular intervals during production. For example, the data may relate to the number of millilitres of battery acid per bottle, or to the minimum temperature of operation of a certain device, etc. The results of each sample of five tests are shown in the figure as dots, with the measurements displayed on a horizontal scale. Six cases are presented, numbered 1 to 6, in each of which the results of 12 samples from consecutive lots are shown. Beneath the dots for each case, curves have been drawn. These represent the hypothetical distribution of the measured characteristic that would be found were it possible to test all the items in the 12 lots that have been sampled.

Consider case 1. There are, of course, considerable differences between the dot patterns of the 12 samples. Yet a certain stability or uniformity in the variation from sample to sample is evident, which is clearly not so in case 5. In case 1 there is no indication that the production process is anything but stable through time, as the samples could quite easily be imagined to have been drawn from the same lot.

If the pattern of variation is stable, then:

a) when all that are available are the measurements of the characteristic in a random sample, it is possible to use these measurements to estimate the distribution curve of the characteristic for the process;

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

Notice that situations a) and b) are the inverse of one another.

In forming the estimate described in a), the larger the size of sample, the more reliable the estimate. For example, consider the distribution curve shown for case 1. Its mean and standard deviation would be estimated more reliably from 12 samples of size 5 than from a single sample of size 5. This is simply common sense but, when it is required to draw inferences about the distribution curve that depend on the *extent* of this reliability, the assistance of statistical theory is necessary. Much of what follows in this British Standard is concerned, directly or indirectly, with the problems of drawing inferences and making decisions related to the process distribution curves, based on limited information. The construction of confidence intervals, prediction intervals and statistical tolerance intervals, addressed in clause **8**, are but three examples of these problems.

The following example illustrates situation b). For safety and ease of transportation, car batteries are supplied dry, together with plastic bottles of acid. With too little acid per bottle, the battery electrodes will not be fully covered, while if there is too much, the cell of the battery could overflow or present the user with the problem of disposing of the surplus acid. Suppose it is known, based on extensive experience, that the extent to which the bottles are filled varies according to the normal distribution curve. Suppose also that the mean and standard deviation of this normal curve are known to be 729.0 millilitres and 2.0 millilitres respectively. Then statistical theory enables such statements as the following to be made.

1) The chance that the mean contents in a random sample of six bottles will fall below 726.5 millilitres is 0.001 1, or such a result may be expected only once in about 900 samples.

2) The chance that *at least one bottle* in a sample of six bottles will contain less than 726.5 millilitres is 0.488 3, i.e. this result is over 400 times as likely as the previous result.

Statistical theory leads one to *expect* these two chances to be entirely different; moreover, statistical theory is able to give precision to the expected.

When this characteristic of stability of distribution obtains, as represented by case 1 of Figure 25, the process will be described as *stable*, or *under statistical control*. It is then possible to make use of the methods of statistical theory described in the following clauses for such purposes of inference or prediction as were referred to under a) and b) above.

Although it is not easy to give a precise non-mathematical definition of what is meant by saying that a process is under statistical control, the concept is not difficult to grasp. It will be illustrated below using Figure 25, by contrasting it with some cases where the pattern of variation from lot to lot is not in statistical control.

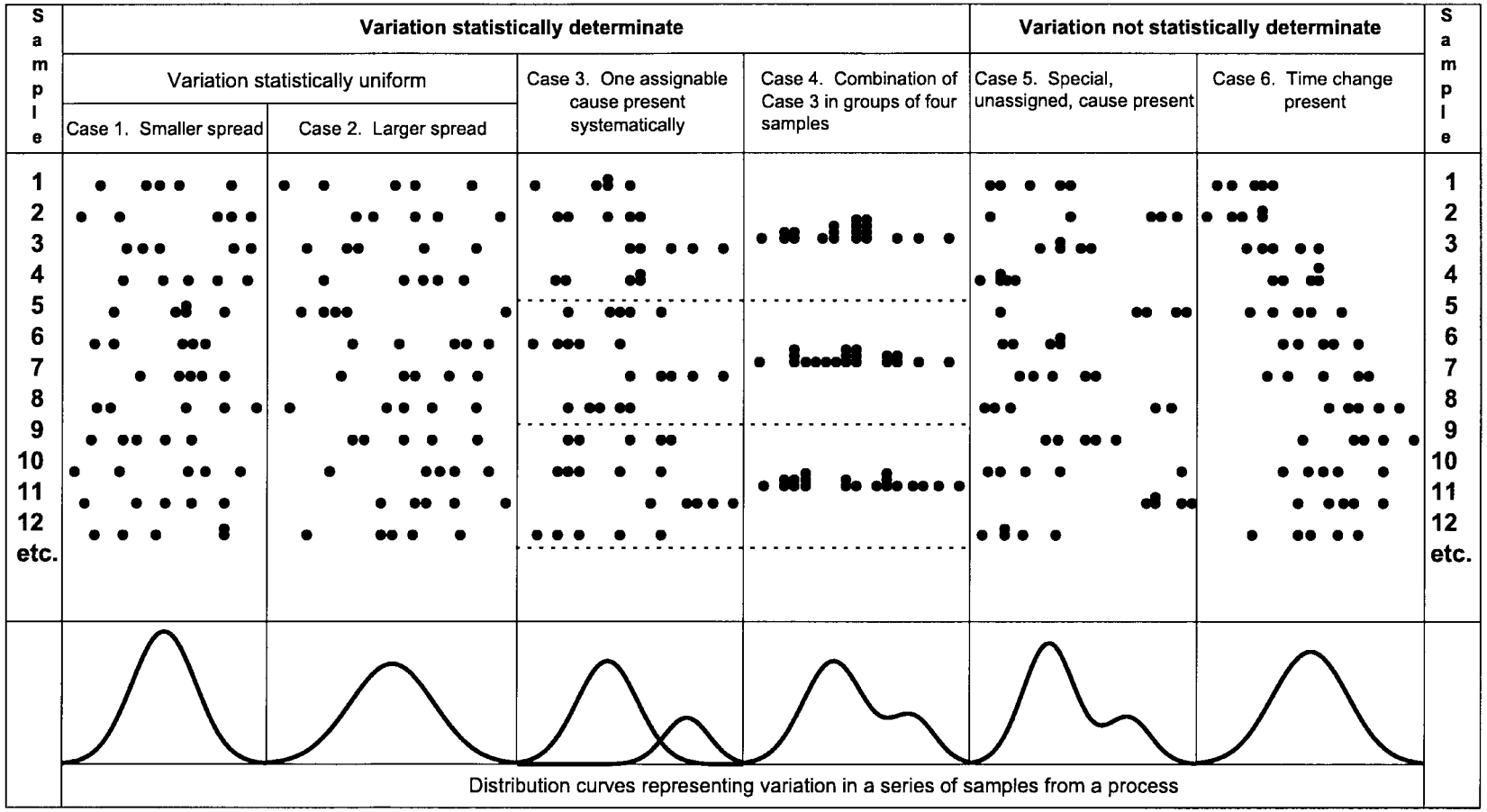


Figure 25 — Diagram indicating types of variation in samples

6.1.2 *Erratic variation*

While the variation in case 1 appears to be in statistical control, the variation for case 5 most certainly does not. A production process that leads without assignable cause to both sample No. 4 and sample No. 11 in case 5 can hardly be considered to be in statistical control. Indeed, the dot patterns suggest that there may be factors at work leading to two centres of variation with sometimes one operating, sometimes the other and sometimes both at the same time. This lack of homogeneity is suggested by the distribution curve for the 12 lots combined, shown beneath the dot patterns. The variation in case 5 may best be described as out of statistical control. Without any further understanding of the factors affecting the lot to lot variation, any attempt to predict the variation in subsequent product from the process is futile.

6.1.3 *Systematic variation*

Another situation is presented by case 3, where samples no. 3, 7 and 11 appear different from the others. However, here there is a systematic repetition in the irregularities, which was not evident in case 5, and it may be the case that it is possible to determine a cause for these differences. If this were so, then the preferred procedure would be to eliminate the cause. However, if this were not possible, the series of lots could be divided into two homogeneous sub-series, within each of which there is statistical control, and to each of which statistical methods could usefully be applied.

Suppose, in case 3, that samples 1, 5, 9, etc. were from sub-lots of material from one machine, samples 2, 6, 10, etc. from sub-lots from a second machine, 3, 7, 11, etc. from a third and 4, 8, 12, etc. from a fourth. If lots were formed by combining one sub-lot from each machine, then the samples of size 20 shown as case 4 could be described as representative of the total output. This illustrates the difference between simple random sampling and representative sampling. Under representative sampling, the output is divided into homogeneous parts from each of which a random sub-sample is drawn of size proportional to the part, and the sub-samples are then combined into a representative sample. Thus, in the example just described, a quarter of each sample of size 20, i.e. five items, are selected from a sub-lot from each machine. Contrast this with simple random sampling, under which every possible sample of the same size from the lot would have exactly the same chance of being selected. Thus, under simple random sampling it would be possible for the sample of 20 items to contain no items at all from the third machine. Conversely, it would be possible for half or more of the items in the sample to come from the third machine. Clearly, such events would be undesirable in the present example.

To illustrate the advantage of representative sampling, consider the manufacture of sheet brass, the thickness of which at the edges is less than at the centre owing to the nature of the rolling process. If sheets are cut into narrower widths, the thickness will vary according to the position from which the strip has been cut. If this variation is recognized, the product will be divided accordingly into parts that will be homogeneous for sampling purposes. However, if it is not recognized, it is then possible that samples would sometimes consist of test pieces all taken from the edges, sometimes all from the centre, and sometimes from both in various proportions. The situation will then be as case 5, with no reliable inference possible from the sample measurements.

The distinction typified by the differences between case 1 and case 3 is an important one. In the former case, any one of the samples may be used to give information regarding the total output from the manufacturing process; in the latter, care needs to be exercised in choosing representative samples. In the one, the variation within and between the samples from individual lots is no different from what might have been expected if a series of random samples had been selected from a consignment formed by first combining and mixing the items from the separate lots. In the other, it is as though the items were drawn from a number of sources of different constitution, the proportion taken from each source being in proportion to the quantity from that source, although the drawing from within each source was at random.

The term *stratified sampling* is often wrongly used as a synonym to representative sampling. In fact, representative sampling is a very simple special case of stratified sampling. Like representative sampling, stratified sampling is used when the product can be divided into relatively homogeneous strata. There the likeness ends, however. The choice of sample size from each stratum under stratified sampling takes into account the cost of sampling an item from each of the strata and prior estimates of the variability within each stratum. The choice is made with the objective of minimizing the cost of achieving a given precision in estimating the average value of the characteristic of interest, or maximizing the precision for a given cost. Representative sampling is therefore a special case of stratified sampling in which the strata have known and equal variability, and the cost of sampling an item from each stratum is the same.

6.1.4 *Systematic changes with time*

Case 6 represents another situation, where there is a systematic change with time taking place in the quality of material produced. If causes can be found for these fluctuations then, if these causes cannot be eliminated, it may alternatively be possible to divide the process output into streams that may separately be considered to be under statistical control. Examples are fluctuations due to known changes in temperature or humidity (perhaps in some textile process) or to differences between operators or shifts. However, if there are irregular fluctuations in time *without* known cause, prediction of characteristics of the process based on sample measurements from a few lots will be impossible. Thus, samples nos. 1 and 2, or again nos. 8 or 9, would not be representative of the process distribution shown below the dot patterns.

6.1.5 *Statistical indeterminacy*

If the total output of a particular article is made up from a number of sources, where each source is under statistical control and the proportion of the total coming from each is known, we have seen that statistical methods can be used to estimate the quality of the total from properly drawn samples. If, on the other hand, there is insufficient information to enable properly representative samples to be drawn, the variation is statistically indeterminate. This indeterminacy may be due to changes in space, e.g. from one machine or supplier to another; or it may be due to changes in quality with time, e.g. changes in the product from one supplier due to seasonal influences or changes in raw material.

6.1.6 *Non-normal variation*

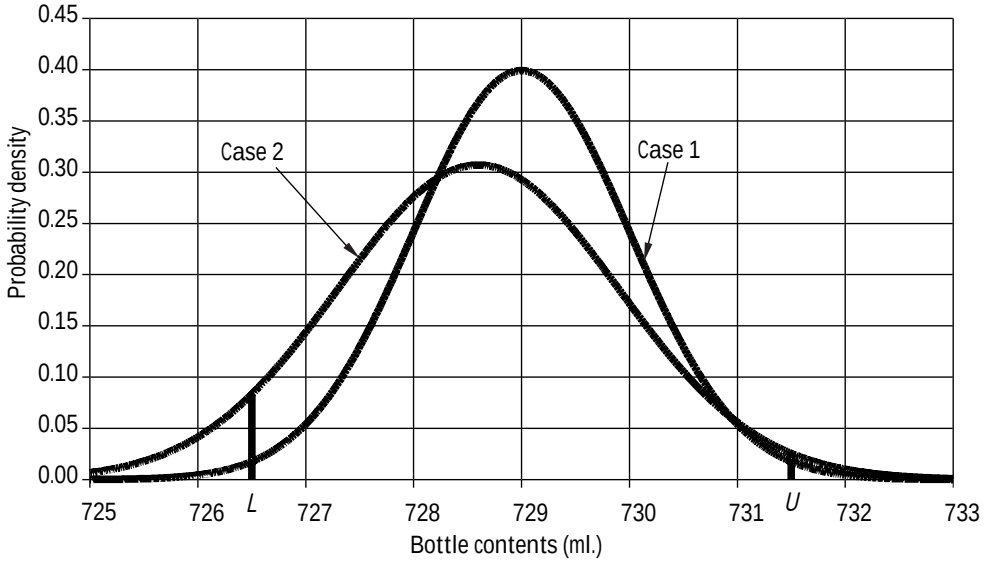
It is important to realize that variation that is under statistical control is not necessarily represented by the normal distribution curve. It is true that the underlying distribution *is* normal in a very large proportion of cases met with in industrial experience. Indeed, most of the methods described in clauses **8** and **9** rely on the variation being approximately of the normal form. Nevertheless, it should be recognized that examples also abound for which the distribution curve of the measured characteristic is far from being symmetrical, e.g. the distributions of lifetimes and breaking loads, which typically have a long tail to the right. Yet provided the distribution remains stable from lot to lot, the concept of a process being in statistical control remains appropriate.

6.1.7 *Quality level and process capability*

There is one further important concept, *process capability*, which needs to be introduced before dealing with the significance of these ideas to the supplier and the customer. Consider case 2, which has not yet been discussed. As in case 1, the variation in case 2 appears to be statistically uniform from lot to lot, i.e. under statistical control. In practice, evidence of stability is not enough. Statistical uniformity does not of itself indicate whether a process is operating at a high or at a low quality level. In order to be able to assess the quality level, information is also required concerning the process mean and the process variation. This information is provided by the sample mean and the sample standard deviation. (Incidentally, with this information, it will also be possible to detect a departure from uniformity during production, which will often enable adjustments to the process to be made to maintain a good quality level. Control chart methods that may be used for this purpose are discussed in clause **10**.)

The process variation for both case 1 and case 2 has been represented by normal distribution curves beneath their respective dot patterns in Figure 25. However, the cases differ in that case 2 has greater variation of individual items within samples than case 1. It has been previously pointed out that a normal distribution curve is completely defined by its mean, μ, and its standard deviation, σ. Denoting the process mean and process standard deviation for case 1 by μ_1 and σ_1, and for case 2 by μ_2 and σ_2, it is evident that the difference between case 2 and case 1 is that σ_2 is greater than σ_1.

Suppose that case 1 and case 2 represent the variation in the amount of car battery acid per bottle from two different filling machines with $\sigma_1 = 1.0$ ml and $\sigma_2 = 1.3$ ml. Suppose also that the mean contents also differ, say $\mu_1 = 729.0$ ml and $\mu_2 = 728.6$ ml, although this is not possible to see from Figure 25 as no scale is given. In both cases, statistical theory could be used to predict the proportion of bottles whose contents lie within *any* given limits. Suppose the specification is for a minimum of $L = 726.5$ ml and a maximum of $U = 731.5$ ml. The situation is shown in Figure 26.



NOTE Comparison of capabilities of a normal distribution having $\mu_1 = 729.0$ and $\sigma_1 = 1.0$ with a normal distribution having $\mu_2 = 728.6$ and $\sigma_2 = 1.3$ when the specification limits are 726.5 and 731.5.

Figure 26 — Contrast of the capabilities of two filling machines

Then evidently the capabilities of the filling machines to satisfy the requirements are different, with the first machine turning out a more homogeneous and more acceptable product. Indeed, it can be seen that the fraction of bottles that violate the lower limit for case 2 is many times that for case 1. Case 1 is therefore said to have greater *process capability*, i.e. the quality level of its output will be better than that of case 2. This distinction between the concepts of the statistical uniformity and the capability of a process is important.

One final remark about statistical uniformity, or statistical control, may be appropriate. They are terms used to describe the variation when the distribution curve *appears* to be stable from sample to sample. This stability is relative to the sampling technique employed, and is sometimes more apparent than real. For example, in a product that is being continuously produced, sampling at short intervals may identify a lack of statistical uniformity, e.g. a high frequency cyclical effect, which sampling at longer intervals could fail to detect.

6.2 Sampling considerations

Consider now the way in which the principles discussed above bear upon the problems of sampling in practice. In general, to what extent are samples drawn to enable statistical theory to be profitably applied? The question is too wide to give a single answer, as the methods of sampling which are practicable can vary enormously from one type of product to another. This notwithstanding, certain illustrations may profitably be presented to show some of the inherent difficulties and to indicate how they may be overcome.

Consider first the situation where the material sampled consists of a number of similar units, either component parts or finished articles. In some instances it will be relatively straightforward to secure a random sample from a single well-mixed source of supply, for example in sampling small engineering parts such as ball bearings, bolts, screws, etc. A supplier who is confident that his process is in statistical control can adopt a simple procedure such as setting aside every 500^{th} or $1\,000^{th}$ item (or whatever the need may be) to form samples for inspection purposes. The danger in such a procedure is of the time interval between the selection of sample items for inspection being in step with any periodic fluctuation in quality that may exist. Were this to occur, the sample may well be biased, in which case the conclusions drawn from it would be misleading. Examples of possible reasons for such fluctuations are diurnal changes in temperature, increasing fatigue or inattention of operators during the course of shifts, or periodic replenishment of the raw material from which the product is made.

More often, however, the problem is not so simple. This is usually the case when sampling needs to be carried out not just to determine the acceptability of a lot but also to determine the grade, and therefore the price, of the product before acceptance by the user. As there is often no evidence available that the supplier's quality level has remained constant, it is important to plan a sampling procedure that will provide a reliable estimate of the quality of each lot, even if each lot is inhomogeneous. In short, care has to be taken to draw a representative sample from each lot.

The following illustration of this point comes from the sampling sub-clause of prEN 12326-1.

> Sampling shall be carried out by selecting slates from each lot separately in a random way so that every slate has an equal chance of being selected. Selected slates shall be marked so as to identify which lot they came from.
>
> When there is a possibility that the slates being tested may contain localized harmful inclusions such as calcite veins or oxidizable pyrite, the preparation of the test pieces shall be modified to ensure sufficient inclusions are contained in the specimen to provide a representative result.

The acceptance procedure itself is not simple.

> Where one or more of the tests do not satisfy the requirements of this standard, the unsatisfactory tests are repeated. If the results of the unsatisfactory test are confirmed, the lot is rejected or re-designated depending on the results.
>
> If the repeated test is satisfactory, a second check is carried out and if the result is satisfactory, the lot is accepted. If the repeated test is unsatisfactory the lot shall be rejected or re-designated.

Different problems arise in sampling where material does not consist of discrete items, but is delivered in bulk, which for one reason or another may not be homogeneous. It is then necessary to withdraw small equal portions of material from a number of different parts of the whole mass. Alternatively, if the material is in movement on conveyors or in barrows, similar portions may be taken at regular intervals during the whole period of movement. The usual practice is then to combine these portions to form initial samples, which are then reduced after thorough mixing to form small quantities of material suitable for analysis in the laboratory. The object of these activities is to obtain final samples that are as representative as possible.

ISO 3082 illustrates the precautions that are necessary when sampling from bulk materials such as iron ore. The standard contains diagrams of many types of sampling and dividing devices, illustrating the difficulties of obtaining representative samples. The following extracts indicate the general considerations for sampling and sample preparation.

> The basic requirement for a correct sampling scheme is that all parts of the ore in the lot have an equal opportunity of being selected and becoming part of the partial sample or gross sample for analysis. Any deviation from this basic requirement can result in an unacceptable loss of accuracy and precision. An incorrect sampling scheme cannot be relied upon to provide representative samples.
>
> The best sampling location to satisfy the above requirement is at a transfer point between conveyor belts. Here, the full cross-section of the ore stream can be conveniently intercepted at regular intervals, enabling representative samples to be obtained.
>
> In-situ sampling of ships, stockpiles, containers and bunkers is not permitted, because it is impossible to drive the sampling device down to the bottom and extract the full column of ore. Consequently, all parts of the lot do not have an equal opportunity of being sampled. The only effective procedure is sampling from a conveyor belt when ore is being conveyed to or from the ship, stockpile, container or bunker.
>
> In-situ sampling from stationary situations such as wagons is permitted only for fine ore concentrates, provided the sampling device, e.g. a spear or auger, penetrates to the full depth of the concentrate at the point selected for sampling and the full column of concentrate is extracted.
>
> Moisture samples shall be processed as soon as possible, and test portions weighed immediately. If this is not possible, samples shall be stored in impervious airtight containers with a minimum of free air space to minimize any change in moisture content, but should be prepared without delay.
>
> Minimization of bias in sampling and sample preparation is vitally important. Unlike precision, which can be improved by collecting more increments or repeating measurements, bias cannot be reduced by replicating measurements. Consequently, the minimization or preferably elimination of possible biases should be regarded as more important than improvement of precision. Sources of bias that can be completely eliminated at the outset by correct design of the sampling and sample preparation system include sample spillage, sample contamination and incorrect extraction of increments, while sources that can be minimized but not completely eliminated include change in moisture content, loss of dust and particle degradation (for particle size determination).

In this example, there is no question of lots of ore being rejected; the sampling is solely to determine the grade and price.

If, from each lot, only one final sample were produced for laboratory analysis, there would be no way of assessing the reliability of the estimates of the lot characteristics. ISO 3084 provides details of how this problem may be handled by the use of interleaved samples. These are "samples constituted by placing consecutive primary increments alternately into two sample containers" where the primary increments are the quantities of ore collected in a single operation of the sampling device. ISO 3084 provides for four scenarios.

a) When lots are frequently delivered, the quality variation may be determined from a large number of lots of almost equal mass by treating each lot separately and making up a pair of interleaved samples for each lot.

b) When large lots are infrequently delivered, the quality variation may be determined from a single lot by splitting the lot into at least 10 parts of almost equal mass and making up a pair of interleaved samples for each part.

c) When small lots are frequently delivered, the quality variation may be determined from several lots of almost equal mass by splitting all the lots involved into a total of at least 10 parts of almost equal mass and making up a pair of interleaved samples for each part.

d) When sampling a wagon-borne lot where increments are taken from all wagons comprising the lot, the quality variation may be determined by treating each lot separately and making up a pair of interleaved samples for each lot.

Instructions are given in ISO 3084 for utilizing the information thus obtained under each scenario.

7 Methods of conformity assessment

7.1 The statistical concept of a population

To some extent, the customer and supplier have different viewpoints when it comes to the question of setting specifications. Broadly speaking, the customer is interested in the whole range of quality of individual items on the market from which he can purchase. The supplier has one eye on the competition, but is also concerned with the statistical control and capability of the production process which can or needs to be maintained in a particular organization or organizations, having regard to technical and economic constraints. In both cases, however, the form of variation in the characteristics of a large collection of individual items is a matter of concern.

In discussing the concept of statistical uniformity, frequent reference has been made to the distribution curves shown at the bottom of the charts in Figure 25. These curves were drawn to represent the frequency distribution of a characteristic that would be obtained if measurements were made on a large collection or aggregation of items. In statistical terminology, the word *population* has been used to describe such a large collection of individual items, each possessing perhaps a number of different variable characteristics.

The use of this term arose because the early development of statistical method was associated with the study of human populations, the individuals forming which were variable and many-charactered. In such a case it is easy to grasp the concept of populations which are homogeneous or heterogeneous, stable or changing, the necessity of sampling, the idea of a representative sample, a biased sample, an adequate sample, and so forth.

Deriving their origin from this special field of application, the terms *sample* and *population* have had associated with them very definite meanings in statistical theory. In the field of industrial production, the meaning of a sample is clear, but the concept of a population will perhaps be more readily understood if a different terminology is employed. It is sensible for the larger collection of items from which the sample is drawn – the statistician's population – to be described differently according to the particular situation under consideration. The terms output, consignment, batch or lot may each be used in their respective places, and no confusion would appear likely to arise since each of the terms will be found to be self-explanatory in its use.

The parallel with the human case can still be drawn. The different suppliers are the sources from which the output or consignments of manufactured items (corresponding to some extent to the different ethnic groups) are supplied.

The customer's interest in the "populations", i.e. the outputs of the various suppliers who provide products of the type he desires to purchase, will depend on a number of considerations.

a) In certain cases, it will be essential for the variation in a characteristic to lie within narrowly defined limits. This will be so for the dimensions of component parts that need to be fitted together, for the analytical properties of certain chemical products, etc. The ideal, from the user's point of view, would probably be attained if the variation in items from all sources could be described by a normal curve with its mean on target and its standard deviation no greater than a certain value.

b) In other cases, wider latitude is permissible, so long as a minimum level is reached by virtually all the items; this is true, for example, when the qualities tested relate to strength or durability. For example, the average and standard deviation of the breaking strength of wire may differ considerably between suppliers, yet still meet the user's requirement.

c) Sometimes what is important to the user is not a particular mean value of a quality in a product, but a limit to the amount of variation about some mean which remains constant from one consignment to another. An example is products requiring craftsmanship in the finishing process, such as plasters or paints. The total material on the market may well be heterogeneous, consisting of several suppliers' outputs, all of which are of differing quality. However, the customer needs to be able to draw continually from one stable source of supply, i.e. to use material for which particular characteristics have a constant mean and low variation.

In all the above cases a statistical methodology is required which will indicate how best to determine from samples whether the output or consignment does in fact conform to the desired standard.

7.2 The basis of securing conformity to specification

7.2.1 *The two principal methods*

The provisions of a specification, the limits for the various quality characteristics, and the sampling technique to be adopted should be designed so as to provide assurance to the customer that each consignment or batch of material which he purchases is up to the stipulated standard. At the same time, the supplier will require to know that the standard prescribed is one which is consistent with the capability of his production processes, and which is economically feasible for him to maintain. There are two principal methods of securing conformity to a specification:

a) by a system of tests of samples taken from batches of finished material. In certain cases, these samples may be drawn at random from the whole bulk of material; in other cases, it may be necessary to take special precautions to ensure that representative samples are obtained. In either event, this method is called *acceptance sampling*;

b) by requiring that records be kept which will provide statistical evidence of both the control and the capability of the manufacturing processes. Such a procedure could form the basis of a guarantee system of specification, so long as occasional audits, independent of the supplier, are made in order to satisfy the certifying authority that the routine tests are actually being carried out in the production facility.

Both these methods can form the basis of a system of quality marking or guarantee to show conformance with a specification. Statistical theory can assist by providing the user with assurance as to the adequacy of the sampling, and the supplier with confidence that no unsuspected variations in his processes are affecting the quality of his product. In many cases, however, it will be found on statistical analysis that acceptance sampling on an adequate scale will be too cumbersome or expensive, or even quite impracticable, while the second method would appear likely to provide effectively for a guarantee system.

It is not appropriate to make too sharp a distinction between supplier and customer, as the supplier will not only be a user of raw materials but will also be interested in the range of quality on the market of the commodities that he is himself providing. Nevertheless, in making comparison of the two methods, it will be convenient to distinguish between questions of special importance to the customer and to the supplier.

7.2.2 *Considerations of importance to the customer*

The following considerations are of importance to the customer.

a) It has already been pointed out that in some cases it may be extremely difficult to obtain a sample that is representative of a consignment. This is well illustrated by the following example taken from Shewhart [1].

Given a consignment consisting of 10 truckloads of boxed material, there being 12 items in a box and roughly 1 000 boxes in a truck, how would it be possible to obtain a representative sample of these 120 000 items? Clearly, if the output were not homogeneous, certain boxes may contain articles of significantly different quality from others. According to the method of packing, these differences may be associated with certain trucks, or parts of a truck, or they may be scattered at random in the process of loading. Again, it may be the case that the articles at the bottom of each box are different from those at the top.

b) Statistical theory may show that the number of items that need to be tested to give the desired degree of information about the lot is prohibitive from an economic standpoint. This is particularly likely to be true where the test required is destructive and where, at the same time, there is considerable variation in the quality characteristic from item to item. For example, to burn out sufficient electric light bulbs to obtain a valid test of the difference in quality between the products of two manufacturers may not be financially viable in some cases.

c) For many processes, the process variation can be controlled successfully and the problem then becomes one of providing assurance that the process mean has not moved too far from the target value. For a continuing series of lots from the same source, the amount of random sampling necessary can be very much reduced if the process variation remains demonstrably constant over time (i.e. from hour to hour and day to day). As soon as it is established that the process standard deviation is constant at a given value, smaller sample sizes can be used on subsequent lots. The process variation would still need to be checked and more intensive sampling resumed if evidence emerges that the process variation is no longer stable.

To reduce his intensity of sampling inspection with safety, it is necessary for the user to know that effective quality control procedures are in place. If the supplier maintains quality control records, what is required is an agreed means of making this information available to the customer. If access to such records is available, the question then naturally arises whether the guarantee system of securing conformity to specification would not be far more satisfactory than the method of testing samples from consignments. This is especially true for those materials for which inspection of the final product entails elaborate and costly procedures.

7.2.3 *Considerations of importance to the supplier*

The supplier is concerned with the day-to-day routine problem of turning out goods that will satisfy the requirements of a specification. As a more distant objective, probably involving research and experimentation, he aims to reduce variation and increase the efficiency of the production process. Points he will consider are as follows.

a) If acceptance sampling is specified, the supplier who does not realise the waywardness of chance when dealing with variable material may find a sample from his product unexpectedly failing to pass specification. If, however, he has studied and measured this variability, he may judge at what level quality should be maintained in order to reduce the risk of rejection to an acceptable level. Without this knowledge, the quality level he is maintaining for safety may in fact be uneconomically high.

b) The form of routine control required under a guarantee system of specification, depending on tests analysed on a statistical basis, is no different from that which is necessary to assure the same level of safety under an acceptance sampling system.

c) Stability in the quality of a manufactured product has a number of advantages to the supplier. Besides its relation to sales owing to user confidence, it may have an important bearing on the economy of management. The following example of a problem that might arise in the production of high-grade cotton fabric illustrates this point.

Owing to uncontrollable faults, a certain percentage of the lengths turned out by looms always needs to be put into a lower quality grade. If this percentage reaches a high figure, the manufacturer is faced with the necessity of disposing of this unwanted burden of low quality material, which is attached to his high-grade produce as an awkward but unavoidable shadow. Clearly, fluctuations in the magnitude of this percentage figure will upset his costing forecasts.

d) The concept of statistical uniformity has so far in this publication mainly been associated with the stability of variation in time. But the methods of statistical analysis which need to be used to decide whether variation is statistically uniform will also be invaluable to the supplier in research and development when he is attempting to reduce variability and to detect and eliminate sources of trouble. For these aspects of the problem, reference should be made to textbooks, journal articles and standards on statistical process control, a brief selection of which may be found in the Bibliography. This is not strictly a problem of securing conformity with a specification, but some indication of its treatment is given in clause **10**.

From these considerations, it will be seen that there are clear advantages in the second method of securing conformity to a specification, namely by requiring that definite evidence be furnished of effective process control during manufacture. For this purpose, statistical theory can suggest systems for routine tests in the workplace. These will go far to arm a certifying authority with competence to assess the quality level of a product that is sold under a quality mark or guarantee.

In conclusion, the advantages of this method may be summarized as follows.

a) It avoids the difficulty that often arises of determining how to draw a representative sample from a lot or consignment.

b) It saves the cost of sampling on the large scale often necessary to give adequate assurance.

c) The amount of sampling necessary will be far less than that required to provide definite protection in the face of erratic quality levels. This is generally true even where it is desirable to carry out occasional tests on samples from lots to gain assurance that the process control remains effective.

d) The form of routine statistical analysis necessary to provide the basis of a system of quality certification is that which a supplier would employ anyway in attempting to increase the efficiency of his production process.

8 The statistical relationship between sample and population

8.1 The variation of the mean and the standard deviation in samples

8.1.1 *General*

In the preceding clauses it has been explained why many of the problems that arise in attempting to achieve effective standardization of production and conformity to specification are essentially statistical in nature. A detailed development here of the relevant statistical theory would be inappropriate, but it is necessary to outline sufficient elements of this theory to clarify the treatment of some typical problems.

Suppose that the variation in initial efficiency of a specified type of electric light bulb is under consideration. If all the light bulbs in a lot of several thousand were tested, it would be possible to calculate the mean efficiency and the standard deviation of efficiency, measured in lumens per watt, for the whole lot. If, however, tests were only made on several samples each consisting of 10 lamps, then a different mean and a different standard deviation would be obtained for each sample. Not only would these means and standard deviations differ among themselves from sample to sample, but also they would not correspond exactly to the values that could, in theory, be determined from the whole lot. It is clearly important to have some means of defining the extent of the differences that can arise through the chance fluctuations of sampling. Some further mathematical results need to be introduced in order to be able to do this in a precise manner.

To avoid ambiguity, it is essential to make a clear distinction in the notation used for the characteristics of the population (lot, consignment) and that used for the characteristics of a sample drawn from this population. The most common notation for this is as follows. For the population, containing N items, the mean is denoted by μ and the standard deviation by σ. For a sample, containing n items, the mean is denoted by $\bar{x}$ and the standard deviation by s. Values of the sample characteristics for different samples are identified by the use of subscripts, for example:

1st sample, size n_1, mean = $\bar{x}_1$, standard deviation = s_1;

2nd sample, size n_2, mean = $\bar{x}_2$, standard deviation = s_2;

3rd sample, size n_3, mean = $\bar{x}_3$, standard deviation = s_3;

4th sample, size n_4, mean = $\bar{x}_4$, standard deviation = s_4.

If a number of random samples of the same size n were drawn from the population, the standard deviation of the resulting values of $\bar{x}$ would be a measure of the magnitude of the error likely to be involved in using the mean of just one sample of n items as an estimate of the population mean μ. This standard deviation of the sample means $\bar{x}_1$, $\bar{x}_2$, ..., etc., is called the standard error of the mean. Similarly, the standard deviation of the sample standard deviations s_1, s_2, ..., etc., is called the standard error of the standard deviation, since it measures the error involved in using s as an estimate of σ.

In general it will not be practicable to take more than one random sample from the population. Fortunately, statistical theory comes to our aid by providing a means of estimating both of these standard errors from the results of one sample. It will be assumed in **8.1.2** and **8.1.3** that the sample size n is small compared to the population size N, say less than one twentieth of N.

8.1.2 *Variation of means*

The average of the sample means from all possible samples of size n from the population equals the population mean, i.e.

$$\text{Mean of } \bar{x} = \mu \tag{1}$$

In fact, to make it clear that the left-hand side of this equation is a population mean rather than a sample mean, a better notation is:

$$\mu_{\bar{x}} = \mu \tag{2}$$

The reliability of the mean varies directly as the standard deviation of the characteristic, x, in the population and inversely as the square root of the sample size, i.e.

$$\text{Standard deviation of } \bar{x} \text{ (i.e. standard error of the mean)} = \frac{\sigma}{\sqrt{n}} \tag{3}$$

Again, to make it clear that the left-hand side represents a population standard deviation rather than one based on a sample, a better notation is:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{4}$$

The variation in the sample means will approximate to a normal distribution except in cases of extremely asymmetrical variation in $\bar{x}$.

In simple terms, the standard error may therefore be interpreted as follows, provided that the sample contains some 20 or more items. Since $\bar{x}$ varies approximately in accordance with a normal distribution about μ with a standard error of $\sigma/\sqrt{n}$, it follows that it is rather unlikely that in any particular random sample the magnitude of the difference $(\bar{x} - \mu)$ will be greater than $2\sigma/\sqrt{n}$, and very unlikely that it will be greater than $3\sigma/\sqrt{n}$. Consequently, when only the data obtained from a sample are available, there is reasonable assurance that the population, lot or consignment mean will not differ from the sample mean $\bar{x}$ by more than ± 2 to ± 3 times $\sigma/\sqrt{n}$. If the standard deviation, σ, is not known from past experience, an estimate of σ obtained from the sample needs to be used. More precisely, tables of multipliers can be derived from the theory of sampling, such as Tables 8 and 9, whose uses are described in **8.1.3**.

8.1.3 *Variation of standard deviations*

The square of a standard deviation is called a variance, i.e. s^2 is the sample variance and σ^2 is the population (lot or consignment) variance. The average of the sample variances over all possible samples of size n from the population equals the population variance, i.e.

$$\mu_{s^2} = \sigma^2 \tag{5}$$

Unfortunately, there is no such simple result for the average of all possible sample standard deviations for samples of size n. In fact, the average value of s is less than σ. This effect is described as a bias, a negative bias in this case. The bias depends on the sample size, and gets smaller as the sample size increases.

Table 8 — Factors for confidence limits for the population mean and population standard deviation

Sample size	Mean				Standard deviation							
	Limits $\mu_1 = \bar{x} - as$, $\mu_2 = \bar{x} + as$				Limits $\sigma_1 = b_1 s$, $\sigma_2 = b_2 s$							
	Chance of error				Chance of error							
	10 %	5 %	2 %	1 %	10 %		5 %		2 %		1 %	
n	a	a	a	a	b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2
5	0.954	1.242	1.676	2.060	0.649	2.373	0.599	2.874	0.548	3.670	0.518	4.396
6	0.823	1.050	1.374	1.647	0.672	2.090	0.624	2.453	0.575	3.004	0.546	3.485
7	0.735	0.925	1.188	1.402	0.690	1.916	0.644	2.203	0.597	2.623	0.568	2.980
8	0.670	0.837	1.060	1.238	0.705	1.798	0.661	2.036	0.615	2.377	0.587	2.661
9	0.620	0.769	0.966	1.119	0.718	1.712	0.675	1.916	0.631	2.205	0.603	2.440
10	0.580	0.716	0.893	1.028	0.729	1.646	0.687	1.826	0.644	2.077	0.617	2.278
11	0.547	0.672	0.834	0.956	0.739	1.594	0.698	1.755	0.656	1.978	0.630	2.154
12	0.519	0.636	0.785	0.897	0.747	1.551	0.708	1.698	0.667	1.899	0.641	2.056
13	0.495	0.605	0.744	0.848	0.755	1.516	0.717	1.651	0.676	1.834	0.651	1.976
14	0.474	0.578	0.709	0.806	0.762	1.486	0.724	1.612	0.685	1.780	0.660	1.910
15	0.455	0.554	0.678	0.769	0.768	1.460	0.732	1.578	0.693	1.734	0.668	1.854
16	0.439	0.533	0.651	0.737	0.774	1.438	0.738	1.548	0.700	1.694	0.676	1.806
17	0.424	0.515	0.627	0.709	0.780	1.418	0.744	1.522	0.707	1.660	0.683	1.764
18	0.411	0.498	0.606	0.684	0.785	1.401	0.750	1.500	0.713	1.629	0.689	1.728
19	0.398	0.482	0.586	0.661	0.789	1.385	0.755	1.479	0.719	1.602	0.696	1.696
20	0.387	0.469	0.568	0.640	0.793	1.371	0.760	1.461	0.724	1.578	0.701	1.667
21	0.377	0.456	0.552	0.621	0.797	1.358	0.765	1.445	0.729	1.557	0.707	1.641
22	0.367	0.444	0.537	0.604	0.801	1.346	0.769	1.430	0.734	1.537	0.712	1.617
23	0.359	0.433	0.524	0.588	0.805	1.336	0.773	1.416	0.738	1.519	0.716	1.596
24	0.350	0.423	0.511	0.574	0.808	1.326	0.777	1.403	0.743	1.502	0.721	1.576
25	0.343	0.413	0.499	0.560	0.811	1.317	0.780	1.392	0.747	1.487	0.725	1.559
26	0.335	0.404	0.488	0.547	0.814	1.309	0.784	1.381	0.751	1.473	0.729	1.542
27	0.329	0.396	0.478	0.535	0.817	1.301	0.787	1.371	0.754	1.460	0.733	1.527
28	0.322	0.388	0.468	0.524	0.820	1.293	0.790	1.362	0.758	1.448	0.737	1.513
29	0.316	0.381	0.459	0.514	0.823	1.287	0.793	1.353	0.761	1.437	0.741	1.499
30	0.311	0.374	0.450	0.504	0.825	1.280	0.796	1.345	0.764	1.427	0.744	1.487

NOTE The fact that s is a biased estimator of σ although s^2 is an unbiased estimator of σ^2 may seem puzzling at first, but it should be remembered that the mean value of a set of numbers is not the same as the square root of the mean of their squares. For example:

(1/5) (1 + 3 + 3 + 5 + 6) = 3.6 whereas $\sqrt{(1/5)(1 + 9 + 9 + 25 + 36)} = 4.0$.

The average of the sample standard deviations over all possible samples of size n from the population is given by the following equation:

$$\mu_s = c_4\sigma \qquad (6)$$

where c_4 is the bias factor and depends on the value of n.

NOTE Relation (6) is only true if the variation among the observations is of the normal form.

Values of c_4 are given for sample sizes from 2 to 30 in Table 9; note that c_4 is approximately equal to $4(n - 1)/(4n - 3)$.

Table 9 — Factors for removing bias from sample standard deviations

Sample size, n	c_4	$1/c_4$	Sample size, n	c_4	$1/c_4$	Sample size, n	c_4	$1/c_4$
			11	0.975 4	1.025 3	21	0.987 6	1.012 6
2	0.797 9	1.253 3	12	0.977 6	1.023 0	22	0.988 2	1.012 0
3	0.886 2	1.128 4	13	0.979 4	1.021 0	23	0.988 7	1.011 4
4	0.921 3	1.085 4	14	0.981 0	1.019 4	24	0.989 2	1.010 9
5	0.940 0	1.063 8	15	0.982 3	1.018 0	25	0.989 6	1.010 5
6	0.951 5	1.050 9	16	0.983 5	1.016 8	26	0.910 1	1.010 0
7	0.959 4	1.042 4	17	0.984 5	1.015 7	27	0.990 4	1.009 7
8	0.965 0	1.036 2	18	0.985 4	1.014 8	28	0.990 8	1.009 3
9	0.969 3	1.031 7	19	0.986 2	1.014 0	29	0.991 1	1.009 0
10	0.972 7	1.028 1	20	0.986 9	1.013 2	30	0.991 4	1.008 7

If the sample standard deviation is to be used as an estimate of the population value, for some applications it is desirable or customary to eliminate the bias. Following Shewhart [2], this is done by taking s/c_4 as the estimate of σ.

With regard to these corrections, the following points should be noted.

a) Unless the sample contains very few items (i.e. unless n is very small), the bias is inconsequential.

b) If n is small, no single estimate of σ can be regarded as satisfactory; what is required is a pair of lower and upper limits σ_1 and σ_2 within which we may feel confident that σ lies. These are so-called "confidence limits", for the calculation of which Table 8 has been given. This table shows very clearly the extent of the uncertainty that remains when σ is estimated from a few observations only.

c) Corrections in the case where σ is estimated from a *number* of small samples are, however, important, for in this case a really reliable unbiased estimate is possible. These corrections are discussed further in **10.7**.

Corresponding to equation (3) there is an approximate theoretical expression for the standard error of a standard deviation, namely:

$$\text{standard deviation of } s \text{ (i.e. standard error of the sample standard deviation)} = \frac{\sigma}{\sqrt{2(n-1)}} \quad (7)$$

i.e.:

$$\sigma_s = \frac{\sigma}{\sqrt{2(n-1)}} \quad (8)$$

The accuracy of this approximation is subject to several limitations:

a) the variation among the observations, x, has to approximately follow the normal distribution;

b) as σ will generally not be known, it will be necessary to substitute s in the right-hand side of (8), so the sample should consist of at least 30 observations.

Subject to these restrictions, the result is nevertheless helpful in giving some idea of the reliability of s as an estimate of σ.

The data in Table 10 illustrates the variation in $\bar{x}$ and s among samples from the same population. A can of tomatoes is taken from a production line once every two hours, and its contents weighed. Four observations therefore become available every 8-hour shift. Observations over 40 shifts are given in the table.

Table 10 — Weight of contents of tomato cans (g)

Shift	Weight of can contents (g)				Shift	Weight of can contents (g)			
1	401.5	401.5	404.8	402.8	21	405.0	405.7	404.1	404.4
2	404.4	403.4	406.3	403.1	22	403.8	405.4	406.5	401.3
3	405.7	405.5	406.1	404.8	23	401.8	404.4	407.6	405.6
4	405.0	402.6	406.6	402.9	24	402.4	401.8	403.8	401.6
5	402.6	404.0	404.4	404.0	25	407.3	404.1	406.3	403.1
6	404.2	403.6	403.7	407.9	26	401.4	407.4	402.1	404.4
7	404.4	405.2	402.5	403.5	27	402.8	403.7	405.5	402.4
8	407.7	403.9	403.8	407.1	28	401.6	406.5	400.8	404.1
9	409.7	400.7	405.0	405.5	29	407.3	401.3	406.1	405.9
10	405.7	400.4	402.3	405.4	30	401.2	405.3	405.2	403.2
11	402.8	403.2	402.3	402.0	31	408.4	403.3	404.1	402.9
12	400.8	406.3	403.6	402.6	32	404.8	404.9	406.0	404.5
13	401.0	403.9	403.0	403.4	33	403.2	402.0	403.4	404.0
14	402.3	405.6	402.5	404.8	34	404.9	400.9	400.9	400.4
15	403.7	404.7	405.8	403.9	35	405.0	402.1	405.6	402.0
16	403.9	402.2	403.7	402.7	36	402.1	403.1	403.8	404.2
17	404.2	404.9	406.3	401.4	37	405.3	403.9	404.7	404.3
18	403.6	404.0	401.0	400.9	38	404.5	401.5	404.7	402.7
19	405.9	403.8	405.6	398.4	39	405.2	399.7	405.1	406.2
20	401.5	401.7	404.0	403.8	40	402.0	400.7	402.6	404.9

Statistical analysis of the type described in **4.4** shows that the variation from item to item is under statistical control during the period covered by the first 40 shifts.

The mean and standard deviation for the 160 weights are μ = 403.84 g and σ = 1.909 g.

The means and standard deviations, $\bar{x}$ and s, of the 40 samples each consisting of 4 test results, are shown in Table 11. They have been grouped together in Table 12 where they form frequency distributions analogous to that in Table 6.

Table 11 — Canned tomatoes data — Mean and standard deviation of four weights per shift (g)

Shift	Mean	s.d.	Shift	Mean	s.d.	Shift	Mean	s.d.	Shift	Mean	s.d.
1	402.65	1.559	11	402.58	0.532	21	404.80	0.707	31	404.68	2.533
2	404.30	1.445	12	403.32	2.297	22	404.25	2.258	32	405.05	0.656
3	405.52	0.544	13	402.82	1.271	23	404.85	2.424	33	403.15	0.839
4	404.28	1.882	14	403.80	1.651	24	402.40	0.993	34	401.78	2.097
5	403.75	0.790	15	404.52	0.954	25	405.20	1.936	35	403.68	1.893
6	404.85	2.050	16	403.12	0.810	26	403.82	2.706	36	403.30	0.920
7	403.90	1.163	17	404.20	2.061	27	403.60	1.378	37	404.55	0.597
8	405.62	2.065	18	402.38	1.654	28	403.25	2.583	38	403.35	1.526
9	405.22	3.680	19	403.42	3.476	29	405.15	2.640	39	404.05	2.942
10	403.45	2.549	20	402.75	1.333	30	403.72	1.941	40	402.55	1.756

Table 12 — Canned tomatoes data — Frequency distribution of individual observations and of means and standard deviations of 4 tests

Weight (g)	Frequency	Mean, $\bar{x}$ (g)	Frequency	Standard deviation, s (g)	Frequency
398.00–398.99	1	400.50–400.99	1	0.500–0.999	11
399.00–399.99	1	401.00–401.49	2	1.000–1.499	5
400.00–400.99	9	401.50–401.99	5	1.500–1.999	9
401.00–401.99	16	402.00–402.49	8	2.000–2.499	7
402.00–402.99	26	402.50–402.99	7	2.500–2.999	6
403.00–403.99	30	403.00–403.49	5	3.000–3.499	1
404.00–404.99	32	403.50–403.99	6	3.500–3.999	1
405.00–405.99	25	404.00–404.49	4		
406.00–406.99	11	404.50–404.99	2		
407.00–407.99	7				
408.00–408.99	1				
409.00–409.99	1				
Total	160	Total	40	Total	40

If the mean and the standard deviation of these distributions are calculated, they may be compared with the theoretical values obtained by inserting the values of μ and σ in relations (2), (4), (6) and (7), as shown in Table 13.

Table 13 — Canned tomatoes data — Comparison of the sample mean and sample standard errors of the means and standard deviations in groups of 4 tests with theoretical results

Measure	Results from Table 9	Results from theoretical formulae
Mean of $\bar{x}$, i.e. μ_x	403.84 g*	403.84 g*
Standard error of $\bar{x}$, i.e. $\sigma_{\bar{x}}$	0.965 g	$\sigma/\sqrt{n} = 1.909/\sqrt{4} = 0.954$ g
Mean of s, i.e. μ_s	1.727 g	$4(n-1)\sigma/(4n-3) = 12 \times 1.909/13 = 1.762$ g
Standard error of s, i.e. σ_s	0.808 g	$\sigma/\sqrt{2(n-1)} = 1.909/\sqrt{6} = 0.779$ g

* Note that these two values necessarily agree, as the mean of the 40 sample means has to equal the mean of the population of 160 tests.

Corresponding figures are seen to be in quite close agreement.

8.2 The reliability of a mean estimated from representative and duplicate sampling

8.2.1 *Representative sampling*

An important extension to some of the previous results is necessary when the items on which the observations or tests have been made are drawn from a number of sources, within each of which there is statistical control, but between which there may be differences in the process averages and variances.

Suppose a population (batch, consignment) consists of a large number of items, of which a proportion p_1 has come from one source, p_2 from a second, etc., and finally p_k from a k^{th} source. Suppose that the mean and standard deviation of the quality characteristic under consideration in material from the first source are μ_1 and σ_1, for the second source are μ_2 and σ_2, etc. The mean for the whole population will then be:

$$\mu = p_1\mu_1 + p_2\mu_2 + \ldots + p_k\mu_k \quad (9)$$

What has been described as a representative sample of n items can then be drawn by taking n_1 items at random from the material coming from source 1, n_2 from the material from source 2, and so on, where:

$$n_1 = np_1,\ n_2 = np_2,\ \ldots,\ n_k = np_k \quad (10)$$

and $n_1 + n_2 + \ldots + n_k = n$ since the proportions sum to unity, i.e. $p_1 + p_2 + \ldots + p_k = 1$.

If $\bar{x}$ is the mean value of the characteristic for the n items of this representative sample, then it may be used as an estimate of the true population mean μ. It is known that the standard error (standard deviation in repeated samples) of $\bar{x}$ is given by the relation:

$$\text{standard error of } \bar{x} = \frac{1}{n}\sqrt{n_1\sigma_1^2 + n_2\sigma_2^2 + \ldots + n_k\sigma_k^2} \quad (11)$$

In practice, it will commonly happen that the standard deviations within the different sources of supply will be approximately the same, i.e. $\sigma_1 = \sigma_2 = \ldots = \sigma_k = \sigma$. Relation (11) then simplifies to:

$$\text{standard error of } \bar{x} = \frac{\sigma}{\sqrt{n}} \quad (12)$$

Results (11) and (12), like (1) to (5), are independent of any assumption of normality in the variation. A number of points of practical importance may be deduced.

a) Provided that the sample can be made properly representative by drawing sub-samples that satisfy condition (10), the reliability of the estimate $\bar{x}$ depends only on the variation *within* each source of supply, and not upon the differences between the mean values $\mu_1, \mu_2, \ldots, \mu_k$.

b) If the values of $\sigma_1, \sigma_2, \ldots, \sigma_k$ are not known from previous experience, they may be estimated from the standard deviations of the sub-samples. In particular, if it is believed that $\sigma_1 = \sigma_2 = \ldots = \sigma_k = \sigma$, then a rapid estimate of σ, which will be generally adequate, can be obtained from the *range* of variation in each of the sub-samples [see **13.3.4**e)].

c) If no attempt is, or can be, made to draw a representative sample satisfying conditions (10), then the mean μ of the population will be estimated with less precision. The whole population of items then needs to be regarded as a single group having a standard deviation σ'. The mean $\bar{x}$ of the n observations in the sample drawn at random from the whole population will have a standard error given by formula (3) in **8.1.2**, i.e.:

$$\text{standard error of } \bar{x} = \frac{\sigma'}{\sqrt{n}} \quad (13)$$

In this case it can be shown that:

$$\sigma' = \sqrt{p_1\sigma_1^2 + p_2\sigma_2^2 + \ldots + p_k\sigma_k^2 + p_1(\mu_1 - \mu)^2 + p_2(\mu_2 - \mu)^2 + \ldots + p_k(\mu_k - \mu)^2} \quad (14)$$

which, when $\sigma_1 = \sigma_2 = \ldots = \sigma_k = \sigma$, simplifies to:

$$\sigma' = \sqrt{\sigma^2 + p_1(\mu_1 - \mu)^2 + p_2(\mu_2 - \mu)^2 + \ldots + p_k(\mu_k - \mu)^2} \quad (15)$$

a quantity clearly at least equal to σ. Hence, if $\mu_1, \mu_2, \ldots, \mu_k$ are not all equal to μ, i.e. the process average changes from one source of supply to another, the mean of the batch will be estimated with less precision.

d) Even when no exact attempt is made to estimate the values of $\sigma_1, \sigma_2, \ldots, \sigma_k$, representative sampling is often employed to ensure that the resulting estimate of the population mean is as reliable as possible. In other words, although no calculations of reliability are made, sampling is in fact planned so that the standard error is given by (11) or (12) rather than (13) with (14) or (15).

These points may be illustrated by the following numerical example. The strength and other properties of bricks depend to some extent on the position of the bricks during firing in the kiln. An investigation has shown that in a particular case the standard deviation of dry strength for the whole batch of bricks from a single firing of a kiln was given by:

$\sigma' = 1\,283$ lbf/in^2

If, however, the cross section of the kiln was divided into nine areas, the averaged standard deviation of brick strengths in a single area was the following:

$\sigma = 737$ lbf/in^2

It follows that the producer, taking say four bricks at random from each of the areas, could obtain from the mean of the 36 test results an estimate of the mean brick strength of the kiln, having a standard error of the following:

$$\frac{\sigma}{\sqrt{36}} \text{ 123 lbf/in}^2$$

NOTE It may be helpful to see this result obtained in two steps, as follows. The standard error for the mean result of tests made on four bricks from any one position in the kiln is $737/\sqrt{4} = 368.5$ lbf/in^2. As nine averages of similar tests are then themselves averaged, the standard error for the mean of the 36 tests will be $368.5/\sqrt{9} = 123$ lbf/in^2.

The user, on the other hand, does not have the opportunity to obtain a representative sample in this way. It is likely, but by no means certain, that neighbouring bricks in the consignment he receives will have come from the same parts of the kiln. The user therefore needs to take a sample from the whole consignment and associate a standard error of $\sigma'/\sqrt{n}$ with the resulting estimate of the mean strength. To obtain an estimate about as reliable as that of the producer, which was based on 36 bricks, the user needs to test about $n = 110$ bricks because, approximately:

$$\frac{\sigma'}{\sqrt{110}} = 123 \text{ lbf/in}^2.$$

It needs to be understood that the above argument holds only when principally it is the *mean* value of a characteristic that it is desired to control. In the example taken for illustration, the mean strength of a consignment of bricks is not in fact the best criterion for assessing quality. Both the mean strength and the standard deviation of strength require control, as they both relate to the proportion of bricks that are below a given strength (see **8.5** and **9.5.2**). The question of representative sampling applied to the simultaneous estimation of the mean and the standard deviation involves other considerations, which cannot be entered into here. The comparison given, however, expresses in numerical form the advantage that follows if process data is obtained at the time of production rather than by the sampling of consignments.

8.2.2 *Duplicate sampling*

For certain products it is the practice not to measure the characteristics of each individual item in the sample but to record a single value which is the grand total of the individual values. For example, the total weight of a sample of n articles may be taken, but not the n separate weights. For sampling bulked materials such as coal, cement and oil, the sampling methods may aim only at obtaining a single total measure as an estimate of quality. However, without additional information, it is impossible to determine the reliability of this single measure. This is because even if results from a number of consignments are collected and compared, the variation between them may be due to changes in the process mean value and not to sampling error.

The only satisfactory method of determining the reliability of a sampling procedure is for the same sampling procedure to be carried out independently several times on the same consignment or batch, and for the standard deviation of these independent results to be obtained. If the process variation remains approximately stable, then the reliability of the sampling procedure may be examined initially and rechecked only occasionally. An economic method of maintaining assurance of the continued reliability is to arrange that independent *duplicate* samples be taken. For example, as described in **6.2** for iron ore, for some products a number of small portions may be taken at regular intervals from a conveyor, alternate portions put into two separate receptacles, and the process of mixing, quartering, etc. and final analysis performed independently in duplicate.

Suppose that x_1 and x_2 are the two test results that are to be used as an estimate of the real quality of the consignment. Their difference may be expressed as $d = x_1 - x_2$. From statistical theory, it is known that the standard deviation of d in the consignment may be expressed in terms of the standard deviation of x in the consignment by the formula:

$$\sigma_d = \sqrt{2}\sigma_x \qquad (16)$$

This result remains true even if the process mean value changes from one consignment to another, provided:

a) that the variation about the mean in the parts of the sampled bulk is approximately the same in all consignments; and

b) that the duplicate samples are independent, e.g. if x_1 is above the real consignment value, then x_2 is as likely to be below as to be above.

It follows from (16) that, if these conditions are satisfied, $\sigma_d/\sqrt{2}$ may be used as a measure of σ_x, the standard error of either of the estimates x_1 and x_2. As it is known that the standard error of the mean of x_1 and x_2 is $\sigma_x/\sqrt{2}$, it follows that $\sigma_d/2$ may be used for the standard error of $\bar{x} = (x_1 + x_2)/2$. In order to keep a check on the continued reliability of the sampling and analytical processes, the successive values of $d = x_1 - x_2$ may be plotted on a control chart (see **10.7**).

8.3 Illustration of the use of the mean weight, and the lowest weight, in a sample of prescribed size of standard specimens of fabric

The principles considered in **8.1** and **8.2** are of equal importance whether the assistance of statistical theory be required in connection with the method of consignment sampling or in assessing how well the process mean and variation are being controlled.

The following is an illustration of the use of the standard error of the mean, i.e. $\sigma/\sqrt{n}$, in consignment sampling. Example 2 in **4.3** gave some figures for the weights of 128 standard specimens from a roll of fabric. A potentially large-scale user first investigates the quality of such fabrics in the marketplace. He then decides that the specification and method of sampling should be as follows: while it will penalize occasionally the producer whose fabric has an average weight of $\mu = 100$ and a standard deviation of individual test specimens of $\sigma = 3.5$, it will penalize less and less frequently as quality improves above this level. He intends to do this by introducing a test clause into the specification such that if the tests on the sample material fail to pass the standard, the whole roll of fabric from which the sample has been taken is to be rejected. Suppose that it is under discussion whether the tests on each roll should be made on $n = 4$, 8 or 16 specimens, and that it is proposed that the clause should specify a *minimum weight which the mean of n tested specimens has to exceed.* The problem is how to determine what this minimum average weight should be in each case.

If the production process is under statistical control, we know that the means of samples of n pieces will be closely represented by a normal curve with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

Table 14 — Fractiles of the normal distribution corresponding to selected confidence levels

Confidence %	Chance of error %	α	One-sided interval u	$\alpha/2$	Two-sided interval u
90	10	0.10	1.281 6	0.05	1.644 9
95	5	0.05	1.644 9	0.025	1.960 0
98	2	0.02	2.053 7	0.01	2.326 3
99	1	0.01	2.326 3	0.005	2.575 8

From Table 14, which gives details of the fractiles of the standard normal probability curve, it may be expected in the long run that, for example:

10 percent of means (i.e. 1 in 10) will fall below $\mu - 1.281\,6\sigma/\sqrt{n}$;

5 percent of means (i.e. 1 in 20) will fall below $\mu - 1.644\,9\sigma/\sqrt{n}$;

1 percent of means (i.e. 1 in 100) will fall below $\mu - 2.326\,3\sigma/\sqrt{n}$; and

0.5 percent of means (i.e. 1 in 200) will fall below $\mu - 2.575\,8\sigma/\sqrt{n}$.

He decides to fix the minimum so that a producer whose standard of quality is represented by $\mu = 100$, $\sigma = 3.5$ will be liable to have only one sample in 20 rejected. The limits were therefore set as follows.

For samples of size 4, $L_4 = 100 - 1.644\,9 \times 3.5/\sqrt{4} = 97.12$.

For samples of size 8, $L_8 = 100 - 1.644\,9 \times 3.5/\sqrt{8} = 97.96$.

For samples of size 16, $L_{16} = 100 - 1.644\,9 \times 3.5/\sqrt{16} = 98.56$.

These limits are shown in the left hand side of Table 15 together with the number of samples, represented in Figure 5 in **4.3.2**, which would be *rejected* if these specification limits were imposed.

Suppose that it was decided to demand a higher quality of material having a mean weight per standard specimen of at least 103, and as before a standard deviation not greater than $\sigma = 3.5$. The specification limits, which are now 3 higher than before, are shown on the right-hand side of Table 15, together with the number of samples (from Figure 5) which would now be *accepted.*

Table 15 — Fabric specimens — Testing rule based on sample mean

Size of sample	Case μ = 100		Case μ = 103	
	Limit	Number of samples in Figure 5 *rejected*	Limit	Number of samples in Figure 5 *accepted*
4	97.12	5 out of 32	100.12	17 out of 32
8	97.96	2 out of 16	100.96	3 out of 16
16	98.56	1 out of 8	101.56	0 out of 8

The advantages of adjusting the limits to suit the sample size are now evident. Note that the mean and standard deviation of the 128 test results are 99.91 and 3.49 respectively. The first case, with μ = 100, illustrates the way in which a producer whose material lies on the borderline will receive broadly similar treatment whatever the sample size. The second case vividly demonstrates how the user may protect himself against receiving material of quality inferior to the standard at which he aims by increasing the number of specimens to be subjected to the test.

In the earlier use of this data in **4.3**, it was suggested that the minimum criterion might be applied to the lowest weight in a sample of n test specimens instead of to the mean. If the weights vary according to a normal distribution with a mean μ and a standard deviation σ, then tables are available[1)] from which such limits can be determined; the limits are of the following form:

$$L = \mu - k\sigma$$

where the value of k depends upon the number, n, of items in the sample and the chance of rejection or of acceptance that it is decided to adopt.

Suppose, for example, that it were again decided to fix a minimum limit such that a producer conforming to a standard of quality represented by μ = 100, σ = 3.5 will tend to have one sample in 20 rejected. Then the appropriate factors, k, are shown in Table 16, together with the resulting critical limits and the number of rejections among the same series of samples, i.e. those of Figure 5 in **4.3.2**. The results of choosing a higher standard of quality, μ = 103, are shown in the right-hand side of the same table.

Table 16 — Fabric specimens — Testing rule based on smallest weight in sample

Size of sample	Factor k	Case μ = 100		Case μ = 103	
		Limit	Number of samples in Figure 5 *rejected*	Limit	Number of samples in Figure 5 *accepted*
4	2.234	92.18	2 out of 32	95.18	23 out of 32
8	2.490	91.28	1 out of 16	94.28	11 out of 16
16	2.726	90.46	0 out of 8	93.46	6 out of 8

It is instructive to compare Tables 15 and 16. As in the case of basing the test on the sample mean, the left-hand side of Table 16 shows that, even when basing the test on the *smallest* weight in the sample, a producer whose material lies on the borderline will receive broadly similar treatment whatever the sample size. The most noticeable difference between the tables is how much better the user safeguards himself against receiving material of inferior quality by using the sample *mean* weight rather than the *smallest* of the weights in the sample.

It would not be justifiable to draw general conclusions from a single practical example. This is particularly true with this example as the test specimens were cut from the same roll, so the variation from specimen to specimen would not have been entirely random. However, the general conclusions that this example suggests are in accordance with what would have been predicted by statistical theory.

The focus of this example has been on protecting the user against receiving material of low weight. Uniformity of weight may, however, be an important characteristic, that is to say it may also be desirable to protect against the acceptance of fabric with a large variation in weight from specimen to specimen. For this purpose, it would be possible to specify some upper limit either to the standard deviation, s, or to the range (i.e. the difference between the heaviest and lightest specimens) in a sample. The question of control of variation is discussed later in connection with control charts (see clause **10**).

[1)] See Bibliography.

8.4 Tests and confidence intervals for means and standard deviations

8.4.1 *Confidence intervals for means and standard deviations*

A sample provides an estimate of the mean and the standard deviation of a variable x in the population from which the sample is drawn, that is to say $\overline{x}$ and s provide estimates of μ and σ. It has been indicated that if the sample does not contain many items, then these estimates may not be very accurate; the inaccuracy is measured by the standard errors, expressions for which have been given in **8.1**. For some practical purposes a rather more precise method of expressing the uncertainty of estimation may be desirable. This can be provided by statistical theory, but only on certain assumptions which are summarized below, and which must not be overlooked in using Table 8. The problem and its solution may be put in the following form.

Given a sample of size n having, for a certain measured variable, a mean $\overline{x}$ and a standard deviation s, to determine:

a) the limits μ_1 and μ_2 between which the population mean μ is likely to lie; and

b) the limits σ_1 and σ_2 between which the population standard deviation σ is likely to lie.

The expression "is likely" needs to be defined in terms of probability. For example, the limits may be chosen in such a way that, using limits derived in the same way on repeated occasions, we would be wrong only one time in 50 (i.e. 2 % of the time) in the long run. Such limits may be calculated as follows:

for the population mean, $\mu_1 = \overline{x} - as$, $\mu_2 = \overline{x} + as$ (17)

for the population standard deviation, $\sigma_1 = b_1 s$, $\sigma_2 = b_2 s$ (18)

where the factors a, b_1 and b_2 are given in Table 8 for four levels of probability and for sample sizes n from 5 to 30.

For larger values of n the following approximations to a, b_1 and b_2 are reasonably accurate:

$$a = \frac{u}{\sqrt{n-3}}, \quad b_1 = \frac{1}{1 + \frac{u}{\sqrt{2n}}} \quad \text{and} \quad b_2 = \frac{1}{1 - \frac{u}{\sqrt{2(n-2)}}}$$

where the values of u are related to the chance of error for two-sided intervals for a standard normal distribution as shown in Table 8.

For example, the value of u for a 10 % chance of error with a two-sided confidence interval is 1.644 9. For a sample of size 30, the values of a, b_1 and b_2 for a 10 % chance of error are calculated from these approximations as 0.317, 0.825 and 1.282, which are reasonably close to the correct values 0.311, 0.825 and 1.280.

These limits on the population mean and population standard deviation are called *confidence limits*, because they are associated with a stated measure of confidence. If, for instance, we have a sample of 10 items, and assert that in the sampled population the mean lies in the range:

$\overline{x} - 0.580s$ to $\overline{x} + 0.580s$

then it can be seen from Table 8 that we should expect such predictions to be correct about 90 % of the time in the long run, and in error 10 % of the time. About half of the 10 % of erroneous assertions would be because $\overline{x} - 0.580s$ exceeds the population mean μ and the other half because $\overline{x} + 0.580s$ was less than μ. On the other hand, if we take the wider range:

$\overline{x} - 0.893s$ to $\overline{x} + 0.893s$

we shall be 98 % confident that we are correct, knowing that there is only a 1 % chance that $\overline{x} - 0.893s$ will exceed μ and a 1 % chance that $\overline{x} + 0.893s$ will be less than μ. The interpretation will be similar for the case of σ.

Table 8 was designed for two-sided intervals, but can be used for one-sided intervals simply by doubling the chance of error. For example, if an upper confidence limit on σ was required at 99 % confidence when the sample size is 15, the chance of error of 1 % is doubled to 2 % to locate the appropriate value of b_2, which is 1.734. Thus, we would have 99 % confidence that σ is less than $1.734s$.

The validity of such confidence limits depends upon certain assumptions, viz.:

1) *that the variation is under statistical control.* For instance, whilst the sample might be any one of those in case 1 or case 2 of Figure 25, we clearly could not expect to derive meaningful limits from any one of the samples in cases 3, 5 or 6;

2) *that the form of variation among the items is represented approximately by the normal curve* (see **5.3.8**);

3) *that the sample has been drawn at random from a much larger population.* This condition is generally satisfied if the sample size is no more than 5 % of the population size. If, for example, a sample of 10 items were to be drawn at random from a lot containing only 20 items, then it would be possible to estimate the μ and σ of this lot from the $\bar{x}$ and s of the sample within considerably narrower limits.

The first assumption is of particular importance. It cannot be emphasized too strongly that if the variation is not under statistical control then it is foolhardy to attempt to predict the characteristics of the population from a sample chosen at random.

In practice, statistical control will rarely be perfect, so it is advisable not to pay too much regard to the precise risks associated with the limits. It is better to regard the constants in Table 8 as part of a useful working tool, whose value will be tested by experience. For the same reason, although the constants are given to three decimal places of accuracy, some common sense is necessary in determining how many figures are worth retaining in the calculated confidence limits.

The following illustration is based on the canned tomatoes data given in Table 10. The unit of measurement throughout is the weight in grams.

The first group of three shifts provide a total of 12 observations, with $\bar{x} = 404.16$ and $s = 1.681$. If we assume that the process is in statistical control, we may use these values to define limits within which we would feel confident that the mean and standard deviation of production lies. Choosing a 98 % confidence level (i.e. a 2 % chance of error), we find from Table 8 that $a = 0.785$, $b_1 = 0.667$ and $b_2 = 1.899$, giving the following values:

i) for the mean of production, $404.16 \pm 0.785 \times 1.681$, i.e. 402.8 to 405.5;

ii) for the standard deviation of production, 0.667×1.681 to 1.899×1.681, i.e. 1.12 to 3.19.

The range of uncertainty is clearly very large. Adding further observations to the group can narrow this range. Suppose that we base the confidence limits on the first six shifts, doubling the number of observations to 24. Calculating from the original data in Table 10, it is found that $\bar{x} = 404.22$ and $s = 1.598$. The constants for 98 % confidence limits are found from Table 8 to be $a = 0.511$, $b_1 = 0.743$ and $b_2 = 1.502$, giving limits as follows:

I) for the mean, $404.22 \pm 0.511 \times 1.598$, i.e. 403.4 to 405.0;

II) for the standard deviation, 0.743×1.598 to 1.502×1.598, i.e. 1.19 to 2.40.

The extra information has narrowed the limits, as one might have expected. In both cases the limits include the values $\mu = 403.8$ and $\sigma = 1.91$ calculated from the first 160 test records.

8.4.2 *Tests for means and standard deviations*

8.4.2.1 *Terminology*

In constructing statistical tests, it is important to be precise about the hypotheses under consideration. Some terminology is helpful here. Generally speaking, the hypothesis of "no difference" or equality of population values is called the *null hypothesis*, and is usually denoted by H_0. The hypothesis against which this hypothesis is to be compared is called the *alternative hypothesis*, and is usually denoted by H_1. The test is performed on the value of a *test statistic* that is calculated from the sample data. The region of variation of the test statistic that leads to rejection of the null hypothesis is called the *critical region*. The probability that the test statistic falls in the critical region when the null hypothesis is true (thereby leading to the erroneous decision to reject the null hypothesis in favour of the alternative hypothesis) is called the *size* or *significance level* of the test, usually denoted by α. Finally, the probability that the test statistic falls in the critical region when the alternative hypothesis is true (thereby leading to the correct decision to reject the null hypothesis in favour of the alternative hypothesis) is called the *power* of the test. The power is usually denoted by $1 - \beta$. A good test will have high power and a low value for the significance level.

A rejection of the null hypothesis when the null hypothesis is true is called a Type I error, or an error of the first kind. A rejection of the alternative hypothesis when the alternative hypothesis is true is called a Type II error, or error of the second kind. It follows that the probabilities of Type I and Type II errors are α and β respectively.

8.4.2.2 *Test of a population mean against a given value*

A simple example will illustrate these concepts. Suppose that the standard deviation of a normal population, σ, is known but that the mean, μ, is unknown. We wish to test the null hypothesis:

H_0: the mean of the normal population is μ_0;

against the alternative hypothesis:

H_1: the mean of the normal population is greater than μ_0.

The significance level of the test is to be 5 %. A random sample of size n is drawn from the population, and its mean $\bar{x}$ calculated.

It is intuitively obvious for this example that the critical region should lie entirely to the right of $\mu_0 + c$, where c is some constant that is greater than zero. For a 5 % significance level, we require $\mu_0 + c$ to be the upper confidence limit on μ with a 95 % confidence level. The appropriate standard normal fractile for a one-sided confidence interval at confidence level 95 % is found from Table 8 to be 1.644 9. As $\bar{x}$ is normally distributed with a mean μ_0 and a standard deviation $\sigma/\sqrt{n}$ under hypothesis H_0, it follows that $c = 1.644\,9 \times \sigma/\sqrt{n}$.

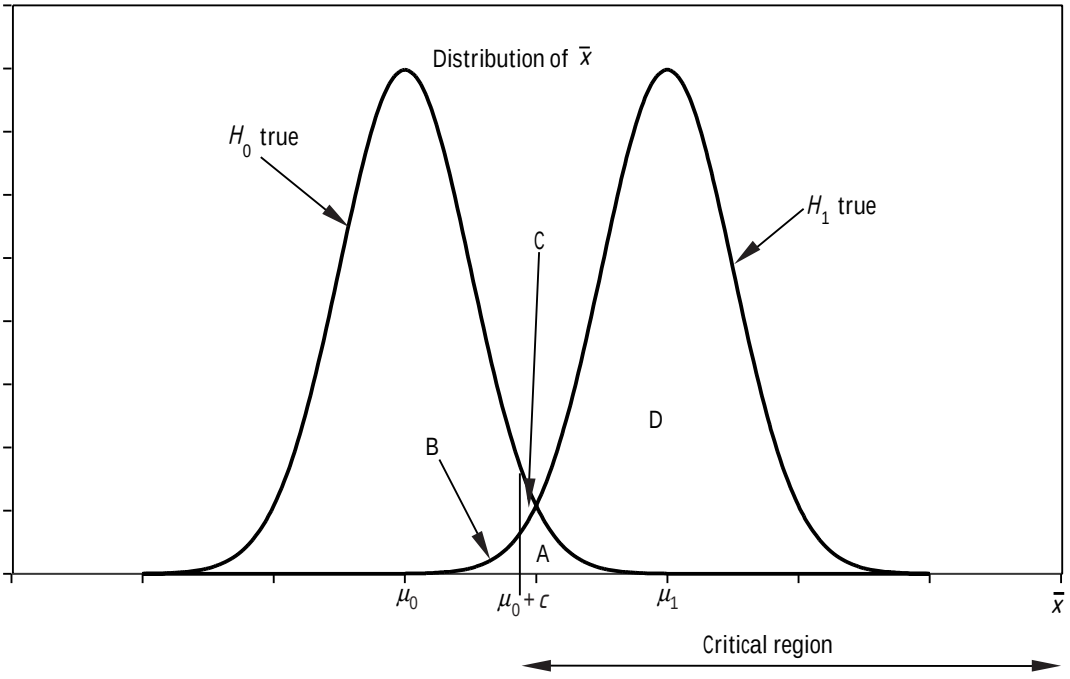


Figure 27 — Illustration of one-sided test

Figure 27 shows the critical region. The area A + C represents the significance level, in this case 5 % or 0.05. The area A + D (= 1 − B) represents the power of the test when $\mu = \mu_1$. To calculate the power of the test we first calculate the standardized difference between $\mu_0 + c$ and μ_1, i.e. $z = \{(\mu_0 + c) - \mu_1\}/(\sigma/\sqrt{n})$. The power of the test is then the area to the right of z under the standard normal curve, which may be found from Table 7, for example.

A more common situation would be where the population mean and standard deviation are both unknown. In this case the multiplier $1.644\,9/\sqrt{n}$ in the above (one-sided) example would be replaced by the value of a given in Table 8 corresponding to the sample size and a $2 \times 5\,\% = 10\,\%$ chance of error. Thus, for sample size 9 the multiplier $1.644\,9/\sqrt{9} = 0.548\,3$ would be replaced by 0.620. The increase in the value of the multiplier reflects the increased uncertainty due to not knowing the value of σ.

8.4.2.3 *Test of the difference between two population means; degrees of freedom*

A few remarks are in order at this point about the concept of *degrees of freedom*. In a single sample of size n, the deviations from the sample mean are restricted in the sense that they need to sum to zero; once $n - 1$ of the deviations are known, the n^{th} can be calculated, so the degrees of freedom are $v = n - 1$. If the sample values are considered as the co-ordinates of a point in n-dimensional space, the deviations from $\bar{x}$ are all constrained to lie in an $(n - 1)$-dimensional plane. The simplest example is when $n = 2$; if $x_1 - \bar{x}$ is plotted on the horizontal axis, and $x_2 - \bar{x}$ on the vertical axis, it will be found that any conceivable pair of sample values (x_1, x_2) will give rise to a point lying on a straight line with a gradient of -1 and passing through the origin.

Degrees of freedom are parameters of a number of important statistical distributions, and therefore form a natural quantity by which to tabulate them. For the straightforward case of a standard deviation in a single sample it makes very little difference whether the tabulation is in terms of sample size n or degrees of freedom v, for in that case $v = n - 1$. But tabulating the distribution of the appropriate statistic in terms of degrees of freedom can facilitate the use of the tables for other cases, for example:

a) for a single sample where the number, say k, of independent constraints is greater than 1. The table could be used in such a case with $v = n - k$;

b) for k samples of sizes $n_1, n_2, ..., n_k$ with different means but equal standard deviations which are to be combined for the purposes of estimating their common standard deviation. The table could be used in such a case with $v = (n_1 - 1) + (n_2 - 1) + ... (n_k - 1) = n - k$ where $n = n_1 + n_2 + ... + n_k$;

The appropriate statistic to use in such problems when σ is unknown is the t-statistic, a tabulation of which is provided in annex B.

Consider case b) with $k = 2$, the comparison of the means of two populations when neither the population means nor the population standard deviations are known. Suppose the hypotheses are as follows:

H_0: $\mu_1 = \mu_2$ against H_1: $\mu_1 \neq \mu_2$.

The sample data are a random sample of size n_1 from the first population and an independent random sample of size n_2 from the second population. The sample means are $\bar{x}_1$ and $\bar{x}_2$, and the sample variances (i.e. squares of the sample standard deviations) are s_1^2 and s_2^2, given by:

$$s_1^2 = \frac{\sum(x_1 - \bar{x}_1)^2}{n_1 - 1} \text{ and } s_2^2 = \frac{\sum(x_2 - \bar{x}_2)^2}{n_2 - 1}$$

Consider the statistic $d = \bar{x}_1 - \bar{x}_2$. Assuming that the sample sizes are small by comparison with their respective population sizes, we know from statistical theory that the population mean of d is as follows:

$$\mu_d = \mu_1 - \mu_2$$

and that the population standard deviation of d is as follows:

$$\sigma_d = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

where μ_1 and μ_2 are the population means and σ_1 and σ_2 are their standard deviations. The hypotheses can be restated as H_0: $\mu_d = 0$ and H_1: $\mu_d \neq 0$. The test is two-tailed, as the critical region of the test (where the truth of H_0 would be in doubt) will clearly consist of the large positive and negative values of d.

We shall consider only the case where the two population dispersions are the same, i.e. $\sigma_1 = \sigma_2 = \sigma$. The formula for σ_d then simplifies slightly to the following:

$$\sigma_d = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

An estimate of σ is s obtained by adding the numerators of the expressions given above for s_1^2 and s_2^2, dividing by the sum of the denominators, and then taking the square root, i.e.:

$$s = \sqrt{\frac{\sum(x_1 - \bar{x}_1)^2 + \sum(x_2 - \bar{x}_2)^2}{(n_1 - 1) + (n_2 - 1)}}$$

with degrees of freedom $v = (n_1 - 1) + (n_2 - 1)$. Substituting for σ in the previous formula, we obtain the following estimate:

$$s_d = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

of σ_d. The upper fractile $t_{1-a/2}$ corresponding to a two-tailed test with the required significance level a is read from the *t*-table in annex B. The null hypothesis is rejected in favour of the alternative hypothesis if the confidence interval for μ_d, with limits $\pm t_{1-a/2}\, s_d$, does not include the value zero.

8.4.2.4 *Power of the test*

As with all statistical tests, the power of the test should be estimated in advance of collection of the data. Too often in practice the power is wholly disregarded in planning a trial and the results turn out to be inconclusive. In this example, the power would have to be estimated rather than determined, as the value of σ is unknown. If the estimated power turned out to be lower than required for detecting a difference of a given magnitude between the population means, consideration should be given to increasing one or both sample sizes. If this is not possible, it may be decided not to waste resources in carrying out the trial or experiment, as the outcome is so unlikely to tell us anything we did not know (or thought we knew) already. There is always the possibility that the power will be found to be higher than required in which case the size of the proposed trial could be reduced.

In many cases, and this example is no exception, the calculation of the power involves relatively advanced statistics, which may explain why considerations of power are often avoided. However, the speed of modern desktop computers now enables the power of any proposed test to be estimated by means of simulation using simple statistical concepts.

8.4.2.5 *Comparison of two means in the case of paired observations*

Increasing the sample sizes is not the only way to increase the power. Another is to improve the precision of the comparisons by eliminating or reducing the effects due to differences between the samples of raw material on which measurements or treatments are carried out. For example, suppose we wish to compare the effect of two fertilisers, A and B, on a certain crop. One approach would be to apply fertiliser A to one random sample of n test plots and fertiliser B to a second random sample of n test plots, the $2n$ plots all coming from the same, fairly homogeneous field, and then to compare the two yields. But plots will inevitably differ with regard to drainage, levels of nutrients, etc. and by pure bad luck we could select two samples such that most of the plots in one sample were inferior to most of the plots in the other. This could compromise the conclusions from the trial. Even if both samples were similar, the plot to plot variation may be enough to reduce the power of the test to an unacceptable level, with commercially significant differences between fertilisers having too high a chance of not being detected.

A simple way of reducing the effect of plot to plot variation from such a test would be to select adjacent *pairs* of plots, applying fertiliser A to one of each pair chosen at random, and fertiliser B to the other. Suppose the yield from the i^{th} pair of plots is x_i for fertiliser A and y_i for fertiliser B. Then the difference $d_i = x_i - y_i$ would be affected hardly at all by plot to plot differences, assuming that adjacent plots are nearly identical. This is called the method of paired comparisons (see BS 2846-6). Provided x and y are independent and have approximately normal distributions, the differences d will be approximately normally distributed about the population mean difference μ_d with a population standard deviation σ_d. The sample mean $\bar{d}$ and sample standard deviation s_d provide estimates of these parameters.

The precise nature of the test will depend on the null and alternative hypotheses and the significance level of the test. If the two fertilisers are new and untried, we may simply wish to determine if one is superior. This could be done using a two-tailed test of H_0: $\mu_d = 0$ against H_1: $\mu_d \neq 0$; alternatively, and equivalently, a two-sided confidence interval for d could be calculated, to see whether or not it included the value zero. On the other hand, fertiliser B may be the standard against which a more expensive new fertiliser A is being tested; in this case, the new fertiliser may need to improve yield by more than an amount c to justify its extra cost. This could be done by means of a one-tailed test of H_0: $\mu_d = c$ against H_1: $\mu_d > c$. Or fertiliser B may be a null treatment, i.e. no treatment at all, in which case we would use a one-tailed test of H_0: $\mu_d = 0$ against H_1: $\mu_d > 0$ to determine if fertiliser A is effective. In all of these cases, Table 8 could be used.

The method of paired comparisons can be even more effective when the same item of raw material can be used for both treatments, e.g. in comparing the results of two test methods, or two measuring instruments, or two laboratories on the same product. The test or measurement process would clearly have to be non-destructive, e.g. measuring the strand widths of samples of tobacco for the purpose of classifying a consignment as pipe or cigarette tobacco, to determine the rate of duty payable.

Any information from previous trials of the likely size of σ_d should be utilized to determine the sample size n that will provide sufficient power.

8.4.2.6 *Comparisons of standard deviations*

We have already seen in **8.4.1** how to set confidence limits for a population standard deviation σ. Testing whether σ is equal to a given value σ_0 can be effected by calculating a confidence interval (σ_1, σ_2) and seeing if σ_0 lies inside the interval. A problem we have not yet addressed is testing for differences between two population standard deviations σ_x and σ_y. The problem is tackled by determining if σ_x/σ_y differs significantly from unity.

Suppose we have a sample of size n_1 from the first population and one of size n_2 from the second population. The respective sample standard deviations are s_x and s_y. Then a $100(1 - a)$ % two-sided confidence interval on σ_x/σ_y is:

$$\frac{s_x/s_y}{\sqrt{F_{n_1 - 1,\, n_2 - 1,\, 1 - a/2}}} \text{ to } \frac{s_x/s_y}{\sqrt{F_{n_1 - 1,\, n_2 - 1,\, a/2}}}$$

where $F_{n_1 - 1,\, n_2 - 1,\, 1 - a/2}$, and $F_{n_1 - 1,\, n_2 - 1,\, a/2}$ are the upper and lower $(a/2)$-fractiles of the F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

Tables of the F-distribution are three-way and therefore too extensive to reproduce here, so to illustrate the method we will simply quote the appropriate fractiles from published tables. Suppose a sample of size 10 from the first population yields $s_x = 10.5$ and a sample of size 16 from the second population yields $s_y = 6.8$. Is this evidence sufficient to conclude with 95 % confidence that there is a difference between σ_x and σ_y?

The significance level of the test is 5 %, or 0.05. From tables of the F-distribution it is found that $F_{9,\,15,\,0.975} = 3.12$ and $F_{9,\,15,\,0.025} = 0.265$. The ratio $s_x/s_y = 10.5/6.8 = 1.544$. A 95 % confidence interval on σ_x/σ_y is therefore $1.544/\sqrt{3.12}$ to $1.544/\sqrt{0.265}$, i.e. 0.874 to 3.00.

Since this range encloses the value 1, we cannot conclude with 95 % confidence that there is a difference between σ_x and σ_y.

One-sided tests can be treated in a similar way by calculating either of the one-sided confidence intervals:

$$\left(0, \frac{s_x/s_y}{\sqrt{F_{n_1 - 1,\, n_2 - 1,\, a}}}\right) \text{ or } \left(\frac{s_x/s_y}{\sqrt{F_{n_1 - 1,\, n_2 - 1,\, 1 - a}}}, \infty\right)$$

and determining whether or not the value 1 lies in the interval.

BS 2846-4 and BS 2846-5 deal with estimation and tests for means and variances with power functions for tests.

8.5 Simultaneous variation in the sample mean and in the sample standard deviation

Until now, we have considered the variation in sample means and standard deviations separately, but for problems of the type to be discussed later, the two need to be treated together. The nature of their relationship can be illustrated most clearly by plotting a point $(\bar{x}, s)$ to represent each sample on a diagram having $\bar{x}$ and s as co-ordinate axes. Such a diagram for the canned tomatoes data is shown in Figure 28, where each of the 40 dots represents a set of four test results. In the centre of the field, shown by a triangle, lies the point (μ, σ) representing the whole aggregate of items sampled; the values used are for all 160 original observations, namely $\mu = 403.8$ grams and $\sigma = 1.91$ grams.

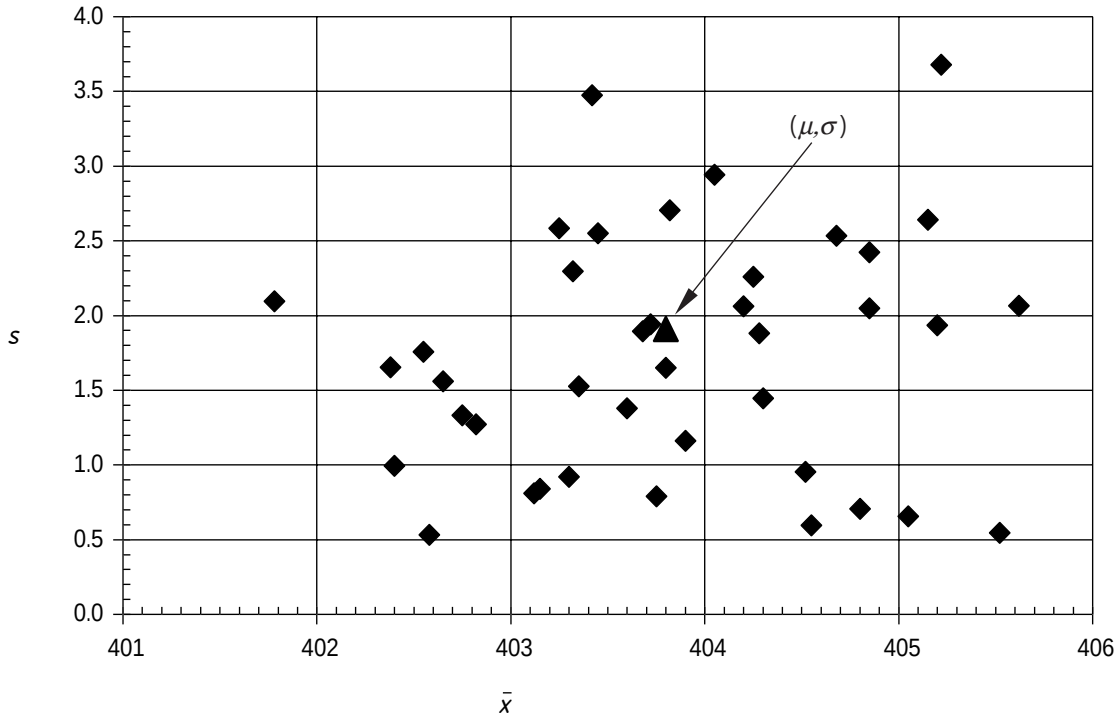


Figure 28 — Scatter chart for canned tomatoes data

Suppose now that an increasingly large number of shifts were sampled, the can contents weighed, the mean and standard deviation for sets of 4 results calculated and the points $(\bar{x}, s)$ plotted on the diagram. An increasing swarm of points would surround the central spot (μ, σ). Assuming production to be under statistical control, theory would allow the prediction, at least approximately, of what may be called the density of this swarm of points at different distances and in different directions from (μ, σ). In other words, we could express the chance that a sample point $(\bar{x}, s)$ would fall in any prescribed region of the diagram.

An illustration of where most of the sample points $(\bar{x}, s)$ would be expected to fall is given by the standardized $(\bar{x}, s)$ control charts of Kanagawa, Arizono and Ohta [3]. These charts, which are based on information theory, are useful for determining the type of departure from control by considering $\bar{x}$ and s simultaneously. A standardized $(\bar{x}, s)$ control chart for sample size 4, with limits for which there is only a 27 in 10 000 risk per observation of a false out-of-control signal, is shown in Figure 29 for the canned tomatoes data; the 40 sample points $[(\bar{x} - \mu_0)/\sigma_0, s/\sigma_0]$ have been plotted on the chart for target values of $\mu_0 = 404.0$ g and $\sigma_0 = 1.90$ g. (The reason for the strange choice of probability, 27 in 10 000, or 2 in 741, will become evident in **10.5** and **10.6**.)

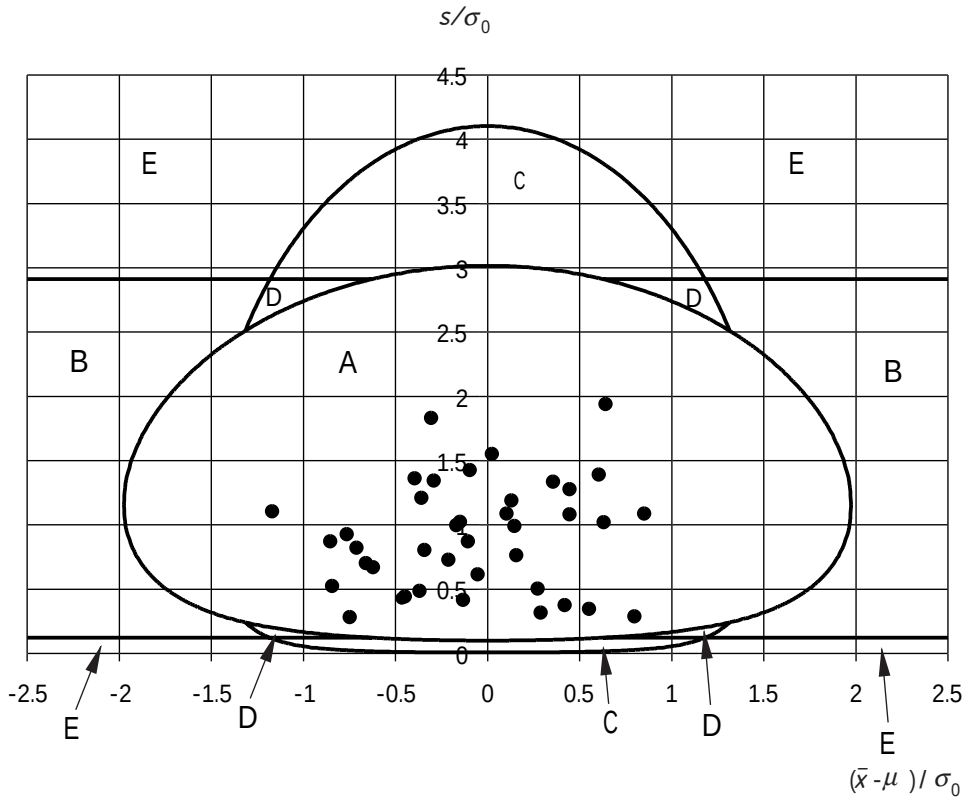


Figure 29 — Standardized control chart for mean and standard deviation

The regions on the chart are identified as follows.

A The process is in control.

B The process is out of control because of a change in the process mean.

C The process is out of control because of a change in the process standard deviation.

D The process is out of control because of a slight change to both the process mean and the process standard deviation.

E The process is out of control in both the process mean and the process standard deviation.

All of the 40 plotted points lie in region A, indicating that the tomato canning process is in control with respect to net weight.

The chart is standardized so that the same chart can be used regardless of the values of μ_0 and σ_0. The chart would only need to be changed if the sample size or false signal rate were changed. Increases in either of these quantities shrink the boundary lines and curves closer to the point with co-ordinates $(\bar{x} - \mu_0)/\sigma_0 = 0$, $s/\sigma_0 = \sqrt{n/(n-1)}$.

What is particularly interesting about this type of chart is the shape of the region A in which most of the standardized sample points $[(\bar{x} - \mu_0)/\sigma_0, s/\sigma_0]$ are expected to lie when a process is under control. Note that the closer s is to the target value σ_0, the more latitude is allowed in $\bar{x}$; similarly, the closer $\bar{x}$ is to the target value μ_0, the more latitude is allowed in s. In other words there is, in a sense, a natural trade-off between estimated departures of μ from μ_0 and estimated departures of σ from σ_0. Traditional control charts, by contrast, treat $\bar{x}$ and s separately (see clause **10**). Superficially the latter approach may seem logical from the point of view that $\bar{x}$ and s are known from statistical theory to be independent in samples from a normal distribution. But it allows no trade-off between the estimated departures from the target values μ_0 and σ_0, and is equivalent to having a rectangular in-control region on an $(\bar{x}, s)$ chart.

The *joint* consideration of $\bar{x}$ and s will be a recurring theme later in this clause and also in discussing methods of determining conformity to specification in clause **9**.

8.6 Tests and confidence intervals for proportions

8.6.1 *Attributes*

For many quality characteristics, it is either impossible or impracticable to obtain a measure of the characteristic on a continuous scale. For example, consider office cleaning services. A random sample of rooms could be inspected after cleaning to check that waste bins had been emptied, filing cabinets and desks had been dusted, and carpets had been vacuum-cleaned. For each of these three characteristics an experienced inspector would have little difficulty determining if the operation had been carried out to a satisfactory standard, and deciding that a room had been satisfactorily cleaned if it passed on all three checks. He would have rather more difficulty in grading the extent to which these tasks had been done on a meaningful continuous scale from, say, zero to one, and combining the grades in a coherent way to come to a decision on whether the cleanliness of the room was satisfactory.

Characteristics such as these, the realizations of which can most naturally be considered to fall into one of two states (pass/fail, go/no-go, ignites/fails to ignite), are called "attributes".

For critical characteristics, e.g. those that may affect the safety of personnel, every effort should have been made to ensure that the proportion of nonconforming items in the population is as near zero as possible. 100 % inspection would be used where practicable, and the critical items removed, in which case the proportion of critical items remaining in the population would be known to be zero. This assumes, of course, that the inspection is 100 % effective.

For non-critical characteristics, there may be a need to estimate the proportion of nonconforming items in the population, to calculate confidence limits on the proportion in the population, to test the proportion against a given value, or to compare two or more proportions.

8.6.2 *Estimating a proportion*

To continue the office cleaning example, suppose that a contractor is responsible for cleaning N rooms, i.e. the size of the population is N. On a particular day, suppose that R of the rooms would fail inspection, i.e. they have not been cleaned to a satisfactory standard. The proportion of the population that would fail inspection is therefore:

$$P = R/N.$$

R is unknown, so n rooms are chosen at random and inspected, with r failing inspection. The question is how best to estimate P.

A close analogy between the treatment of attributes and variables is possible here. Suppose the state of a room is characterized by a variable X taking the value 0 if a room is satisfactorily cleaned and 1 otherwise. The population of X values then consists of R ones and $(N - R)$ zeros, while the sample consists of r ones and $(n - r)$ zeros. Denote the sample values of X by x. Then the sum of the sample values of X is equal to r, i.e.:

$$\sum x = r.$$

The sample mean, $\bar{x}$, is therefore given by the following:

$$\bar{x} = \sum x/n = r/n = p, \text{ say} \qquad (19)$$

The sum of X in the population is equal to R, i.e.:

$$\sum X = R.$$

The population mean, $\bar{X}$ is therefore given by the following:

$$\bar{X} = \sum X/N = R/N = P \qquad (20)$$

It was stated in **8.1.1** that the sample mean is an unbiased estimator of the population mean. Here we use $\bar{x}$ from (19) as an unbiased estimator of $\bar{X}$ from (20), which translates into using the sample proportion p as an unbiased estimator of the population proportion P. The lack of bias holds good even when the sample size is a large proportion of the population size. For our example, if a sample of 50 rooms reveals that 2 were inadequately cleaned, it would be estimated that 2 out of 50, i.e. 4 % of the rooms in the population of rooms under consideration were unsatisfactory.

8.6.3 *Confidence intervals for a proportion*

Given that a random sample of size n contains r nonconforming items, it may be required to provide an interval, say P_1 to P_2, within which we may have a given confidence that the true proportion, P, of nonconforming items from the production process lies. Suppose the confidence is denoted by $100(1 - \alpha)$ %, with the chance of error of 100α % being equally divided between the case $P < P_1$ and the case $P > P_2$. These limits can be interpreted as follows.

i) If P was as low as P_1, there would be a probability of only $\alpha/2$ of finding r or more unsatisfactory items in a sample of size n.

ii) If P was as high as P_2, there would be a probability of only $\alpha/2$ of finding r or fewer unsatisfactory items in a sample of size n.

The probabilities in i) and ii) are calculated from a distribution called the *binomial* distribution. For small sample sizes, the values of P_1 and P_2 satisfying i) and ii) may be found in published tables. For larger sample sizes, approximate values may be read from published charts.

It is instructive to see how far the analogy between the treatment of attributes and variables can be extended to provide approximate confidence limits. To emulate the procedure in **8.4**, we require an estimate of the standard deviation of p, the estimated proportion. As the values of x are all zero or one and the square of zero is zero and the square of 1 is 1, we have the following situation:

$$\sum x^2 = \sum x = r$$

The sample standard deviation of x (see **5.2**) is therefore given by the following expression:

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

so an estimator of the standard deviation of $\bar{x}$ (or p) is as follows:

$$s/\sqrt{n} = \sqrt{\frac{\frac{r}{n}\left(1 - \frac{r}{n}\right)}{n - 1}} = \sqrt{\frac{p(1 - p)}{n - 1}}$$

This is the point from which the analogy becomes rather stretched. Confidence limits on P can be obtained by assuming the distribution of p is approximately normal. A two-sided confidence interval for P would then be of the form (P_1, P_2) where:

$$P_1 = p - u\sqrt{\frac{p(1 - p)}{n - 1}} \text{ and } P_2 = p + u\sqrt{\frac{p(1 - p)}{n - 1}}$$

where u is the upper $(\alpha/2)$-fractile of the standard normal distribution. If a one-sided confidence interval were required, then only P_1 or P_2 would be required; the chance of error would then only apply at one end of the interval, so the appropriate value of u would be the upper α-fractile of the standard normal distribution (see Table 14). Unfortunately, the normal approximation to the distribution of p is poor unless either P is close to one half or the sample size is quite large. In cases where this is not true, the use of this approximation should be confined to cases where only rough approximations to P_1 and P_2 will suffice.

Much effort has been devoted in the past to obtaining accurate approximations for P_1 and P_2, generally involving methods of improving the closeness of the normal approximation. See, for example, Molenaar [4] and Blyth [5].

8.6.4 *Comparison of a proportion with a given value*

Another common problem is how to determine whether a sample proportion differs from a given population value by more than can be attributable to chance. For example, would three substandard items in a sample of size 30 be sufficient to provide 95 % confidence that the percentage of substandard items in the population under consideration exceeded 3 %? There are two ways of answering this question. The first is to determine from the sample results the lower one-sided 95 % confidence limit on the percentage in the population, answering "yes" if 3 % was below this value and "no" otherwise. The second, a kind of inversion of the first, is to determine the probability of finding three or more substandard items in a sample of 30 when the percentage in the population is 3 %, answering "yes" if this probability is below 0.05 and "no" otherwise. The methods are essentially equivalent, but the latter has the advantage that it provides an actual measure of the confidence, rather than the result that it exceeded or did not exceed 95 %.

To answer the question in this specific case, the first method produces a lower confidence limit at 95 % confidence of 2.7 % substandard items in the population. (This can be found from published tables, e.g. ISO 11453.) As 3 % lies within the confidence interval, we would conclude that the sample result is compatible with a population percentage of 3 %. Alternatively, the probability of a random sample of 30 items containing three or more substandard items when the percentage in the population is 3 % can be shown to be 0.06. As this is greater than 0.05, the sample result does not provide sufficient evidence to conclude with 95 % confidence that the percentage of substandard items in the population exceeds 3 %. In fact we would only have confidence $100(1 - 0.06) = 94$ % that such a conclusion was correct.

It may seem somewhat surprising that a sample result of 3 in 30, i.e. 10 %, is insufficient to provide very high confidence that a population percentage exceeds 3 %. This goes to show how important it is to take into account the sample size when assessing a sample result.

8.6.5 *Comparison of two proportions*

Another related group of questions that can be answered by the use of appropriate statistical methods concerns whether the difference between two sample proportions is more than can be attributable to chance. Suppose that a random sample of size n_1 is taken from one population and a random sample of size n_2 from another population. (Usually n_1 and n_2 will be chosen to be equal.) Suppose further that the numbers of items with a given characteristic in the samples are determined to be r_1 and r_2. The two sample proportions are therefore $p_1 = r_1/n_1$ and $p_2 = r_2/n_2$. If P_1 and P_2 denote the unknown population proportions, the various questions that may be asked on the basis of the sample evidence are:

a) what confidence may we have that P_1 is different from P_2?;

b) what confidence may we have that P_1 exceeds P_2?; or

c) what confidence may we have that P_1 is less than P_2?

If p_1 is less than p_2 in case b), or p_1 is greater than p_2 in case c), then we could answer "not very much" without having to carry out any statistical calculations at all. The same would be true if p_1 and p_2 are approximately equal. In all other cases the answer may be determined by the use of tables of a distribution called the *hypergeometric* distribution. Special tables have been developed for directly determining the significance of any differences between p_1 and p_2 when n_1 and n_2 are small. For larger values of n_1 and n_2 a number of approximate methods have been devised.

8.6.6 *Sample size determination*

For tests on proportions, as with tests on means and variances, it is important to keep in mind the probability of detecting a difference of a size that would be considered important in practice, i.e. the power of the test. There may be technical or economic reasons why the sample or samples have to be limited in size, in which case it is useful to know to what extent this limits the power. If not, joint consideration of the required power and significance level of a test will enable an appropriate sample size (or sizes) to be determined, either from published tables or from approximate formulae. The formulae can look a bit daunting at first, but most are straightforward to use, albeit requiring a little care.

For example, suppose we wish to test the hypothesis that two population proportions, P_1 and P_2, are equal against the hypothesis that P_1 is greater than P_2, assuming that the sample size is to be the same from both populations. The common sample size is required that will provide a confidence $100(1 - \alpha)$ % of accepting the equality hypothesis when it is true, and provide a power $100(1 - \beta)$ % of concluding that there is a difference when P_1 and P_2 take certain different values (with P_1 greater than P_2). Then Walters [6] has shown that the approximate sample size can be found as the solution in n to the equation:

$$n = \frac{1}{2}\left(\frac{u_{1-\alpha} + u_{1-\beta}}{\sin^{-1}\sqrt{P_1 - 1/(2n)} - \sin^{-1}\sqrt{P_2 + 1/(2n)}}\right)^2 \qquad (21)$$

where $u_{1-\alpha}$ and $u_{1-\beta}$ are respectively the upper α and β fractiles of the standard normal distribution. Consider the case of a significance level of 5 % and a power of 90 % to detect a difference if $P_1 = 0.8$ and $P_2 = 0.6$. Setting $\alpha = 0.05$ and $\beta = 0.10$ and inserting the values of u from the left-hand side of Table 14, then (21) becomes:

$$n = \frac{1}{2}\left(\frac{1.644\,9 + 1.281\,6}{\sin^{-1}\sqrt{0.8 - 1/(2n)} - \sin^{-1}\sqrt{0.6 + 1/(2n)}}\right)^2 = \frac{4.282\,2}{(\sin^{-1}\sqrt{0.8 - (1/2n)} - \sin^{-1}\sqrt{0.6 + 1/(2n)})^2} \qquad (22)$$

Equation (22) can be solved iteratively by first guessing a value of n, then evaluating the right-hand side to give a new value of n, and repeating until the values of n converge. Suppose we start with an initial guess of $n = 50$. It can be verified that successive iterations of (22) give $n = 109, 96, 98, 98$. Further iterations are pointless, as they will evidently all produce $n = 98$. Thus, 98 is the appropriate size of random sample from each population.

Not only does this approximate method provide a solution in a matter of only a few iterations, but it is also very accurate. Formula (21) can also be used for two-sided alternative hypotheses by replacing α by $\alpha/2$.

ISO 11453 provides methods of estimation, testing, setting of confidence limits and sample size determination for problems relating to proportions.

8.7 Prediction intervals

8.7.1 *One-sided prediction interval for the next m observations*

We may sometimes wish to determine the value of an upper limit, T_U, based on the results of a random sample of size n from a normal population, in such a way that we would have a given level of confidence that none of the next m random observations from the same normal population will exceed T_U. In general, this upper limit is given by the formula:

$$T_U = \bar{x} + qs$$

where $\bar{x}$ and s are the sample mean and the sample standard deviation and q is a factor that depends on the sample size n, on the number m of future observations and on the level of confidence required.
Table 17 shows the values of this factor for a range of values of n and m for a confidence level of 95 %.

Table 17 — Factors, q, for calculating one-sided prediction intervals — Confidence level 95 %

Sample size	Number of future observations, m							
n	5	10	20	50	100	200	500	1 000
5	3.787 9	4.417 8	5.028 8	5.793 8	6.337 5	6.852 5	7.493 4	7.951 5
10	2.886 8	3.284 1	3.669 9	4.159 3	4.512 5	4.851 0	5.276 9	5.583 8
20	2.574 4	2.890 7	3.194 0	3.577 7	3.855 7	4.123 7	4.463 2	4.709 7
50	2.415 3	2.689 8	2.948 8	3.272 0	3.504 4	3.727 8	4.011 2	4.217 5
100	2.366 1	2.627 7	2.872 7	3.175 9	3.392 5	3.599 8	3.861 6	4.051 7
200	2.342 2	2.597 6	2.835 7	3.129 0	3.337 6	3.536 5	3.786 9	3.968 0

Note the way in which the factor inflates as n decreases (due to having less information on which to base the prediction) or as m increases (due to being more ambitious in what the interval is to include).

From the symmetry of the normal distribution, it will be evident that a value of q that provides a given confidence that none of the next m observations *exceed* the upper limit T_U provides the same confidence that none of the next m observations are *less than* a lower limit, T_L, given by:

$$T_L = \bar{x} - qs$$

To illustrate the use of one-sided prediction intervals, suppose that a retailer has complained to its supplier that several size 12 ladies' jumpers of a particular style have had bust sizes above the nominal maximum of 92½ cm. The supplier has 1 100 jumpers remaining out of a batch of this size and style, all of which were made under the same conditions, and decides to check the bust sizes of a random sample of 100 of them. None of the 100 measurements was found to exceed 92½ cm. Past supplier data suggests that the bust sizes tend to be approximately normally distributed, and a normal plot of the 100 measurements gives no grounds to doubt the assumption of normality. The sample mean and standard deviation turn out to be $\bar{x} = 90.1$ cm and $s = 0.4$ cm respectively. The factor for a one-sided prediction interval with $n = 100$ and $m = 1\,000$ is seen from Table 17 to be 4.051 7. The supplier can therefore be roughly 95 % confident that none of the remaining 1 000 garments have bust measurements in excess of $90.1 + 4.051\,7 \times 0.4 = 91.7$ cm.

As 91.7 cm is well below the nominal maximum of 92½ cm, the supplier continues to supply the retail trade with jumpers from this batch.

8.7.2 *Two-sided prediction interval for the next m observations*

Alternatively, it may be required to determine both a lower limit T_L and an upper limit T_U from our initial sample of size n, such that we have a given confidence that none of the next m observations will lie outside the interval (T_L, T_U). These limits are given by the following:

$T_L = \bar{x} - rs$

and

$T_U = \bar{x} + rs$

where r, like q, is a factor depending on n, m and the required level of confidence. Table 18 shows the values of r for two-sided prediction intervals with confidence 95 % for some values of n and m.

Table 18 — Factors, *r*, for calculating two-sided prediction intervals — Confidence level 95 %

Sample size	Number of future observations, *m*							
n	5	10	20	50	100	200	500	1 000
5	4.577 3	5.228 6	5.851 7	6.624 0	7.169 8	7.685 4	8.326 1	8.783 7
10	3.321 0	3.717 3	4.102 5	4.590 5	4.942 4	5.279 3	5.702 9	6.008 2
20	2.902 1	3.207 6	3.502 9	3.878 4	4.151 4	4.415 1	4.749 8	4.992 9
50	2.693 4	2.952 3	3.198 9	3.509 0	3.733 5	3.950 2	4.226 0	4.427 3
100	2.629 8	2.874 3	3.105 5	3.394 1	3.601 6	3.801 2	4.054 3	4.238 6
200	2.599 0	2.836 6	3.060 3	3.338 3	3.537 2	3.727 9	3.968 9	4.144 0

As an example of two-sided prediction intervals, suppose that it is required to verify on a sample basis that a batch of 250 pairs of size L men's trousers have waistbands all in the range 86 cm to 92 cm. A random sample of 50 pairs yields $\bar{x}$ = 88.8 cm and s = 0.78 cm, with none of the individual measurements outside the specified range. The appropriate factor from Table 18 with n = 50 and m = 200 is 3.950 2. Assuming a normal distribution of waistband measurements, a two-sided prediction interval is found to be $88.8 \pm 3.950\,2 \times 0.78$, i.e. 85.7 cm to 91.9 cm. Since the lower limit of this prediction interval violates the lower specification limit, the supplier decides it is in his best interests to check the other 200 pairs individually before shipping them.

For both one-sided and two-sided prediction intervals, it is also possible to provide factors which assure with a given confidence that no more than 1, or no more than 2, etc. of the next m observations will fall outside the limits. However, the case of zero out of m is generally of the most interest. It is also possible to provide factors for the case where the process standard deviation σ is known, or at least presumed to be so, which will tend to lead to smaller prediction intervals.

See Hahn [7, 8], Hahn and Nelson [9] and Hahn and Meeker [10] for further details.

8.7.3 *One and two-sided prediction intervals for the mean of the next m observations*

It may on the other hand be required to provide a prediction interval for the *mean* of the next m observations. Prediction intervals for the mean can be determined more readily from standard tables, being based upon the t-distribution. A one-sided upper limit for the mean at confidence $100(1 - \alpha)$ % is given by:

$$T_U = \bar{x} + t_{n-1,\,1-\alpha} s \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)}$$

where $t_{n-1,\,1-\alpha}$ is the upper α-fractile of the t-distribution with $n - 1$ degrees of freedom. A corresponding one-sided lower limit for the mean of the next m observations is given by:

$$T_L = \bar{x} - t_{n-1,\,1-\alpha} s \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)}$$

Note that $t_{n-1,\,1-\alpha}$ only depends on two quantities, namely the sample size and the required confidence level, as a result of which tables of t are rather more compact than the three-way tables needed for q and r.

Two-sided prediction intervals on the mean are given by (T_L, T_U) where:

$$T_L = \bar{x} - t_{n-1,\,1-\alpha/2} s \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)}$$

and

$$T_U = \bar{x} + t_{n-1,\,1-\alpha/2} s \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)}$$

It should always be borne in mind that departures from normality could cause considerable errors in the prediction intervals, particularly when these intervals extend far outside the range of the sample values.

8.8 Statistical tolerance intervals

8.8.1 *Statistical tolerance intervals for normal populations*

We have seen in **8.7** that a prediction interval is an interval, derived from a sample, within which a specified finite *number* of future observations may be asserted to lie with a given confidence. A statistical tolerance interval is also derived from a sample, but is an interval within which a specified *proportion* of the population values may be asserted to lie with a given confidence.

The name of these intervals is unfortunate, as it can be misconstrued to mean the interval between the tolerance limits specified by the user. In fact, the limits of a statistical tolerance interval, like those of a prediction interval, will vary from sample to sample. As a statistical tolerance interval is asserted to include, or *cover*, a proportion of the population, an alternative name that is sometimes used is statistical *coverage* interval.

For populations that are normally distributed, the intervals are constructed in much the same way as prediction intervals, but with different values for the factors by which the standard deviation is multiplied. Published tables address four cases:

a) one-sided limits when the process standard deviation is known, of the form $\overline{x} - b_1\sigma$ or $\overline{x} + b_1\sigma$;

b) two-sided intervals when the process standard deviation is known, of the form $(\overline{x} - b_2\sigma, \overline{x} + b_2\sigma)$;

c) one-sided limits when the process standard deviation is unknown, of the form $\overline{x} - b_3 s$ or $\overline{x} + b_3 s$;

d) two-sided intervals when the process standard deviation is unknown, of the form $(\overline{x} - b_4 s, \overline{x} + b_4 s)$;

where the constants b_1, b_2, b_3 and b_4 depend on the sample size, the coverage and the required level of confidence.

A simple example will illustrate this type of interval. A customer who has received a batch of 12 000 bobbins of cotton yarn decides to check on its breaking load distribution. He takes a random sample of 24 bobbins, and cuts from each a test piece of length 50 cm at about 5 m distance from the free end. The central part of each test piece is tested for breaking load. The unit of measurement is the centinewton. The sample mean and standard deviation turn out to be $\overline{x} = 249.8$ and $s = 31.4$. From previous experience, it is known that the distribution of breaking loads closely approximates a normal distribution. From tables, it is found that the constant for a one-sided statistical tolerance interval for sample size 24 for coverage 95 % and confidence level 95 % is $b_3 = 2.310$. The lower statistical tolerance limit is therefore $249.8 - 2.310 \times 31.4 = 177.3$. The customer can therefore be 95 % confident that at least 95 % of the breaking loads are in excess of 177.3 centinewtons.

Suppose that the customer felt confident enough from previous batches from the same supplier to assume that it was only the mean that varied from batch to batch, and that the cotton yarn coming from the production process had a breaking load with a constant standard deviation. Then the appropriate constant would be $b_1 = 1.981$. Note that this is considerably smaller than the corresponding value 2.310 of b_3 for the case of unknown process variability. This is because the extra information leads to a smaller safety margin being required. Suppose that σ is known to be 33.2. This produces a statistical tolerance limit of $249.8 - 1.981 \times 33.2 = 184.0$. The customer could now be 95 % confident that at least 95 % of the breaking loads are in excess of 184.0 centinewtons.

If there is any doubt about the constancy of σ, it should be assumed to be unknown and case c) or d) used as appropriate.

8.8.2 *Statistical tolerance intervals for populations of an unknown distributional type*

Even if the form of the distribution of values of the characteristic in the population is in doubt, it is still possible to construct one- and two-sided statistical tolerance intervals. Instead of being based on statistics such as $\overline{x}$ and s, they are based on what are known as the *order statistics*, that is to say individual sample values after they have been sorted and numbered in ascending order. Any single or pair of order statistics can be used to provide a statistical tolerance interval but, of course, the largest and/or the smallest provide the greatest coverage. The penalty in not knowing the distributional form is that the statistical tolerance intervals will be rather wider than they would otherwise have been, or require larger sample sizes. To give some idea of the numbers involved, a sample of size 93 is required in order to have 95 % confidence that the interval formed by the largest and smallest observations covers 95 % of the population values. This rises to a sample of size 473 when the coverage increases to 99 %, and to a sample of size 4 742 for coverage of 99.9 %.

Tables of factors for statistical tolerance limits for the normal distribution may be found in Odeh and Owen [11], Hahn and Meeker [10] and in ISO 16269-6. The latter also provides tables of minimum sample sizes required for a selection of coverages and confidence levels, for both the one- and two-sided cases, when the population distribution is of unknown form.

8.9 Estimation and confidence intervals for the Weibull distribution

8.9.1 *The Weibull distribution*

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 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. The probability density function for this form of the Weibull distribution is as follows:

$$f(x) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta - 1} e^{-(t/\alpha)^{\beta}} \text{ for } t \geq 0$$

Figure 30 shows the way this density function changes shape for the case $\alpha = 1$ as β increases from ½ to 4.

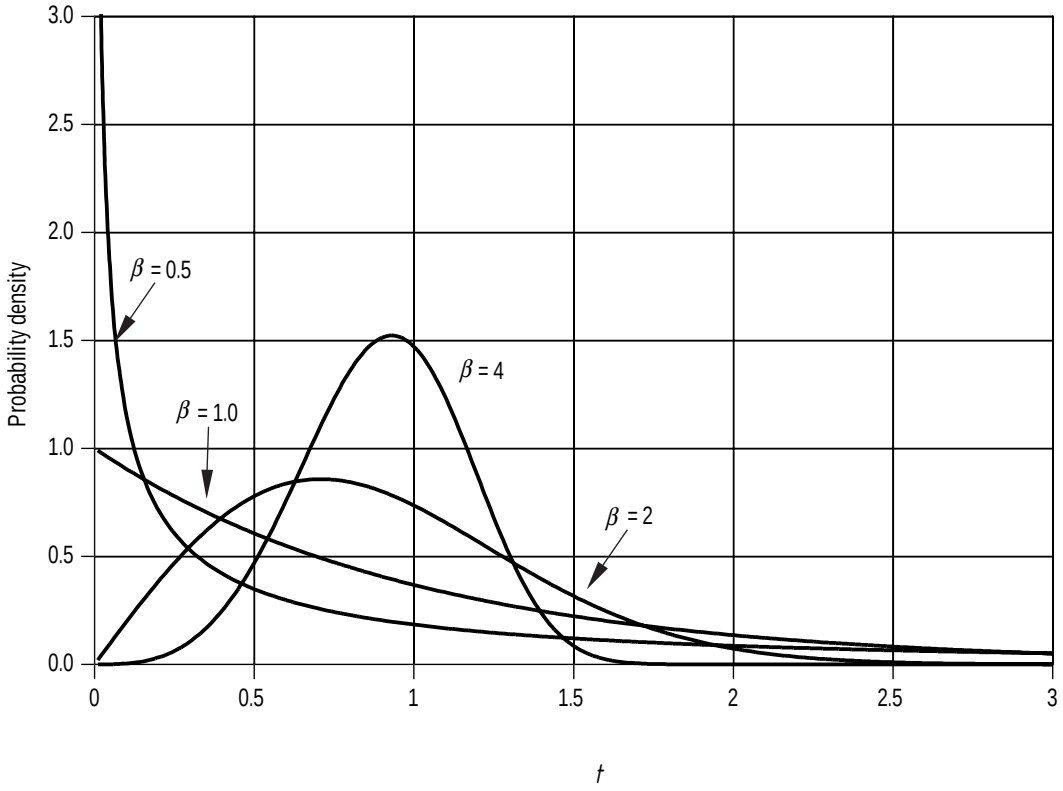


Figure 30 — Comparison of Weibull distributions all with $\alpha = 1$

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 the following equation:

$$F(t) = 1 - e^{-(t/\alpha)^{\beta}} \text{ for } t \geq 0$$

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

$$R(t) = 1 - F(t) = e^{-(t/\alpha)^{\beta}} \text{ for } t \geq 0$$

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 hours, at which 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.

8.9.2 *Goodness-of-fit tests*

Although it is somewhat subjective, the easiest way to check if the Weibull distribution will provide a reasonable fit to a given set of data is to plot the data on Weibull probability paper. This type of graph paper has been specially devised so that data from a Weibull distribution tend to lie on a straight line, having a log log scale on the $F(t)$ axis and a log scale on the t axis. It is based on the fact that:

$$\ln \ln \left(\frac{1}{1 - F(t)} \right) = \ln \left((t/\alpha)^{\beta} \right) = \beta \ln (t) - \beta \ln (\alpha)$$

which is a straight line relationship between $\ln \ln \left(\frac{1}{1 - F(t)} \right)$ and $\ln (t)$.

As with the normal distribution, the plotting procedure is as follows. The n sample values are first arranged in ascending order to give the *order statistics* $t_{[1]}, t_{[2]}, ..., t_{[n]}$, i.e. such that $t_{[1]} \leq t_{[2]} \leq ... \leq t_{[n]}$. For each $t_{[i]}$, the point with co-ordinates $[t_{[i]}, i/(n + 1)]$ is plotted.

NOTE 1 When the sample values come from a Weibull distribution, the line joining successive points on Weibull probability paper tends to be more straight if $(i - 0.3)/(n + 0.4)$ is plotted on the vertical axis instead of $i/(n + 1)$.

NOTE 2 A similar improvement for the use of normal probability paper can be achieved by replacing $i/(n + 1)$ by $(i - \frac{3}{8})/(n + \frac{1}{4})$.

Numerical tests are also available for testing the goodness of fit of the Weibull distribution. These generally require a fair amount of computation involving the order statistics.

8.9.3 *Parameter estimation*

Many numerical procedures have been proposed for the estimation of the Weibull parameters. Some involve iterative solution, others involve the use of special tables. A rough graphical approach is as follows.

a) Plot the data on Weibull probability paper, and fit a straight line to the points by eye.

b) Read off from this line the value of t at which the cumulative probability is 1.0 %. Denote this by $t_{0.010}$.

c) Read off the value of t at which the cumulative probability is 63.2 %. Denote this by $t_{0.632}$.

d) Estimate α as $t_{0.632}$.

e) Estimate β as $4.6/\ln(t_{0.632}/t_{0.010})$.

8.9.4 *Confidence intervals*

Detailed discussion of point estimation and confidence interval determination for quantities related to the Weibull distribution is beyond the scope of this standard. We content ourselves with merely listing the most important ones. A variety of methods for the calculation of point estimates and confidence intervals exist for:

a) the parameters α and β;

b) the mean time to failure;

c) fractiles of the time to failure;

d) the reliability at time t.

Two international standards, IEC 1649 and prEN 12603, provide procedures for the calculation of point estimates and confidence intervals for the Weibull distribution. IEC 1649 also contains a valuable annex explaining the reasons for the particular choice of methods. Two well-known books on the subject are Mann *et al* [12] and Lawless [13].

8.10 Distribution-free methods: estimation and confidence intervals for a median

So far, we have considered inference from a sample about characteristics of a population when the population distribution is known to belong to a particular family of distributions, e.g. the normal or Weibull families. However, if the form of the population distribution is unknown, statistical methods can still be brought to bear in drawing inferences about a population distribution. Such methods, because they do not depend on the form of population distribution, are called *distribution-free* methods. The advantage of distribution-free methods is that they have greater integrity when there is any doubt at all about the form of the population distribution. The disadvantage is that confidence intervals for probabilities and fractiles are wider than would be the case using methods specially tailored to the specific family of distributions.

An example of the use of a distribution-free method is the determination of confidence limits for the population median (i.e. the value of the characteristic under consideration that divides the total frequency into two halves) when the distributional form is unknown. The median of a population may be of more interest than the mean when the distribution is highly skewed, which can cause the mean to be unduly affected by a small number of extreme values as is the case, for example, for income distributions. To obtain a distribution-free confidence interval, the values of the characteristic in a random sample of size n are first ranked in ascending order of magnitude to give the order statistics $x_{[1]}, x_{[2]}, ..., x_{[n]}$. A symmetrically positioned pair of order statistics $(x_{[k]}, x_{[n + 1 - k]})$ is then used as the pair of confidence limits.

The smaller the value of k, the larger the confidence that the population median will be included in the interval. For example, consider the case with $k = 1$, providing confidence limits $x_{[1]}$ and $x_{[n]}$. These limits will only fail to include the population median if all n sample values lie above the median or all lie below. As the chance of each original observation lying below the population median is one half and the chance of lying above it is one half, the chance of the population median not being included in the interval is $(½)^n + (½)^n = (½)^{n-1}$.

Our confidence on any one occasion that the population median is included between $x_{[1]}$ and $x_{[n]}$ is therefore $1 - (½)^{n-1}$, which as n takes the values 2, 3, 4, 5, ..., etc. gives confidence levels (in percentage terms) of 50 %, 75 %, 87.5 %, 93.75 %, ..., etc. One-sided distribution-free confidence intervals can be constructed, of the form $(a, x_{[n]})$ or $(x_{[1]}, b)$ where a and b are the smallest and largest possible values of the characteristic in the population. These confidence levels represent the *largest* confidence levels for the given sample sizes; in other words, the confidence that distribution-free confidence intervals can provide is limited by the sample size.

Note that these confidence levels are entirely independent of the distributional form of the population. The only assumption made in the above argument is that the probability of a sample value lying on either side of the population median is one half, which requires the distribution to be continuous at that point. Note also that the effect of increasing k is to decrease the width of the confidence interval at the cost of decreasing the confidence level.

Published tables provide, for moderate sample sizes and popular confidence levels, the largest value of k that will provide at least the required confidence. For large sample sizes a number of approximations have been developed. ISO 16269-7 gives tables for k for sample sizes up to 100 and, for use with larger samples, approximations for confidence level $1 - \alpha$ of the following form:

$$k = \frac{1}{2}\left(n + 1 - u\left(1 + \frac{0.4}{n}\right)\sqrt{n - c}\right)$$

where

- k is to be rounded down to the next whole number;
- u is the standard normal deviate corresponding to an upper tail area of α for a one-sided confidence interval and $\alpha/2$ for a two-sided interval;
- c is a constant depending on u.

The values of u and c are provided for eight different confidence levels and for one- and two-sided intervals. It is claimed in the standard that in all these cases the formula yields the correct value of k for sample sizes up to at least 280 000, enough for most practical purposes! For illustration, for a two-sided confidence interval with confidence level 99 % the values of u and c are given as 2.575 829 30 and 1.74. Thus, for a sample size of 200, the appropriate value of k is:

$$\frac{1}{2}\left(200 + 1 - 2.575\,829\,30 \times \left(1 + \frac{0.4}{200}\right)\sqrt{200 - 1.74}\right) = ½(201 - 36.341\,5) = 82.33$$

rounded down, i.e. $k = 82$. It can therefore be asserted in general with at least 99 % confidence that the confidence interval $(x_{[82]}, x_{[119]})$ from a sample of size 200 includes the population median.

The eight decimal places for u are only necessary when obtaining k with high accuracy for very large sample sizes, and may be reduced to two or three decimal places if an approximate value for k is adequate.

Similar methods can be used to determine confidence intervals on other fractiles of the population.

9 Acceptance sampling

9.1 Methodology

In clause **8** a variety of statistical tests and intervals have been described. In the example in **8.3** it was shown how easy it can be to select an inferior criterion for assessing the quality of a lot, even in the simplest case where there is a single-sided requirement on a single quality characteristic. The difficulties in selecting a sound criterion are compounded if there are multiple quality characteristics, perhaps some with single and the others with double specification limits, particularly if not all of these characteristics are independent. In the face of the multiplicity of potential applications and the many techniques from which to choose, a general approach to the problem of assessing quality is desirable.

The supplier will naturally concentrate his attention on keeping the mean of each quality characteristic as close as possible to a target value, and the standard deviation as small as possible (see clause **10**). The customer, on the other hand, will be principally concerned with the quality level of submitted product, i.e. the percentage of nonconforming items or the number of nonconformities per 100 items. Nonconformity is defined as departure of a characteristic from specification, and the probability of such a departure will generally depend on the mean, μ, and standard deviation, σ, of the characteristic in the population. For example, it can be seen from Table 14 that if a lower specification limit, L, has been set then, provided $\mu - 2.053\,7\sigma \geq L$, no more than about 2 % of product will be outside specification if the distribution of the characteristic is normal. If L was equal to 100, say, then a combination such as $\sigma = 1$, $\mu = 102.6$ would provide a similar quality level to the customer as the combination $\sigma = 1.2$, $\mu = 103.1$.

This suggests the following general approach to the assessment of product acceptability: use the sample information to estimate the proportion of product that is outside specification, and accept the batch only if the estimate is below a given maximum value. A judicious choice of this maximum value will provide a given level of assurance that not more than a given proportion of product is outside specification. It turns out that this approach does indeed lead to efficient use of the sample information and to intuitively sensible sampling procedures.

Sometimes the testing of the estimated quality level against a maximum value is done implicitly. For example, for sampling by attributes, the unbiased estimate of the fraction nonconforming is:

$$\hat{p} = \frac{r}{n}$$

where r is the number of nonconforming items in a sample of size n from the lot. Suppose that the maximum value of $\hat{p}$ for which lot acceptance takes place is denoted by p^*. Then a lot is only accepted if $\hat{p} \leq p^*$, i.e. $r/n \leq p^*$, i.e. $r \leq np^*$, i.e. $r \leq c$ where c is the largest whole number less than or equal to np^*. In practice, the acceptance criterion in this situation is always expressed as $r \leq c$ (or $r \leq$ Ac, see **9.4.1**), so it is not immediately obvious that it conforms to the suggested general approach.

Because these sampling methods are used to determine whether or not a population (lot, batch, consignment) of product should be accepted, they are referred to as *acceptance sampling* methods. The methods described in clause **9** are primarily for application to a continuing series of lots from the same supplier, although the case of isolated lots or short series is also considered.

9.2 Rationale

The emphasis in industry moved during the latter half of the 20th century from inspection of final product towards improving the manufacturing processes and controls and producing more robust product designs. When the supplier's quality control system can provide assurance that a process is in statistical control with sufficiently low variability in the quality characteristic or characteristics, sampling inspection of the final output from the process would simply be a waste of resources, merely confirming what was already known. Many mature industrial processes are in this happy state, which should be the ultimate aim for any process. Relying on acceptance sampling to assure the customer that he is getting what he wants is nowadays rightly seen as an inferior state, wasteful in terms of re-inspection, rework, scrap and extra administrative costs, not to mention the loss of customer confidence and competitive edge. Acceptance sampling has almost become a taboo topic in some quarters, being akin to an admission of failure to get the production processes into shape.

So what reasons are there for continuing with this apparently outmoded practice? There are several. The first is that the process may be immature, with unexpected teething troubles arising from time to time that would not necessarily be picked up in production under the existing process controls. Another is that some of the processes involved may be state-of-the-art, perhaps using materials whose properties are not yet fully understood. (This is sometimes the case in defence industries, due to the constant push at the limits of technology and changes to the specification in order to produce devices that are superior to those of the perceived adversary.) Another reason is that it may be necessary to guard against human fallibility and unpredictability, for example where the production item is a complex and delicate assembly of components and where a warranty system is inappropriate.

It is important to emphasize that acceptance sampling should not be seen as a means of sorting good lots from bad. (Indeed, Mood [14] showed that if the process quality level remains constant, then there is no correlation between the fraction nonconforming that finds its way into a sample and the fraction nonconforming in the remainder of a lot. It follows that, unless the sample size is a large proportion of the lot size, the sample results from a stable process are a poor determinant of the quality of the whole lot.) Rather, acceptance sampling should operate under the presumption that the lot is expected to be acceptable and be seen as a precautionary measure, to detect a deterioration in quality level that could not have been detected by any of the existing process controls.

9.3 Some terminology of acceptance sampling

9.3.1 *The AQL*

Some of the aversion to acceptance sampling was due to most international standards on acceptance sampling being indexed by the acceptable quality level (AQL). There were two main objections to the use of AQLs. One was the name itself; the idea that any quality level other than perfection should be considered acceptable or satisfactory in the modern era became outmoded, as one of the basic tenets for surviving in a global economy is the need to strive for continuous quality improvement. The other was the definition of an AQL as "when a continuing series of lots is considered, a quality level which for the purposes of acceptance sampling is the limit of a satisfactory process average". This definition had been deliberately worded to indicate that the acceptability was to be construed as only for the purpose of identifying a suitable sampling" plan, not in any absolute sense. But for the most part the words "for the purposes of acceptance sampling" have been either ignored or misunderstood by commentators.

In 1998, in order to emphasize that the AQL should not be interpreted as a desirable quality level, the meaning of the acronym was revised by ISO/TC 69/SC 5, the ISO subcommittee responsible for developing and maintaining international standards on acceptance sampling. It now stands for "acceptance quality limit" defined as the "quality level that is the worst tolerable process average when a continuing series of lots is submitted for acceptance sampling".

9.3.2 *Limiting quality (LQ)*

When acceptance sampling is applied to a single lot, or to a short series of lots, the concept of an AQL is inappropriate as there is no longer a continuing series. The principal index to classical sampling plans for isolated lots is the limiting quality (LQ) which is the quality level for which the probability of acceptance is a specified low value, usually 10 %. [This is the same thing as the lot tolerance percent defective (LTPD), a term which is now rarely used.] The LQ is chosen by the customer to be an unsatisfactory lot quality level at which lots would be expected to have a high probability of failing the acceptance criterion. For a continuing series of lots, the corresponding unsatisfactory *process* quality level is called the limiting quality level (LQL). More generally, for both lots and processes this quality level is referred to as the consumer's risk quality (CRQ).

9.3.3 *Classical versus economic methods*

Classical acceptance sampling methods are typically selected with little attempt to balance the costs of sampling and inspection against the savings due to more reliably accepting good product and rejecting bad. The reasons for this are fourfold. Firstly, experience has indicated that the classical methods provide sampling procedures that are not that far from the economic optimum over a wide range of scenarios. Secondly, in order to be able to determine the optimal level of inspection for an economic plan, the cost of non-acceptance of good lots and the cost of acceptance of bad lots need to be known. The latter cost is typically particularly difficult to ascertain in most cases, partly because it depends on the extent to which the nonconforming items are out of specification. Thirdly, a presumption has to be made about the distribution of incoming quality. Fourthly, even when an assessment can be made of all these quantities, the economic sampling procedure generally depends on finding the minimum of a complicated formula, requiring too high a level of mathematical sophistication for the typical user.

For the above reasons, we shall only consider classical types of acceptance sampling scheme below. Readers interested in further details of the economic approach are referred to Wetherill and Chiu [15] and von Collani [16]. Some flexibility to lower or raise the amount of inspection on economic or other grounds is provided by the choice of inspection level, which is described next.

9.3.4 *Inspection levels*

The two best-known acceptance sampling systems, the British versions of which are BS 6001-1 for sampling by attributes and BS 6002 for sampling by variables, provide three general inspection levels, I, II, and III, and four special inspection levels, S-1 to S-4. If the inspection level is not specified, it is assumed that general inspection level II is to be used. If better discrimination between good and bad quality is required, perhaps because the supplier has a history of erratic quality, inspection level III may be chosen. Conversely, if a lower level of discrimination is adequate, inspection level I may suffice. Sometimes even the lower sample sizes required by inspection level I are uneconomic when the inspection is expensive or involves destructive testing, or unnecessarily costly in view of the excellent quality history of similar products, the reputation of the supplier or the low importance of the characteristics under consideration. In such cases, one of the levels S-1 to S-4 may be selected, as long as it is understood that the discriminatory ability (i.e. the power) of the sampling scheme tends to diminish as one moves from S-4 to S-1.

The inspection level in combination with the lot size determines a sample size code letter, which is then used in conjunction with the AQL to look up the parameters of the sampling plans of an acceptance sampling scheme.

9.3.5 *Inspection severity and switching rules*

At the start of sampling inspection, when it is believed that the quality level of a process is satisfactory, so-called *normal inspection* is used in BS 6001-1 and in BS 6002. If the results from a predetermined number of lots under normal inspection indicate that the quality level of the process is less than satisfactory, then the severity of the inspection is increased to *tightened inspection*. A tightened inspection plan will usually have the same sample size as the corresponding normal inspection plan, but with a stricter acceptability criterion. If the results from a predetermined number of lots under normal inspection indicate that the quality level of the process is very good, then the severity of the inspection may be decreased to *reduced inspection*. Thus, each normal inspection plan has a corresponding tightened and reduced inspection plan. Each group of three such plans is called a *sampling scheme*. The standards BS 6001-1 and BS 6002 are collections of sampling schemes of a particular type, and are called *sampling systems*. The rules for moving between the plans that make up a scheme are called the *switching rules*. Whereas the normal inspection and tightened inspection plans are mandatory parts of a BS 6001-1 or BS 6002 scheme, the reduced inspection plan is discretionary. Thus, a sampling scheme from these standards consists of at least two plans. (It is sometimes overlooked that the plans were not designed to be used in isolation.)

A switch is also made from tightened inspection to discontinuation of inspection if the sample quality levels fail to improve sufficiently quickly. The supplier then needs to act to resolve any problems with the process before acceptance sampling may be resumed. Tightened inspection is then applied with the switching rules reset in the same way as if there had just been a switch from normal inspection.

9.3.6 *Use of "non-accepted" versus "rejected"*

New users of acceptance sampling standards may be bemused to find the word "non-accepted" used where the word "rejected" may seem more appropriate. There is a good reason for distinguishing these terms. The term "rejected" implies that the user is not prepared to accept the lot under any circumstances. However, when a lot is non-accepted, this only means that it has failed the acceptance criterion of the sampling inspection plan. It does not preclude the customer and the supplier from coming to some arrangement to accept the lot on concession, e.g. at a reduced price, or for a different use such as training, or after some remedial action.

9.4 Acceptance sampling by attributes

9.4.1 *General*

The concept of an attribute has already been discussed in **8.6.1**. For the moment, suppose that items have a single quality characteristic that is an attribute; multi-attribute cases are discussed in **9.6.3** and **9.6.4**. For the purposes of discussion, suppose also that quality is measured in terms of percent nonconforming rather than nonconformities per 100 items. (The methods for nonconformities per 100 items for a single attribute are very similar.)

It is important to distinguish the case of an isolated lot from that of a continuing series of lots. For an isolated lot, the primary purpose of acceptance sampling will be to provide assurance that the *lot* fraction nonconforming is no worse than the limiting quality (LQ). Tabulated plans for isolated lots or short series of lots are therefore indexed by LQ and lot size. For a continuing series of lots, AQL-indexed plans, which are also indexed by lot size, are designed to provide protection against lots being accepted when the *process* quality level is worse than the AQL.

The performance of a sampling plan can be assessed in both cases by considering its *operating characteristic* (OC) curve. This is a graph of probability of acceptance against quality level. Note that if the lot is of size N, there are only a finite number of possible values of the lot fraction nonconforming, namely 0, $1/N$, $2/N$, ..., $(N - 1)/N$, 1. Strictly speaking, the operating characteristic curve for isolated lots is therefore not really a curve, for it will only exist at these values, i.e. it will appear as a series of dots. This type of OC curve is called Type A. By comparison, the process quality level could be any value in the range from 0 % to 100 %, so the operating characteristic appears as a curve, called Type B. Figure 31 shows both types of OC curve for plans with a sample of size $n = 32$ drawn from a lot of size $N = 100$, where the lot is accepted when there are no more than Ac = 2 nonconforming items in the sample. Ac is called the *acceptance number* of the plan.

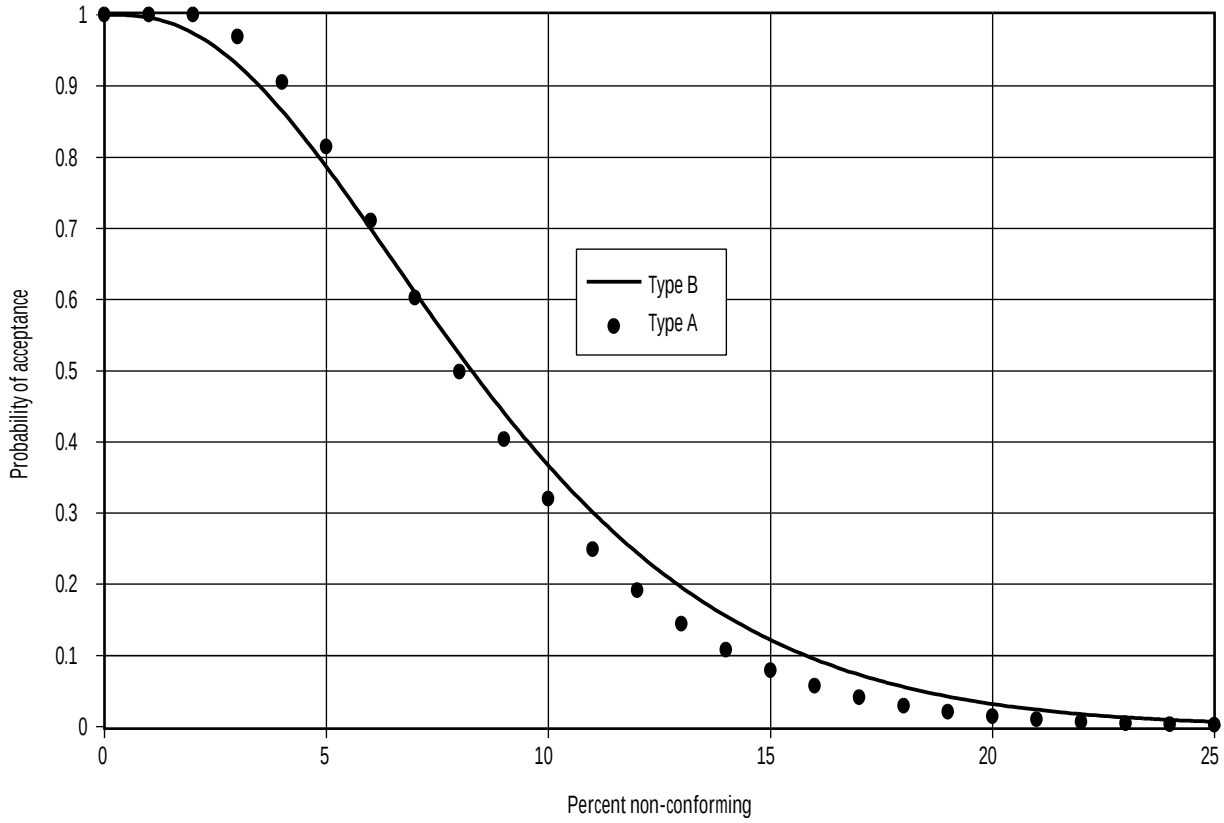


Figure 31 — Type A and B OC curves for n = 32, Ac = 2, N = 100

9.4.2 ***Single sampling***

The plans in **9.4.1** are both called *single sampling* plans, as the decision whether to accept or not to accept the lot depends on the results of a single sample. BS 6001-1 provides master tables of single sampling plans indexed by AQL for a continuing series of lots. (See **9.4.10** for LQ-indexed plans.) To find the appropriate plan, first the lot size and inspection level are used to determine the appropriate *sample size code letter* from Table 1 of the standard. Then this code letter together with the AQL are used to determine the sample size and acceptance number from Table 2-A for normal inspection, Table 2-B for tightened inspection or Table 2-C for reduced inspection.

These master tables have a very simple structure. The sample sizes are restricted to the set of 17 *preferred* sample sizes 2, 3, 5, 8, 13, 20, 32, 50, 80, 125, 200, 315, 500, 800, 1 250, 2 000, 3 150. These roughly form a geometric series with common ratio $10^{1/5} = 1.585$. The AQLs run from 0.01 % to 1 000 % also roughly as a geometric series with the same common ratio. (Above the AQL of 10 %, the plans are for use only for nonconformities per 100 items.) The result of this is that the acceptance numbers, which are also restricted to a series of preferred values, are the same along diagonals of the tables.

BS 6001-1 has recently undergone substantial revision. One of the changes is the introduction of optional *fractional acceptance number* plans between the diagonals for acceptance numbers zero and one, where in the previous edition there were arrows pointing upwards or downwards. They operate as follows. An acceptance acceptance number of $\frac{1}{k}$ when the sample size remains constant from lot to lot means that the present lot can be accepted if:

a) the sample from the present lot contains no nonconforming items; or

b) the sample from the present lot contains one nonconforming item, and the samples from the immediately preceding $(k - 1)$ lots contain no nonconforming items between them.

The fractional acceptance number plans are given in Tables 11-A, 11-B and 11-C for normal, tightened and reduced inspection. The fractions used are ⅓ and ½ for normal and for tightened inspection; for reduced inspection, where there are three diagonals between the acceptance numbers zero and one, the fractions used are $^1/_5$, ⅓ and ½.

The reason for introducing fractional acceptance number plans is that there is such a big difference between the OC curves for acceptance numbers zero and one, and often the desired OC curve is somewhere in between. Figure 32 illustrates how rapidly OC curves change shape in this range by showing the Type B OC curves for the plans with sample size 32 and acceptance numbers 0, ⅓, ½ and 1.

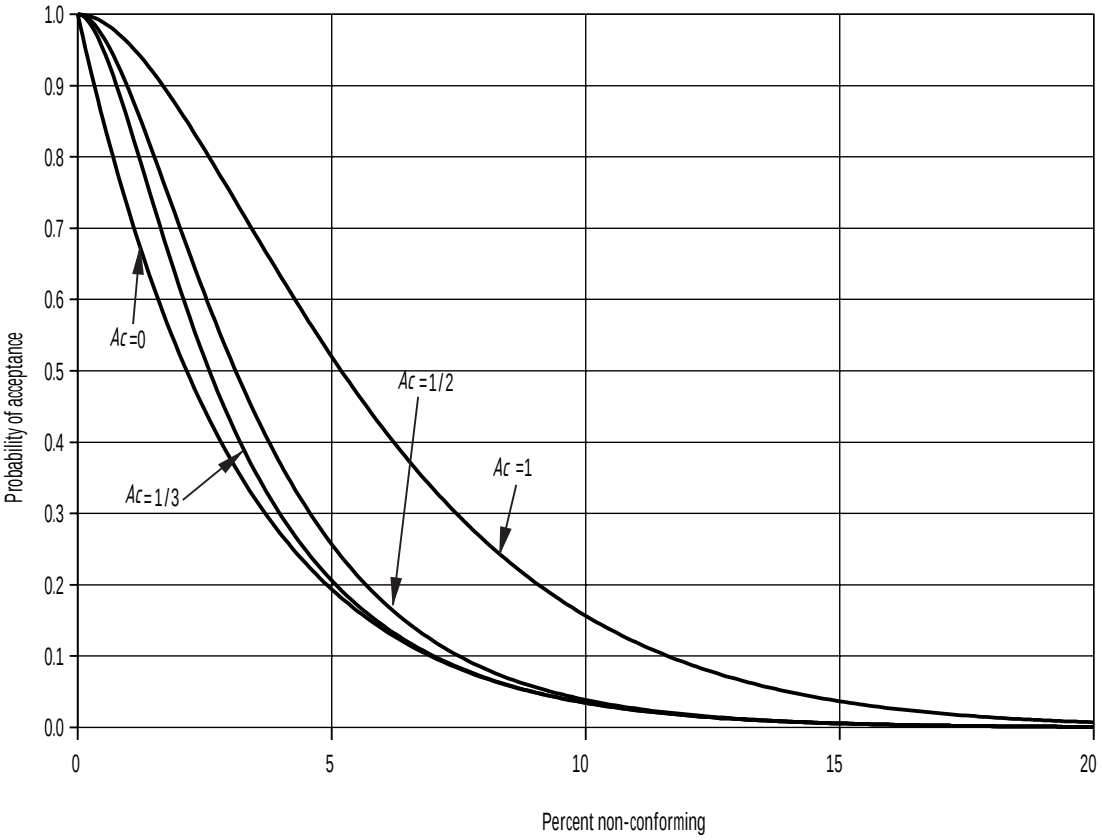


Figure 32 — Type B OC curves for Ac = 0, ⅓, ½ and 1

If the lot sizes change sufficiently to cause the sample sizes to vary from lot to lot, then the determination of whether the acceptance number for the present lot should be zero or one becomes rather more complicated. It is based on a cumulative *acceptance score*. This score is reset to zero whenever there is a switch to a different severity of inspection or whenever a nonconforming item is found. For the current lot it increases by 2, 3, 5 or 7 whenever the tabulated acceptance number is $^1/_5$, ⅓, ½ or at least 1 respectively, but remains unchanged if the tabulated acceptance number is 0. If the tabulated acceptance number is a whole number, it is used irrespective of the acceptance score, but if the tabulated acceptance number is a fraction, the acceptance number is taken to be zero if the score is 8 or below, and one otherwise. An added complication when the sample size changes is that the switching rules also become more complicated, requiring the maintenance of a *switching score*.

9.4.3 *Double sampling*

Double sampling plans provide one means by which the average amount of sampling may be reduced. Note that it is only the *average* that is reduced; for some lots the amount of sampling will be greater than for single sampling.

A double sampling plan by attributes works as follows. In general it has five parameters, which may be denoted n_1, n_2, c_1, c_2 and c_3. A random sample of size n_1 is taken from the lot and the number of nonconforming items d_1 is counted. There are three possible outcomes at this stage.

a) If $d_1 \leq c_1$ then the lot can be accepted without further sampling.

b) If $d_1 \geq c_2$ then the lot can be non-accepted without further sampling.

c) If $c_1 < d_1 < c_2$ then no immediate decision can be taken on the acceptability of the lot.

In case c), another random sample, this time of size n_2, is selected and the number of nonconforming items, d_2, in the sample is counted. The total number of nonconforming items found in the two samples is $d_3 = d_1 + d_2$. If $d_3 \leq c_3$ then the lot is accepted, otherwise it is non-accepted.

In summary, if the evidence from the first sample is very good or very bad, then an immediate decision can be taken. When the evidence is inconclusive, then a further sample is necessary to resolve the matter.

The integers c_1 and c_3 are the acceptance numbers of the plan. The integers c_2 and $c_3 + 1$ are the rejection numbers. Note that $c_2 - c_1$ has to be at least equal to 2, otherwise a decision on lot disposition will always be reached from the results of the first sample.

The procedure for nonconformities is the same as this except that "nonconforming items" is replaced throughout by "nonconformities".

In standards on sampling by attributes, the five parameters of each double sampling plan are chosen so that the OC curve of the double sampling plan roughly matches the OC curve of the corresponding single sampling plan. For simplicity and ease of operation, this matching is generally constrained so that the sample sizes n_1 and n_2 are equal to one another and so that the acceptance and rejection numbers are identical along diagonals of the master tables. Denoting the corresponding single sample size by n_0, it turns out that the double sample sizes are typically given by $n_1 = n_2 \cong 0.63n_0$. It follows that average savings in inspection effort of up to about 37 % of the single sample size may be achieved by using double sampling instead of single sampling, depending on the submitted quality.

The disadvantages of double sampling are their increased administrative and logistical requirements, which often lead to double sampling being impracticable. For example, suppose the acceptance test is to determine whether a device can survive 1 000 hours at 200 °C. It may be possible to test the devices simultaneously, so the test of a single sample would take 1 000 hours. However, if double sampling were used and the result from the first sample was inconclusive, then a second sample would be necessary. Testing of the second sample may not even be able to start at once if the test facility needs to be booked in advance. Coupling this with the time that the second sample requires to be tested, a decision on lot disposition will be substantially delayed. Meanwhile the lot will need to be stored somewhere, awaiting shipment.

Double sampling plans for sampling by attributes may be found for normal, tightened and reduced inspection in Tables 3-A, 3-B and 3-C of BS 6001-1 respectively, and in equivalent standards.

9.4.4 *Multiple sampling*

Multiple sampling takes the idea of double sampling a stage further. A k-stage multiple sampling plan has sample size n_i and acceptance and rejection numbers Ac_i, RE_i at the i^{th} stage, for $i = 1, 2, ..., k$. The present edition of BS 6001-1 has five-stage multiple plans with the sample sizes the same at each stage, and each equal to about one quarter of the corresponding single sample size. These five-stage plans represent an improvement over the seven-stage plans of the previous edition of BS 6001-1, in terms of both practicality and match with the OC curves of the corresponding single sampling plans. Again, the sets of acceptance and rejection numbers are kept the same along diagonals of the master tables, which are given as Tables 4-A, 4-B and 4-C in the standard.

As may be expected, multiple sampling plans provide a further reduction in average inspection requirements compared to double sampling plans. They are worthwhile provided the gains are not outweighed by logistical and administrative difficulties. At perfect quality, there may be as much as a 75 % saving in inspection costs when compared with single sampling plans with acceptance number greater than 5. For multiple sampling plans matching single sampling plans with acceptance numbers of 5 or lower, the maximum saving will be nearer to 50 % as a decision to accept will not be possible after the first multiple sample.

It is instructive to compare the properties of single, double and multiple sampling plans at different quality levels. Consider, for example, sample size code letter L in combination with an AQL of 2½ %. The plans to be compared are:

Single sampling: $n = 200$; Ac = 10, Re = 11;

Double sampling: $n_1 = n_2 = 125$; $Ac_1 = 5$, $Re_1 = 9$; $A_{C2} = 12$, $Re_2 = 13$;

Multiple sampling: $n_1 = n_2 = n_3 = n_4 = n_5 = 50$; $Ac_1 = 0$, $Re_1 = 5$; $Ac_2 = 3$, $Re_2 = 8$; $Ac_3 = 6$, $Re_3 = 10$; $Ac_4 = 9$, $Re_4 = 12$; $Ac_5 = 12$, $Re_5 = 13$.

Figure 33 shows the OC curves of the three plans. It can be seen that there is a very good match between all three.

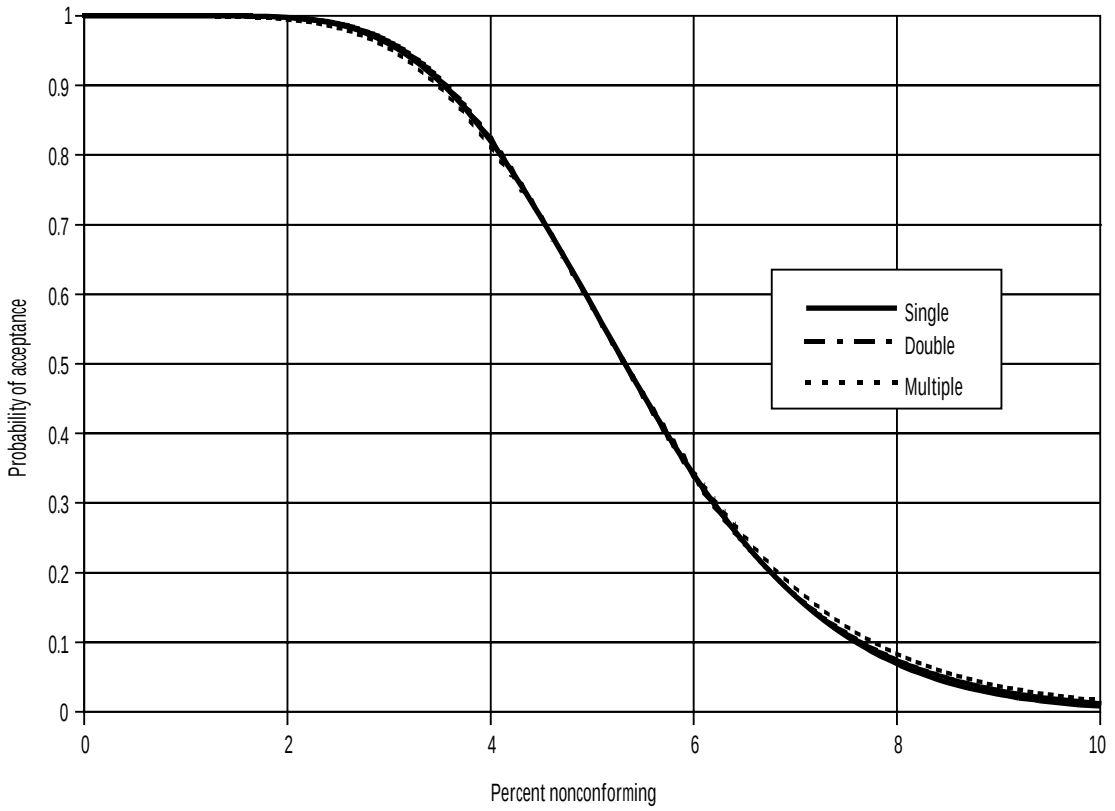


Figure 33 — OC curves for single, double and multiple sampling plans for sample size code letter L and AQL 2½ %

Figure 34 shows the average number of items that will be inspected at different quality levels, the so-called average sample number (ASN) curves, for the three types of plan.

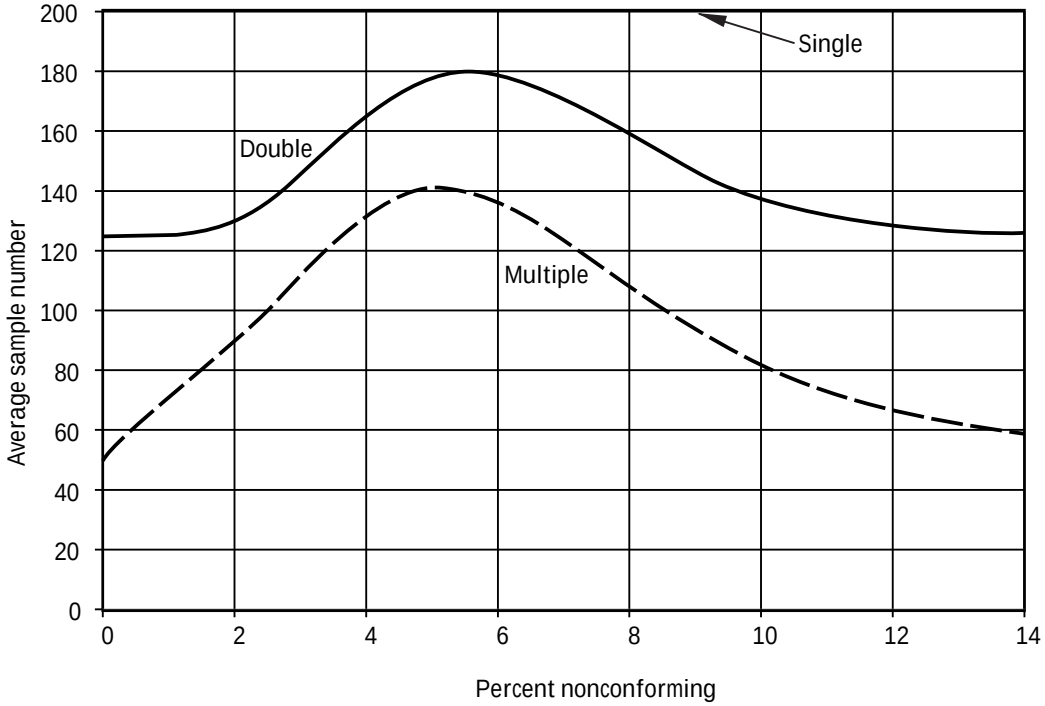


Figure 34 — ASN curves for single, double and multiple sampling plans for sample size code letter L and AQL 2½ %

Sometimes, particularly with destructive inspection, the main disincentive to using double and multiple sampling plans is the possibility that more items will need to be inspected than would be the case with single sampling. Figure 35 shows the probability that the corresponding single sample size is exceeded for the double and multiple sampling plans. In the case of double sampling plans in general, this is the probability of needing a second sample to come to a decision. In the case of this multiple sampling plan, it is the probability of needing all five samples to come to a decision. (Note that for multiple sampling plans where the single sample size is not divisible by four, needing the fourth sample from the BS 6001-1 multiple sampling plans may lead to the single sample size being exceeded, but not significantly.)

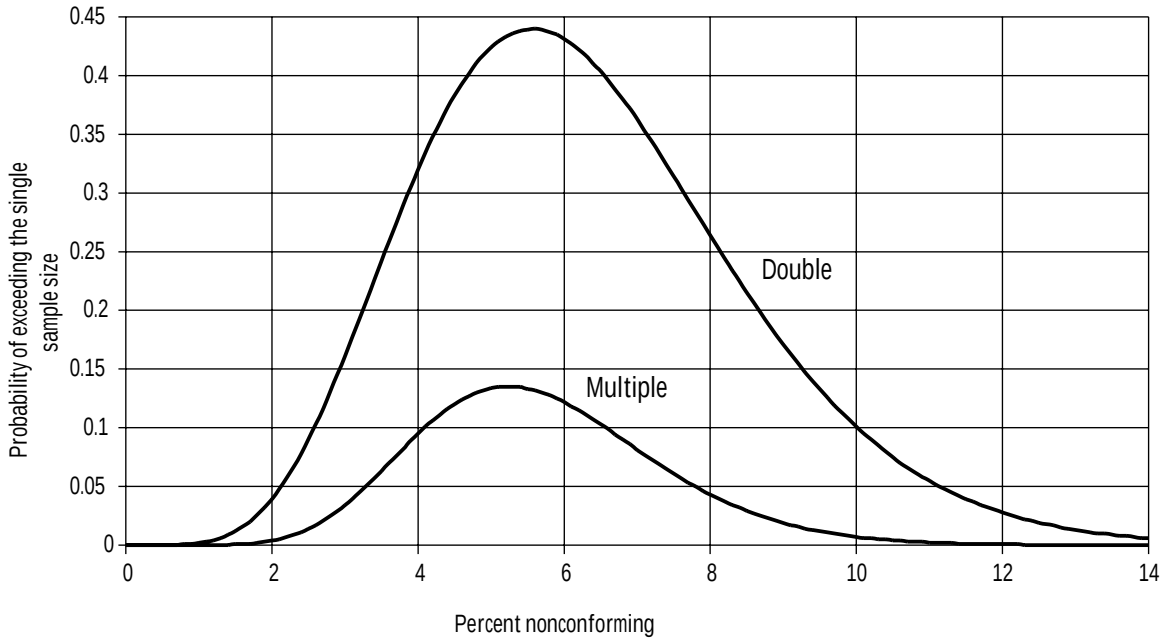


Figure 35 — Curves for the double and multiple sampling plans for sample size code letter L and AQL 2½ % showing the probability of needing to inspect significantly more sample items than under single sampling

Figure 35 shows clearly that another advantage that multiple sampling has over double sampling is a very substantial reduction in the chance of needing to inspect significantly more items than under single sampling.

9.4.5 *Sequential sampling*

The ultimate multi-stage procedure is to inspect items one at a time, making a decision after each inspection either to accept the lot, not to accept the lot or to continue sampling. This is called sequential sampling. Wald [17] devised an approximate method of determining the acceptance and rejection numbers at each cumulative sample size that provide specified values of the overall supplier's and customer's risks. It turns out that, on a graph of cumulative number of nonconforming items against cumulative sample size, the boundaries of the "continue sampling" region are parallel straight lines, as shown in Figure 36.

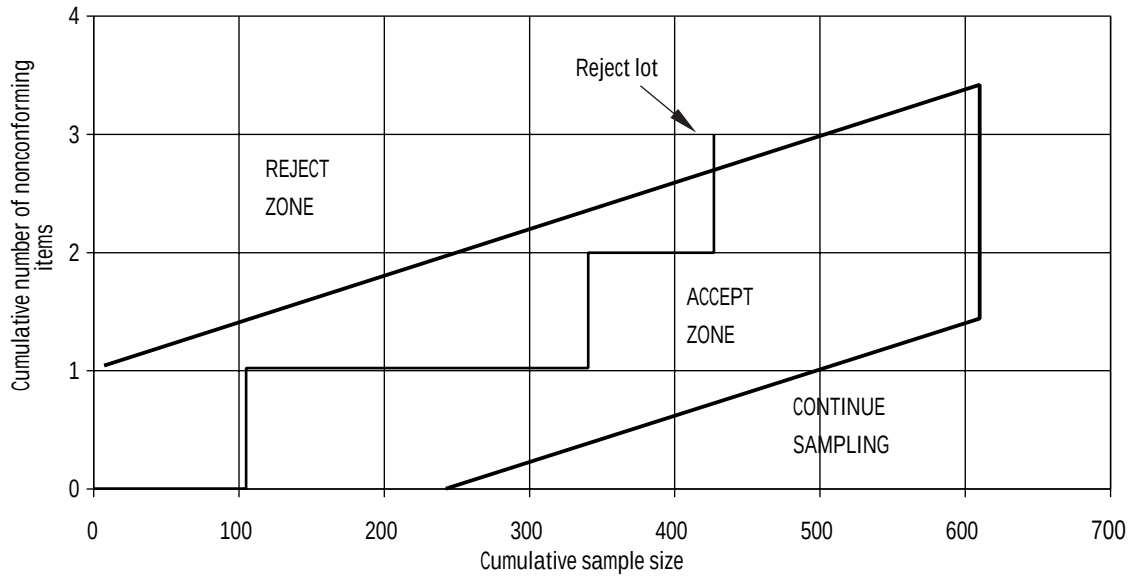


Figure 36 — Example of sequential sampling by attributes for percent nonconforming

The diagram for nonconformities looks similar, except that jumps of more than one nonconformity on the vertical axis are possible for an increase of one in the cumulative sample size.

The first edition of ISO 8422 was based upon the Wald approximation. Baillie [18, 19] demonstrated that, although Wald's method works well when the supplier's and customer's risks are no more than about 1 % or 2 %, the method can be very inaccurate when these risks are 5 % and 10 % respectively, as they were designed to be in ISO 8422. In fact, the resulting plans often have a supplier's risk much less than 5 %, while the customer's risk sometimes exceeds 10 %, the net effect being to require more inspection than necessary on average, i.e. a higher ASN.

Another complication is that the plans in ISO 8422 are curtailed at 1½ times the corresponding single sampling size, further distorting the risks from the design values.

For example, the Wald approximation to the plan for nonconforming items for a nominal 5 % supplier's risk at a quality level of 0.1 % nonconforming and a nominal 10 % customer's risk at 1.0 % nonconforming turns out to have a 2.92 % customer's risk and an 11.13 % supplier's risk. However, by suitable choice of the plan parameters (to three decimal places of accuracy) it is possible to achieve risks of 4.99 % and 10.00 % respectively and consequentially lower values of the ASN.

By taking the approximate values of the parameters from the Wald approximation and iteratively adjusting the gradient and position of the parallel lines, plans for the second edition of ISO 8422 have been obtained that will have supplier's and customer's risks much closer to the nominal values.

9.4.6 *Continuous sampling*

When items are produced in a continuous stream, there may be no natural way of dividing production into lots for the purposes of acceptance sampling. It is for such cases that continuous sampling plans were devised. The first of these, and the best known, is the Dodge [20] CSP-1 plan. This works as follows. First, a sampling frequency, f, and a clearance number, i, are specified. Then 100 % inspection begins. Once i successive conforming items have been inspected, 100 % inspection ceases and sampling inspection begins, with items being selected for inspection with probability f. As soon as a nonconforming item is found, inspection reverts to the 100 % level.

Many variations on this theme have been developed subsequently, with extra sampling frequencies and different rules for returning to 100 % inspection. The US Military Standard MIL-STD 1235A contains five types of continuous sampling plan, indexed by the average outgoing quality limit (AOQL), i.e. the worst possible average outgoing quality over all values of incoming quality p. These plans were designed to have AOQLs that matched the AOQLs of the standard single sampling attributes schemes such as those in BS 6001-1. All suffer from the same disadvantages: phases of 100 % inspection, which may be impracticable or uneconomic, and rapidly changing requirements for inspection personnel.

Beattie [21] proposed a different type of continuous sampling plan based on cumulative sums (cusums, see **10.7.3**) on the number of nonconforming items. A cumulative sum is begun at a specified sampling frequency of one item in f. At the conclusion of accumulating and inspecting each sample of given size, n, the cumulative sum is increased by an amount $(d - k)$, where d is the number of nonconforming items in the sample and k is a specified target reference value. Until the cusum reaches or exceeds a specified upper limit, h, product is accepted. When h is reached or exceeded, a new cumulative sum is started at a specified value, $h + h'$; this could be at a different sampling frequency, but inspection requirements can be kept constant by keeping to the same frequency. Product is non-accepted until the second cusum reaches as low as, or goes below, h. Then a new cusum is started at zero, and the process starts all over again.

The problem with designing a Beattie-type system of acceptance sampling plans is determining how to index them, i.e. determining which performance requirements should be mapped into values of n, f, h and h'. The probability of acceptance at a given quality level does not mean quite the same as for lot-by-lot inspection, as it is the probability of acceptance of an item, rather than a lot. Read and Beattie [22] introduced this probability of acceptance as the Type C OC curve. They defined it, for any quality level p, as the ratio of the average run length (ARL) for the cusum in the acceptance zone to the sum of the ARLs in the acceptance and rejection zones, i.e.:

$$P_a(p) = \frac{ARL_A(p)}{ARL_A(p) + ARL_R(p)}$$

Points on this OC curve could be specified as a way of identifying performance requirements. A related requirement, assuming rectification of nonconforming items found in samples in the acceptance zone and all product in the rejection zone, would be a specified AOQL.

Wadsworth and Wasserman [23], based on the work of Wasserman [24], devised design guidelines for Beattie-type cusum procedures for normally distributed variables and for variables with the Poisson distribution. They proposed these as the basis of a national or international standard. This is presently under consideration by ISO.

9.4.7 *Skip-lot sampling*

It has already been mentioned that inspection severity can be optionally switched to reduced inspection when the quality of successive lots remains consistently at a high level. For some types of product, the savings in switching to reduced inspection may not be very great. For example, the lot may still be delayed while inspection takes place. Moreover, the inspectors may still need to travel to the place of manufacture; or alternatively the sample, however much smaller than under normal inspection, may still need to be transported to a test facility. In short, many of the fixed costs may still have to be borne.

In response to these concerns, Dodge [25] developed the first skip-lot plan. In essence, it was a CSP-1 plan applied to lots instead of items; it was intended for use on homogeneous bulk materials from a reliable source where a single test result from each lot would determine its acceptability. Dodge and Perry [26] extended the idea as an overlay to a reference plan for lots consisting of discrete items. It operates as follows. The reference plan is used on successive lots until i successive lots have been accepted, at which point skip-lot sampling is introduced with a fraction f of lots, which are chosen at random, undergoing inspection still in accordance with the reference plan. As soon as a lot is non-accepted, application of the reference plan to every lot is resumed, and the process is repeated. By this means, some of the fixed costs may be avoided.

Liebesman and Saperstein [27] and Liebesman [28] developed a more sophisticated three-state procedure, which has been implemented as BS 6001-3. To quote from Liebesman [28]:

> "Three states are defined as part of the skip-lot standard: (1) state 1, lot-by-lot sampling, (2) state 2, skip-lot sampling, and (3) skip-lot interrupt. Qualification takes place during state 1 and requires acceptance of 10 lots in a row, the last two satisfying an individual lot criterion and the cumulative number of defects in the 10 lots satisfying a limit number criterion. When the program for a product is in state 2, interrupt occurs when a lot fails to satisfy the individual lot criterion. The program then transfers to state 3. During state 3, the product either re-qualifies for skip-lot having 4 lots in a row accepted with the last two satisfying the individual criterion; or the product becomes disqualified if a lot is rejected or the product is in state 3 for 10 lots."

Within the skip-lot state, there are four levels for the sampling frequency f, namely ½, ⅓, ¼ and $^1/_5$. The procedure is designed to encourage the supplier to maintain a quality level at half the AQL or better.

9.4.8 *Audit sampling*

There is a need for sampling procedures suited to formal, systematic inspections such as reviews or audits, to relieve the user from the problem of determining the appropriate sample size from formulae. ISO 2859-4 has been developed in response to this need. The plans are designed to provide a risk of less than 5 % of wrongly contradicting a correctly declared quality level. In order to keep the sample sizes to reasonable levels, a relatively large risk is allowed of failing to contradict an erroneously declared quality level. Three levels of discrimination are provided, and the plans are recognizable as a subset of the familiar single sampling plans of BS 6001-1. Part 4 is couched in terms of percent nonconforming items, but by a simple change of wording throughout it can also be applied to nonconformities per 100 items.

Note that, strictly speaking, audit sampling is hypothesis testing rather than acceptance sampling despite the fact that the plans have been drawn from BS 6001-1, which is an acceptance sampling standard.

9.4.9 *Sampling for parts per million*

For very good quality levels, typically measured in nonconforming items per million items, there are two difficulties with sampling inspection by attributes. One is the very large sample sizes that would be required to have supplier's and customer's risks as small as is generally the case with acceptance sampling plans. The second is that, quite often, no nonconforming items will be found; the result of this is that the unbiased estimator of the process fraction nonconforming, formed by dividing the number of nonconforming items in the sample by the sample size, will often take the unrealistic value zero.

ISO 14560 (under development) presents plans for this situation for lot-by-lot sampling. The first problem is overcome by allowing larger supplier's and customer's risks. Instead of the usual values for these risks of around 5 % and 10 % respectively, they are increased to as much as 10 % and 20 %. The second problem is overcome by using an estimator that will overestimate p, the process fraction nonconforming, about half the time and underestimate it about half the time. The approximate formula for this estimator given in the standard is:

$$\hat{p} = \frac{x + 0.7}{n + 0.4} \times 10^6 \text{ items per million items}$$

where x is the number of nonconforming items found in a sample of size n.

The plans are indexed by limiting quality level. Smaller sample sizes are required at better quality levels, thereby providing an incentive to the supplier to improve quality.

9.4.10 *Isolated lots*

When only one lot is being supplied, or a short series of lots, the protection afforded to the customer by the switching rules no longer applies. Attention then focuses on ensuring that any individual lot of a quality worse than a specified value has a low probability of acceptance. BS 6001-2 is a sampling system indexed by the limiting quality. The probability of acceptance at the LQ is usually no greater than 10 %, but in some cases it is as high as 13 %. Two procedures are given. Procedure A is intended for use when the supplier and the customer both wish to regard the lot in isolation. Procedure B is for use when the supplier considers the lot to be one of a continuing series while the customer regards the lot to be received in isolation.

The first edition of BS 6001-2 was designed for inspection for percent nonconforming. The second edition is planned to also include inspection procedures for nonconformities per 100 items.

9.4.11 *Accept-zero plans*

Two of the quality world's buzz-phrases that originated in the late 20th century are "get it right first time" and the "zero defects philosophy". In conjunction with the need to strive for ever-higher levels of quality, these words have often been taken to imply that if one or more nonconforming items are found in a sample, then the lot should not be accepted. In other words, the implication for sampling by attributes was interpreted to be that only plans with acceptance number zero should be admissible. Such plans are referred to as *accept-zero* plans.

A system of accept-zero sampling plans should consist of rules for switching from one sample size to another in response to quality history, if necessary supported by master tables of sample sizes. Ideally, the switching rules would guarantee some property of the outgoing product. Squeglia [29] roughly matches accept-zero plans to the normal inspection single sampling plans of BS 6001-1 at the LQ, but provides no switching rules whatsoever. US Military Standard MIL-STD-1916 provides accept-zero plans for normal, tightened and reduced inspection with similar switching rules to those of BS 6001-1. The MIL-STD plans provide seven "verification" levels (rather than AQLs) and five sample size code letters. Unfortunately, the guidance as to choice of verification level is merely that higher numbered levels require more inspection and should be applied to more important characteristics; it is not clear precisely what each verification level is designed to achieve.

Klaassen [30] derived a remarkably simple formula for determining the accept-zero sample sizes from successive lots that would be needed to guarantee an average outgoing quality limit (AOQL). Defining K, the "credit", as the total number of items accepted since the last lot non-acceptance, he showed that the sample size required to guarantee an AOQL of a is given by the smallest integer, n, such that:

$$n \geq \frac{N}{(K + N)a + 1}$$

where N is the next lot size. This assumes that lots that are non-accepted when $K = 0$ are 100 % inspected and that all conforming items found in such lots are accepted. The advantages of this method are threefold:

a) no tables are required for its implementation;

b) a single quantity, K, is sufficient to summarize the quality history; and

c) the AOQL guarantee is valid regardless of the sizes of successive lots or the sequence of lot qualities, rendering the method virtually abuse-proof.

To illustrate this method, suppose for a certain item that it is required that the average quality reaching the market place does not exceed 1.5 % nonconforming in the long term, i.e. $a = 0.015$. Suppose that a supplier always submits lots of the same size, N. K is initialized to zero. Suppose first that $N = 200$. If no nonconforming items are found in each sample, the sample sizes found by using the Klaassen formula are:

50, 29, 20, 16, 13, 11, 10, 8, ...

Even if the lot sizes are huge, the sample sizes for this AOQL for successive lots never increase above:

67, 34, 23, 17, 14, 12, 10, 9, ...

However, whenever a nonconforming item is found in a sample, the lot has to be screened and conforming items accepted, K has to be reset to zero and the sample size for the next lot needs to immediately return to the one at the beginning of the sequence.

Note that the method does not guarantee that the outgoing quality will not exceed the AOQL over any particular sequence of lots. The guarantee applies to the *long-term average*, or *expected*, outgoing quality over the whole sequence. Over a short series of lots there will be an appreciable probability that the AOQL will be exceeded, but this probability will tend to zero as the length of the series increases.

The possibility of a future ISO or British Standard for this method is under active consideration.

9.5 Acceptance sampling by variables — Single quality characteristic

9.5.1 *General*

For quality characteristics that are variables distributed according to a known family, it is possible to utilize this extra information to produce sampling plans that are more efficient. Most procedures for acceptance sampling by variables are for data from normal distributions, and discussion in this clause will be confined to the normal distribution case. Where possible, the procedures will be explained with reference to BS 6002. In that standard, as in BS 6001, the choice of inspection level and lot size determines a sample size code letter which, in conjunction with an AQL and an inspection severity, determines the sampling plan.

The procedures are classified as Form 1 or Form 2, depending on whether the process fraction nonconforming is estimated implicitly or explicitly. In all cases, it is assumed that there is a continuing series of lots and that the process fraction nonconforming is the subject of assessment. Consequently, the methods have much in common with statistical tolerance intervals (see **8.8.1**). Type A OC curves are not relevant to acceptance sampling by variables, as the presumption that the sampled population is normal cannot be true if the population is a finite lot.

The control of double specification limits on a variable is treated in one of three ways:

a) *combined control* is when a single AQL is applied to the sum of the process fractions nonconforming below the lower specification limit or above the upper specification limit;

b) *separate control* is when one AQL is applied to the lower specification limit and another AQL is applied at the upper specification limit;

c) *complex control* is when the one AQL is applied at either the lower or upper specification limit and a larger AQL is applied to the sum of the process fractions nonconforming beyond both of the specification limits.

The 1993 edition of BS 6002 only contains Form 1 procedures for a single variable. The next edition will have a Part 1 concerned with Form 1 procedures for a single variable and a single AQL. Part 2 will provide a more comprehensive coverage, using Form 2.

The symbols L and U will be used to denote the lower and upper specification limits.

Finally, a distinction is made between the "*s*" method and the "σ" method. The expression "*s*" *method* indicates that neither the process mean nor the process standard deviation is known. The expression "σ" *method* indicates that the process mean is unknown but the process standard deviation may be presumed to be known; in practice this will mean that σ is known within a small margin of error.

9.5.2 *Single sampling plans by variables for known process standard deviation — The "σ" method*

For a single, normally distributed quality characteristic with known σ, the acceptability of a lot may be determined as follows. Suppose that the Form 1 acceptance sampling plan has been determined, say by reference to BS 6002. The plan will consist of a sample size, n, and an *acceptability constant*, k. A random sample of size n is drawn from the lot, and the sample mean, $\bar{x}$, is calculated. The *quality statistic*:

$$Q_L = \frac{\bar{x} - L}{\sigma}$$

is calculated for a lower specification limit, and/or:

$$Q_U = \frac{U - \bar{x}}{\sigma}$$

for an upper specification limit.

For a single specification limit, the lot is acceptable only if the quality statistic exceeds k. For double specification limits, and before inspection begins, it first needs to be checked that σ is not so big that the AQL requirements are impossible to meet under tightened inspection. This is done by comparing σ with $(U - L)$ times the tabulated value of the standardized maximum (allowable) process standard deviation (MPSD). The reason for this is that it is pointless to begin sampling inspection if the process variation is too large for the switching rules to function. For separate control of double specification limits, there will be two acceptability constants, say k_L and k_U, corresponding to the AQLs at each limit; in this case, the lot is acceptable only if both quality statistics exceed their respective acceptability constants. For combined control, both quality statistics have to exceed k. For complex control involving a separate AQL requirement on, say, the lower specification limit, the acceptance criterion would be similar to that for separate control except that k_U would correspond to the combined part of the requirement. Similar remarks apply to complex control involving a separate AQL requirement on the upper specification limit.

For the σ known case, the calculation of the quality statistics for each sample may be avoided. For example, the acceptability criterion $Q_L \geq k_L$ may be converted into $\bar{x} \geq L + \sigma k_L = \bar{x}_L$, say, which can be calculated in advance. Similarly, $Q_U \geq k_U$ may be converted into $\bar{x} \leq \bar{x}_U$.

For Form 2, the acceptability constants are maximum acceptable values of the estimated process quality level. They will be denoted by the symbol p^*. The quality statistics are calculated in the same way as for Form 1. Denoting a quality statistic in general by Q, the process fraction beyond a single specification limit is estimated by the area under the standard normal curve above the value $Q\sqrt{n/(n-1)}$. Denoting these estimates at the lower and upper specification limits by $\hat{p}_L$ and $\hat{p}_U$, the lot acceptance criteria become:

a) for a single lower specification limit, $\hat{p}_L \leq p^*$;

b) for a single upper specification limit, $\hat{p}_U \leq p^*$;

c) for combined control of double specification limits, $\hat{p}_L + \hat{p}_U \leq p^*$;

d) for separate control of double specification limits, $\hat{p}_L \leq p_L^*$ and $\hat{p}_U \leq p_U^*$;

e) for complex control of double specification limits, either $\hat{p}_L \leq p_L^*$ and $\hat{p}_L + \hat{p}_U \leq p^*$ or $\hat{p}_U \leq p_U^*$ and $\hat{p}_L + \hat{p}_U \leq p^*$.

9.5.3 *Single sampling plans by variables for unknown process standard deviation — the "s" method*

When the process standard deviation cannot be presumed to be known, it is estimated by the sample standard deviation, s. The quality statistics become the following:

$$Q_L = \frac{\bar{x} - L}{s} \text{ and } Q_U = \frac{U - \bar{x}}{s}$$

The Form 1 acceptability constants (k values) become larger than for the "σ" method and the Form 2 acceptability constants (p^* values) become smaller, to allow for the increased uncertainty in the estimation of the process quality.

Consider first Form 1. For a single specification limit, the acceptance criterion is similar to those for the "σ" method, i.e. the lot is acceptable if $Q \geq k$. Figure 37 shows an acceptance chart for a lower specification limit on a graph of $\bar{x}$ against s, for sample size code letter G on normal inspection (giving sample size 18) with an AQL of 1 % (giving $k = 1.77$). The accept zone is bounded by the line $\bar{x} = L + ks$, where the lower specification limit L has been taken to be 30 units.

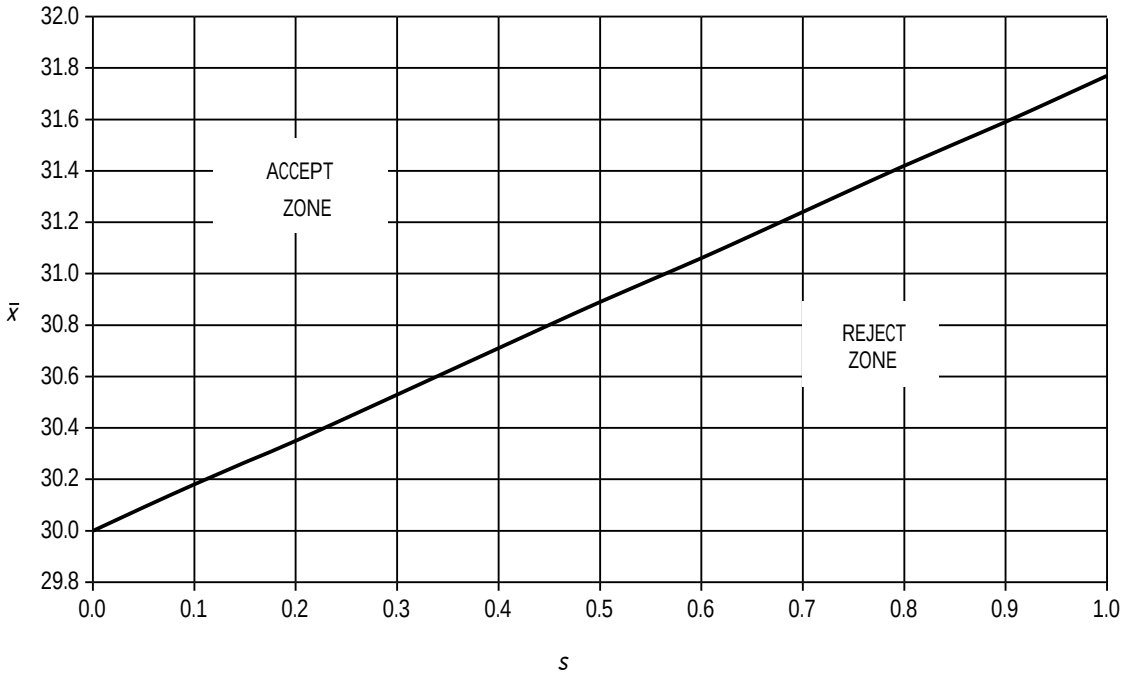


Figure 37 — Acceptance chart for a lower specification limit

For double specification limits, no check on σ can be carried out before inspection begins because σ is unknown. Nevertheless, an initial test may still be carried out on the process potential by comparing s with a maximum (allowable) sample standard deviation (MSSD). The MSSD is found by multiplying $(U - L)$ by a tabulated standardized value. For separate control of double specification limits, the lot is acceptable only if $Q_L \geq k_L$ and $Q_U \geq k_U$. An acceptance chart for separate control is shown in Figure 38; it can be seen in this case that the accept zone is bounded by two straight lines.

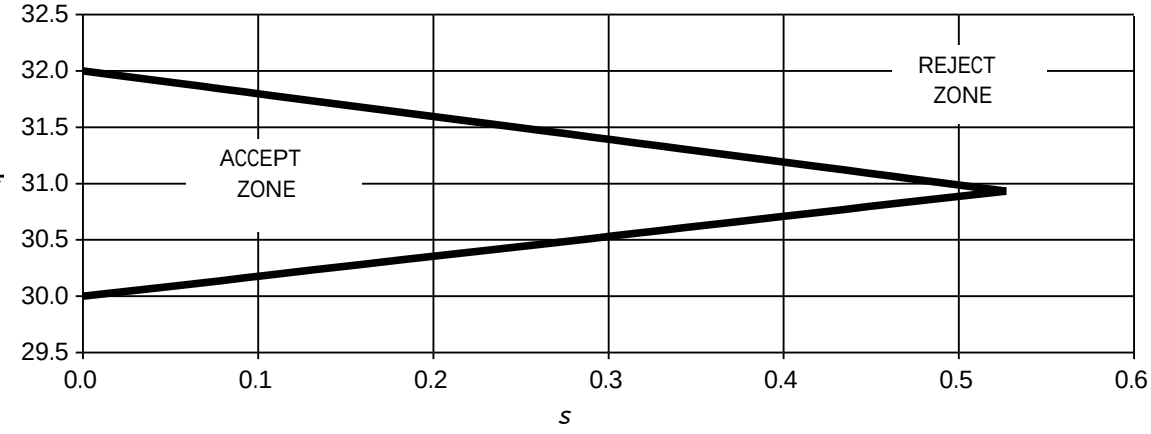


Figure 38 — Acceptance chart for a double specification limits with separate control

For combined control, the acceptability of the lot is determined by plotting the point with standardized co-ordinates $[s/(U - L), (\overline{x} - L)/(U - L)]$ on a standardized chart for the given sample size and AQL. The lot is accepted if the point lies within the acceptance region. Figure 39 shows such a chart for sample size 18 with an AQL of 4 %.

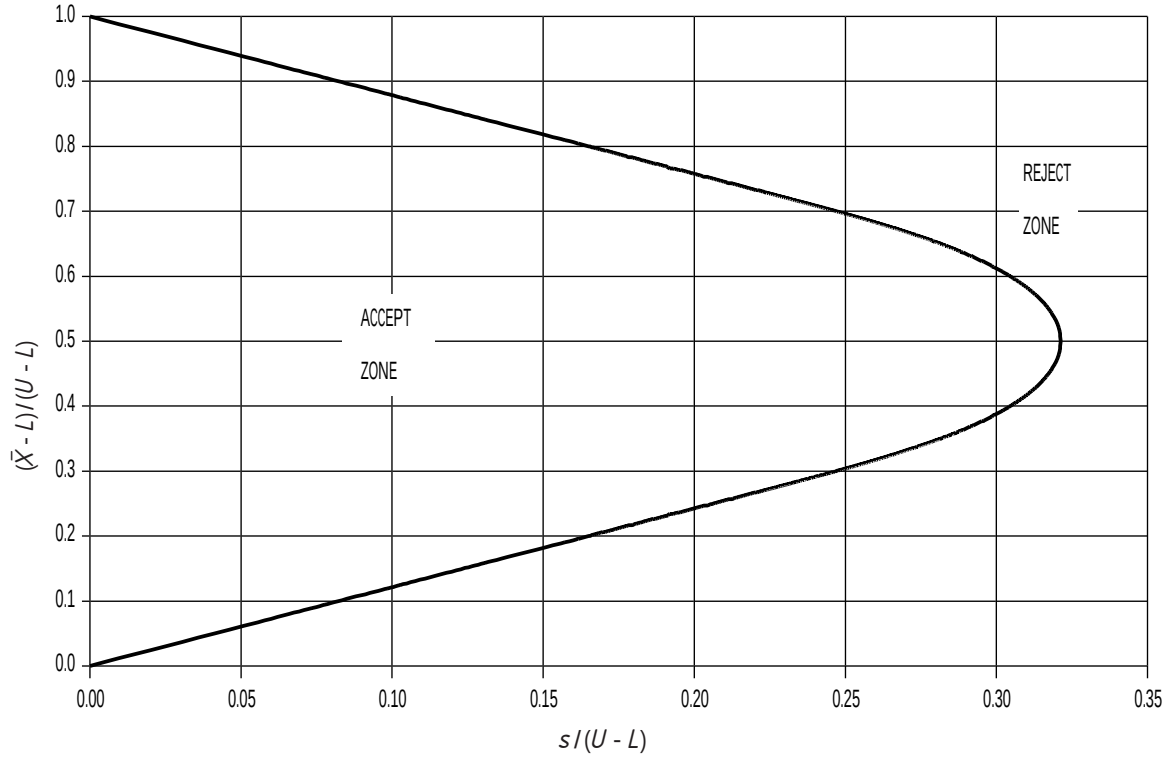


Figure 39 — Standardized acceptance chart for sample size 18 for double specification limits with combined control at an AQL of 4 % under normal inspection

For complex control, lot acceptability is determined in the same way as for combined control except that part of the acceptance region is eliminated in accordance with the requirement on the single specification limit. Figure 40 shows the acceptance region for sample size code letter G on normal inspection for a 1 % AQL at the upper specification limit and a 4 % AQL overall.

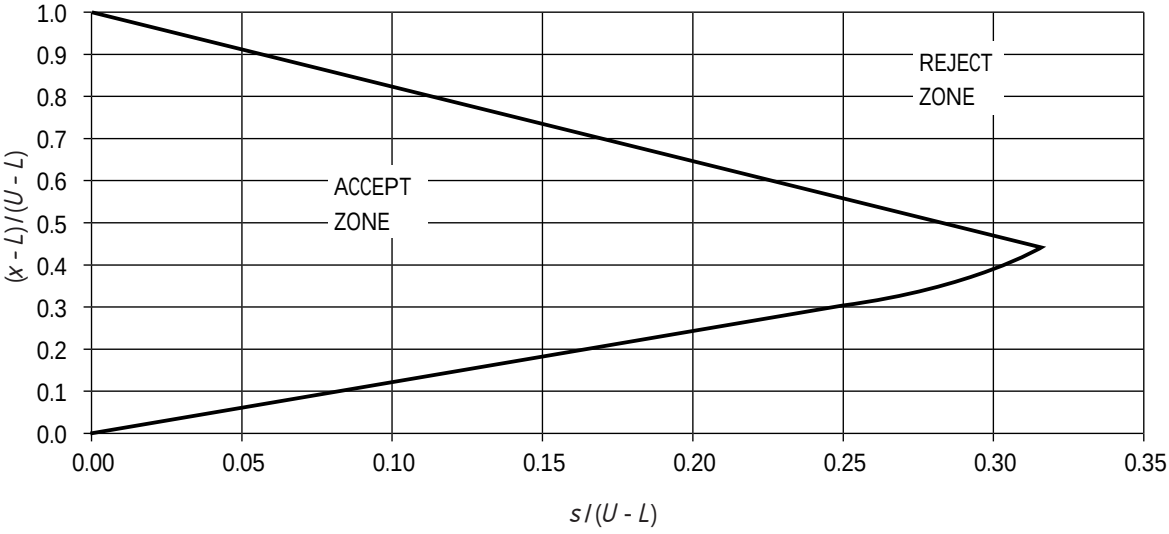


Figure 40 — Standardized acceptance chart for sample size 18 for double specification limits with complex control at an AQL of 1 % for the upper limit and an AQL of 4 % overall under normal inspection

For Form 2, the procedures are identical to a) through e) of **9.5.2** except in one respect. For the "*s*" method, the process fraction nonconforming beyond a single specification limit is estimated by the area to the left of the value $½ - ½Q\sqrt{n}/(n-1)$ under a symmetrical *beta* distribution that has both parameters equal to $½(n-2)$. In order to avoid the need to use tables of the beta distribution to implement Form 2 plans, Baillie [31] has developed a normal approximation for use when $n > 4$, which works as follows.

a) Set $x = ½ - ½Q\sqrt{n}/(n-1)$. If $x \leq 0$, then $\hat{p} = 0$, or if $x \geq 1$, then $\hat{p} = 1$; in both cases, no further steps are necessary. Otherwise, continue to step b).

b) Set $y = d_n \ln\{x/(1-x)\}$ where $d_n = ½\sqrt{(n-3)\left[1 + \frac{1}{3\{(n-3)^2+1\}}\right]}$.

c) Set $w = y^2 - 3$.

d) If $w > 0$, set $t = \frac{y}{1 + w/\{12(n-1)\}}$, otherwise set $t = \frac{y}{1 + w/\{12(n-2)\}}$.

e) Then $\hat{p}$ is approximated by the area to the left of t under the standard normal curve {usually denoted $\Phi(t)$}.

This approximation is quite accurate, guaranteeing a maximum absolute error of not more than 0.000 4 for sample size 5, 0.000 2 for sample size 6 and 0.000 1 for sample sizes of 8 or more. Values of d_n for selected values of n are given in Table 19.

Table 19 — Values of d_n for estimating the process fraction nonconforming

Sample size	d_n	Sample size	d_n	Sample size	d_n
3	0.318 310	15	1.734 040	70	4.092 828
4	0.551 329	18	1.937 919	75	4.242 777
5	0.731 350	20	2.062 737	95	4.795 926
6	0.880 496	25	2.346 014	100	4.924 516
7	1.009 784	30	2.598 669	125	5.522 742
8	1.125 182	35	2.828 887	150	6.062 225
9	1.230 248	40	3.041 751	160	6.265 024
10	1.327 276	45	3.240 676	200	7.017 865
13	1.583 745	50	3.428 086	250	7.858 138

To illustrate, suppose that there is a single, lower specification limit $L = 32$ and a sample of size 18 has a mean $\bar{x} = 34.1$ and a standard deviation $s = 0.93$. Then $Q = (\bar{x} - L)/s = (34.1 - 32)/0.93 = 2.258$. The value of $x = ½ - ½Q\sqrt{n}/(n-1)$ is found to be 0.218. From tables or a computer program, it may be found that the area to the left of 0.218 under a symmetric beta distribution with both parameters equal to $½(n-2) = 8$ is 0.007 4.

If neither the appropriate tables nor software are available, the normal approximation is found as follows, starting at step b). d_{18} is found from Table 19 to be 1.937 919. Then y is calculated as $1.937\,919 \times \ln(0.218/0.782) = -2.475\,4$. Then $w = y^2 - 3 = 3.128$. As $w > 0$, we set $t = y/[1 + w/\{12(n-1)\}] = -2.475\,4/(1 + 3.128/204) = -2.438$. From normal tables, it is found that the area under the normal curve to the left of -2.438 is 0.007 4. Thus, $\hat{p} = 0.007\,4$, in complete agreement with the exact method.

9.5.4 *Double sampling plans by variables*

A double sampling plan by variables can be formulated in either Form 1 or Form 2. Consider for illustration the case of σ unknown for a single lower specification limit L. Form 1 will be described here. A double sampling plan by variables has five parameters, namely the two sample sizes n_1 and n_2 and the three acceptability constants k_a, k_r and k_c. A sample of size n_1 is drawn, and the quality statistic $Q_1 = (\bar{x} - L)/s_1$ is calculated. If $Q_1 \geq k_a$, the lot is accepted. If $Q_1 \leq k_r$, the lot is non-accepted. If $k_r < Q_1 < k_a$, another sample, this time of size n_2, is drawn and its mean $\bar{x}_2$ and standard deviation s_2 are calculated. The combined mean is calculated as:

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

and the combined standard deviation as:

$$s_c = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

The combined quality statistic is calculated as:

$Q_c = (\overline{x}_c - L)/s_c$

If $Q_c \geq k_c$, the lot is accepted; otherwise it is non-accepted.

The procedure for the "σ" method is similar except that $Q_1 = (\overline{x}_1 - L)/\sigma$ and $Q_c = (\overline{x}_c - L)/\sigma$.

Hamaker [32] investigated the matching of σ-known double sampling plans by variables to σ-known single sampling plans. He observed that there were three requirements that are partly contradictory:

a) a reasonably close match between the OC curves;

b) a worthwhile reduction in the average sample number;

c) a low frequency of second sample (FSS);

and developed rules that provided a sensible balance between them. Baillie [33] extended Hamaker's results for $n_1 = n_2$ to the case of unknown σ. ISO 3951-3 will be developed in due course to provide double sampling plans by variables.

9.5.5 *Sequential sampling plans by variables for known process standard deviation*

ISO 8423 provides curtailed sequential sampling plans for inspection by variables when the process standard deviation is known, for single and for double specification limits. The acceptance chart for a single specification limit is similar in appearance to the attributes sequential chart (see Figure 36), the difference being that the cumulative sum of the *leeway* is plotted on the vertical scale. (The leeway is defined as $U - x$ for an upper specification limit and $x - L$ for a lower specification limit, where x is the measured value of the variable.) Thus, the increments on the vertical scale are not constrained to be integers, and can even be negative if x lies outside specification.

It was observed in **9.4.5** that the Wald approximation has been found to be poor for sampling by attributes when the supplier's and customer's risks are of the order of 5 % and 10 %. The same has been found to be true for sampling by variables so, like ISO 8422, ISO 8423 is being revised to provide a better match with the corresponding single sampling plans.

9.5.6 *Accept-zero plans by variables*

Accept-zero plans provide for lot acceptance if there are no nonconforming items in the sample. Denoting the smallest and largest observations in a sample by $x_{[1]}$ and $x_{[n]}$, this requires $x_{[1]} \geq L$ for a lower specification limit, $x_{[n]} \leq U$ for an upper specification limit, and both inequalities to be satisfied in the case of double specification limits.

Klaassen's [30] credit-based method of guaranteeing an AOQL for accept-zero plans for sampling by attributes was described in **9.4.11**. Effectively it provides a switching rule between sample sizes in response to perceived quality history. Baillie and Klaassen [34] have generalized this result to the case of guaranteeing an AOQL for any acceptance sampling plan that includes an accept-zero requirement, and applied the general method to the following three cases, with c any positive constant:

a) $x_{[1]} \geq L$ (for sampling by attributes);

b) $x_{[1]} \geq L + c\sigma$ (for sampling by variables with known σ); and

c) $x_{[1]} \geq cs$ (for sampling by variables with unknown σ).

Again, as with sampling by attributes, the AOQL guarantee requires lots that are rejected when the credit is zero to be 100 % inspected, with acceptance of all conforming items found in such lots. As expected, sampling by variables when the value of σ is presumed known requires smaller sample sizes than sampling by variables when the value of σ is unknown, which in turn requires smaller sample sizes than sampling by attributes.

9.6 Multiple quality characteristics

9.6.1 *Classification of quality characteristics*

Most products have more than one quality characteristic, all of which need to conform to specification if an item is to be classed as conforming. Some of these characteristics may be of greater importance, and may therefore need to be controlled more tightly. This is achieved by classifying the quality characteristics into class A for those of the highest level of importance, class B for the next level of importance, etc., and applying a low AQL to class A, a larger AQL to class B, etc. Sampling inspection schemes are then applied to the classes independently; for example, it would be possible for classes A and C to be on normal inspection while class B is on tightened inspection. The acceptance criteria for all classes have to be satisfied for a lot to be classified as acceptable. The following discussion of multiple quality characteristics is on the treatment of a single class of quality characteristics, where by definition all the quality characteristics in the class are of approximately equal importance to the integrity of the product.

9.6.2 *Unifying theme*

As before, the unifying theme of the discussion for nonconforming items will be the comparison of $\hat{p}$, the estimated process fraction nonconforming from the sample, with p^*, a specified maximum value. As a rough rule, we can set:

$$p^* = (\mathrm{AC} + ½)/n \qquad (23)$$

where the reference single sampling plan by attributes has sample size n and acceptance number Ac.

9.6.3 *Inspection by attributes for nonconforming items*

9.6.3.1 *Independent attributes*

Consider first the simplest case where there are k quality characteristics, all of which are attributes. First suppose that the attributes are independent, i.e. the probability of any one of the attributes in the class being out of specification is constant, regardless of the state of any of the other attributes in the class. Suppose also that in a sample of size n it is found that there are r_1 items that are nonconforming on attribute 1, r_2 items that are nonconforming on attribute 2, ..., r_k items that are nonconforming on attribute k. The estimate of the probability of *conformance* on the i^{th} attribute is estimated by $(1 - r_i/n)$. As the attributes are independent, the estimated overall probability of an item conforming to all the specifications is the product of such estimated probabilities, *viz.* $(1 - r_1/n)\ (1 - r_2/n)\ ...\ (1 - r_k/n)$. Subtracting this from 1, it is seen that the overall probability of an item *not* conforming to at least one of the specifications is estimated by:

$$\hat{p} = 1 - (1 - r_1/n)\ (1 - r_2/n)\ ...\ (1 - r_k/n).$$

The acceptance criterion $\hat{p} \leq p^*$ therefore becomes:

$$1 - (1 - r_1/n)\ (1 - r_2/n\ ...\ (1 - r_k/n) \leq p^*.$$

Provided all the fractions r_i/n are small, it can be shown by expanding the product term that this inequality is approximately the same as:

$$(r_1 + r_2 + ... + r_k) \leq n\,p^*, \text{ i.e. } r \leq c$$

where r is the total number of items that are out of specification with respect to each attribute, summed over attributes, and c is the largest whole number less than or equal to $n\,p^*$.

9.6.3.2 *Dependent attributes*

Now suppose that the attributes are dependent. The estimate of the process fraction nonconforming is $\hat{p} = d/n$, where d is the number of nonconforming items in the sample. The acceptance criterion $\hat{p} \leq p^*$ then becomes $d \leq n\,p^*$, i.e. $d \leq c$.

9.6.3.3 *Example*

To illustrate the difference between the independent and dependent cases, suppose that a single sampling plan under normal inspection is to be used, with a sample size code letter F and an AQL of 4 %. From Table 2-A of BS 6001-1 it is found that the sampling plan is $n = 20$, Ac = 2. From (23):

$$p^* = 2½/20 = 0.125.$$

Suppose that an item has two quality characteristics that are both attributes. A sample of size 20 yields one item that is nonconforming on both attributes and one item that is nonconforming on one attribute. Assuming independence between the attributes, the estimate of the process fraction nonconforming would be:

$$\hat{p} = 1 - (1 - 2/20)(1 - 1/20) = 0.145.$$

On the other hand, assuming dependence, there are only two nonconforming items in the sample of size 20, so the estimate of the process fraction nonconforming would be:

$$\hat{p} = 2/20 = 0.100.$$

As $p^* = 0.125$ we see that the lot would be non-accepted if the attributes were considered to be independent, but accepted if they were considered to be dependent. On reflection, this is not a surprising result. On the evidence from the example, when the two attributes are dependent there seems to be a tendency for both attributes to be out of specification at the same time. Treating the nonconformities as independent in such a situation leads to some double counting.

9.6.4 *Inspection by attributes for nonconformities*

Suppose that the sample contains a total of r_1 nonconformities on attribute 1, r_2 nonconformities on attribute 2, ..., r_k nonconformities on attribute k. The rate of nonconformity on the i^{th} attribute is estimated by r_i/n. If the attributes are independent, these estimated rates are added to give an estimated overall rate of nonconformity per item of r/n, where $r = r_1 + r_2 + ... + r_k$ is the total number of nonconformities in the sample.

On the other hand, if the attributes are dependent, the estimated rate is simply the total number of nonconformities divided by the sample size, i.e. r/n again. It follows that the multi-attribute acceptance criterion is $r \leq$ Ac; this is regardless of the number of attributes and whether or not they are independent.

9.6.5 *Independent variables*

The principle for k independent variables is the same as for k independent attributes inspected for nonconforming items. The process fraction nonconforming is estimated as:

$$\hat{p} = 1 - (1 - \hat{p}_1)(1 - \hat{p}_2)...(1 - \hat{p}_k)$$

where $\hat{p}_i$ is calculated as explained in **9.5.2** for σ known and in **9.5.3** for σ unknown.

Form 2 plans by variables will be presented in BS 6002-2, which is presently under development.

9.6.6 *Dependent variables*

For dependent variables, it is theoretically possible to carry out acceptance sampling, but impracticable without the use of suitable software, as the formula for $\hat{p}$ is a multidimensional integral over a complicated region. For further information, the reader is referred to Baillie [35]. If the correlation between the variables is not strong, they may be treated as independent without much danger of reaching the wrong decision on lot acceptability if the decision is not marginal. If the correlation between the variables is strong, then the variables can be converted to dependent attributes, and treated as described in **9.6.3.2**, although this is an inefficient use of the data.

9.6.7 *Attributes and variables*

Baillie [36] presented master tables of "attriables" plans and procedures for use when the quality characteristics in a class consist of at least one attribute and at least one variable. The plans are only suitable when it is practicable to have a larger sample size for the attributes than for the variables. The implementation of the plans is necessarily complicated, and would need to be supported by suitable software, particularly if there are two or more dependent variables.

10 Statistical process control (SPC)

10.1 Process focus

The question to what extent it is possible to obtain from a sample a reliable estimate of the quality characteristics of products lots has been discussed from various points of view in the preceding clauses. It has been shown that if the variation among individual units is considerable, it may not be an economic proposition for the customer to sample and test sufficient items to provide the desired degree of assurance regarding the consignment.

Furthermore, what if the correct technical decision, on the basis of an "after the event" sample, is to "*reject*" the consignment? All too often, the correct business decision has to be "*accept*" because of logistics, time and other constraints. It is therefore inevitable that attention should be focused on ways of securing and demonstrating conformity to specification which involve the requirement that statistical methods be deployed at the place and time of the process activity giving rise to the product or service. This is recognized in the following circumstances:

a) generic quality management system requirements such as ISO 9001:2000. This standard recognizes that any activity that receives inputs and converts them to outputs can be considered as a process. For organizations to function effectively, they have to identify and manage numerous linked processes. Often the output from one process will directly form the input of the next process. ISO 9001:2000 is based on the "process approach" to management which involves the systematic identification and management of the processes employed within an organization, and the interactions between such processes. Indeed it asserts that all the requirements of a quality management system for achieving conformity of product* may be placed within a process model such as that shown in Figure 41.

*NOTE The term product in the ISO 9000 family has four generic categories: hardware, software, services and processed materials.

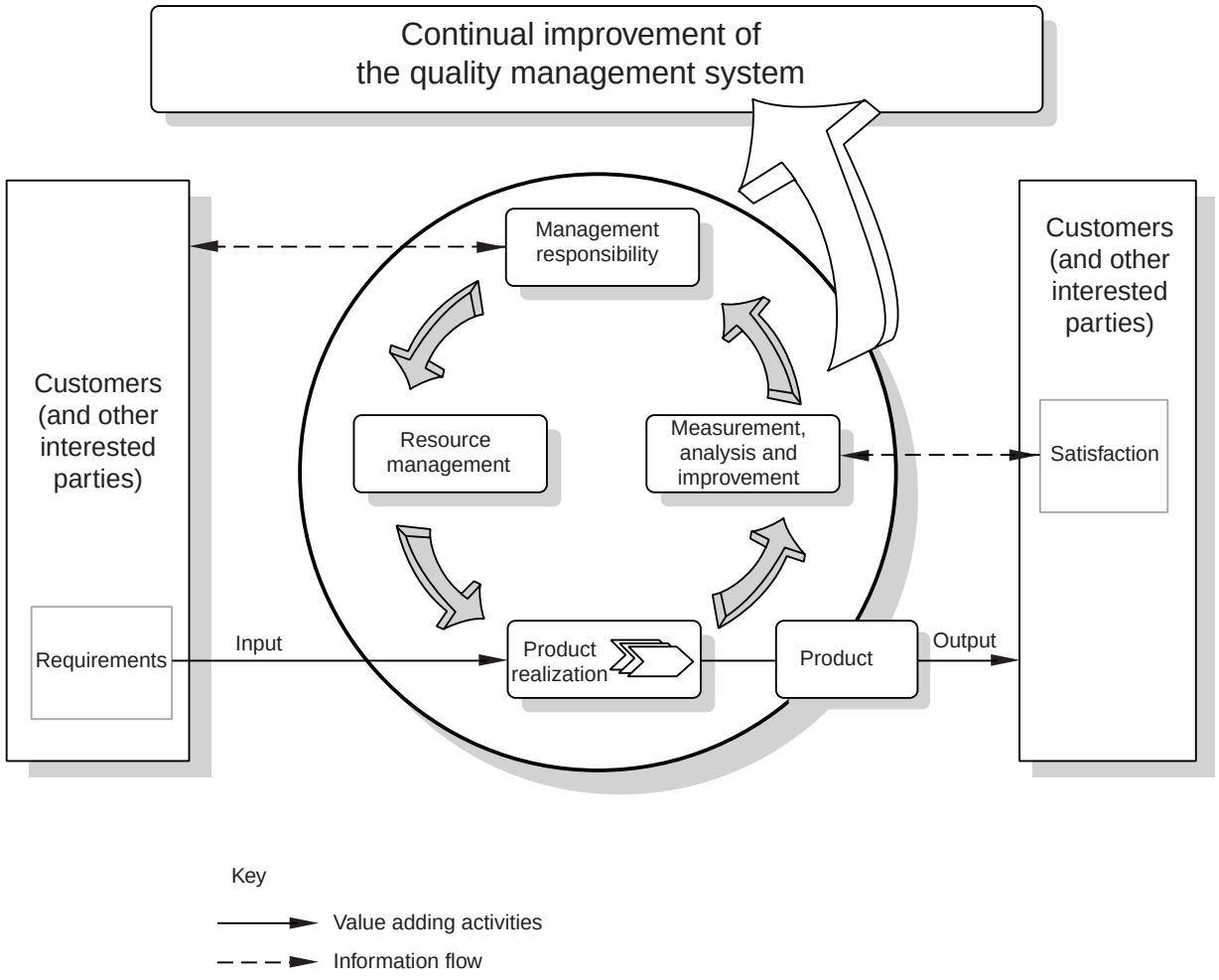


Figure 41 — ISO 9001:2000 Quality management process model

More specifically, there is a requirement clause in ISO 9001:2000 for measurement, analysis and improvement. It states that the organization should define, plan and implement the measurement and monitoring activities needed to assure conformity and achieve improvement. This should include the determination of the need for, and use of, applicable methodologies including statistical techniques.

ISO 9001:2000 requires that an organization should, for example:

1) apply suitable methods for measurement and monitoring of those realization processes necessary to meet customer requirements. These methods shall confirm the continuing ability of each process to satisfy its intended purpose;

2) measure and monitor the characteristics of the product to verify that requirements for the product are met. This shall be carried out at appropriate stages of the product realization process;

3) collect and analyse appropriate data to determine the suitability and effectiveness of the quality management system and to identify improvements that can be made;

4) plan and manage the processes necessary for the continual improvement of the quality management system;

5) facilitate the continual improvement of the quality management system through the use of analysis of data, corrective and improvement action, amongst others.

b) numerous sectors, for example, such as medical devices, aerospace and automotive have more prescriptive quality system requirements than ISO 9001:2000. Taking automotive, for example, three major USA based suppliers have jointly produced the QS-9000 quality systems requirements, together with supporting documentation which includes manuals on: *statistical process control* (SPC) [37] and *measurement system analysis* (MSA) [38] which provide a unified formal approach to both SPC and MSA in the automotive industry.

Regardless of whether or not the application of statistical process control is explicitly stated in system or product requirements, the organization dedicated to "never ending improvement" or aiming for "world class" will recognize its key role in improving business performance. This is illustrated by an example from the aircraft supply industry.

Example: Steel tube dimensions

A steel tube supplier to the aircraft industry buys steel strip from the steel maker by the kilogram and converts strip into tube to sell by the metre. This organization recognized that, by managing variation better, more metres could be produced per kilogram of strip. It aimed for preferred minimal values for outside diameter and wall thickness commensurate with the need to maintain dimensional specification requirements. This aim decreased as they identified and progressively reduced variation using statistical process control methodology. Controlling the new minimum size and its variation using statistical process control it was then able to produce a lighter, more consistent product with a saving of some £¼ m per year.

This demonstrates the ability of statistical process control to both increase profits for the user organization and value to its customers.

These considerations, amongst others, have given rise to a growth in the development and widespread application of statistical process control methods.

10.2 Essence of SPC

The primary operational tool of SPC is the control chart. The first question to be answered is: what is to be its basic purpose? The reason for this is that there are two fundamentally different approaches to control charting. One approach aims at directly controlling to specification using control limits based upon, and set *inwards* from, specified tolerance limits. Such an approach is described in ISO 7966:1993. The other approach uses control limits based entirely on process performance. These control limits are set *outwards* from the mean value of the characteristic plotted to an extent based on the inherent variability of the process with no regard to specified tolerance limits. Such an approach is described in ISO 8258:1991.

Many organizations using tolerance-based control charts have had to abandon their use in favour of performance based statistical process control, for contractual reasons, to meet customer requirements expressed in current quality systems standards. Others have chosen to adopt performance-based control charts, for a number of reasons, such as the following:

a) the recognition that first class (A1) quality for a characteristic is achieved only by realizing preferred value and that there is a progressive deterioration in quality as one moves away from this value towards a specification limit, even though one may, technically, still be "in tolerance". Meeting tolerance then becomes a "minimum" standard which may just be tolerated. It is not a standard of excellence. In a competitive climate there can be considerable advantage in aiming for preferred value with minimum variation;

b) the acknowledgement that a tolerance-based control chart does not provide information on the sources of variation in the process essential for control and improvement purposes. The focus on the classification of the process output purely in terms of specification is in direct contrast to the focus of the performance-based control chart which is on the discrimination between common and special cause variation in the process;

c) an appreciation that there are two kinds of people and two types of variation:

1) *technical and managerial people* who are responsible for the process and for the presence of *inherent/common cause* variation and its reduction;

2) *operational people* who work in the process who can best observe and report on *special cause* variation through the use of performance-based control charts.

Point b) recognizes that the primary operational role of a control chart is to discriminate between special and common cause variation. Common cause variation is generally outside the remit of people who work in the process. Suppose that *operational people* have established, using a performance-based control chart, that a process is in statistical control, namely, that no special causes of variation are present. Then, and only then, *technical and managerial people* can use the control chart data to compare the magnitude of the residual common cause variation present with specified limits, using capability analyses as described in clause **11**. Standardized quality capability indices for the characteristic may then be generated and any necessary improvement actions initiated.

Such process capability analysis brings out another very important aspect of the overall role of SPC. A primary role of a control chart, in an operational sense, by its very name, is to control; namely to inhibit change. The removal of special cause variation to bring a process back into control does not actually improve the process, it only returns it to its original state. This, however should not blind one to the fact that often the objective of SPC, in an overall sense, is to improve process performance by inducing change. Such betterment, through common cause reduction does not have to await special cause removal. A significant improvement in process performance is evidenced in a control chart by an "out of control" situation, as is a significant deterioration. Hence, the control chart has a built in statistical test of significance.

These features are demonstrated in Figure 42 for an underwear making-up process.

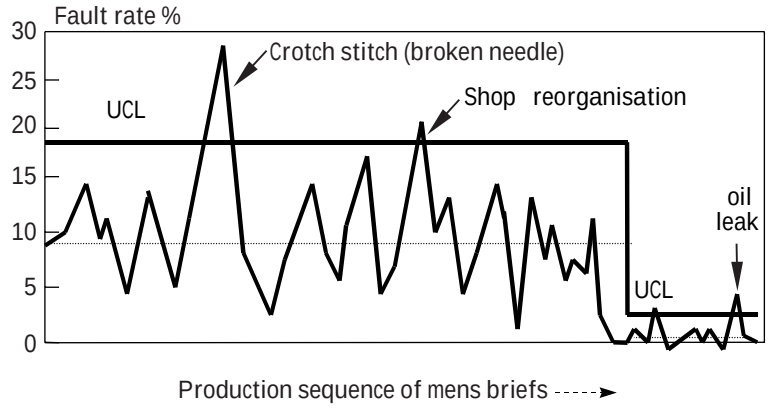


Figure 42 — Control chart for nonconforming underwear

Figure 42 shows:

— undesirable "out-of-control" situations (points above the UCLs – upper control limits) due to special causes: a broken needle, a sewing shop reorganization and oil leak;

— a desirable "out-of-control" situation (more than 7 consecutive points below the original centre line) due to a management led major training and personal development initiative which gave rise to a reduction in common cause variation from a nearly 10 % fault rate to less than 1 %.

This example brings out why it is important to differentiate between special and common cause variation. The sporadic special cause variation is due to specific assignable activities attributable to a machinist and direct support personnel. The overall performance of nearly 10 % fault rate, however, is a result of common causes endemic in the system, which is a management responsibility.

Without this perspective, using a control chart, it is usual, in such a situation, not to consider the impact of common causes on the performance of individuals.

10.3 Statistical process control or statistical product control?

At this stage it is necessary, too, to distinguish between statistical process control and statistical product control. Much, so called, statistical process control is, strictly speaking, after the event statistical product control. Figure 43 illustrates this point.

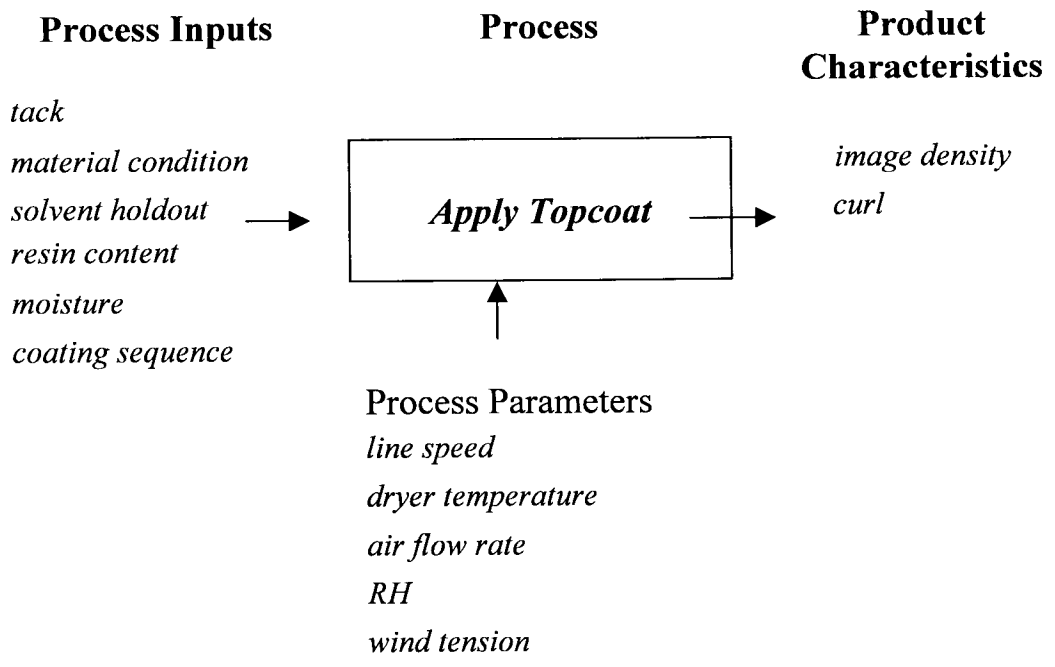


Figure 43 — Outline of process of applying a top coat to a photographic film

In such a process SPC is quite frequently applied to the product characteristics such as, image density and curl. Superficially, from the standpoint of the customer, this may appear quite acceptable. However, it is clearly not nearly as effective as control of the process parameters and process inputs that affect these product characteristics. After the event detection of unsatisfactory product may give rise to delays in shipment and increased production cost which, in turn, results in a decreased profit margin for the supplier and/or increased price to the customer.

Why then is true process control not practised more often? The primary reason is that for a large proportion of processes the technical relationship between process parameters/process inputs and product characteristics is not known. This is why the prior application of statistical experimental designs (commonly termed DoE: design of experiments) often leads to a more purposeful and effective application of SPC. This aspect is dealt with in clause **12**.

This shows that it is necessary to realize that:

a) every process generates information that can be used to control and improve its performance;

b) there is a need to develop informed perceptive observers using appropriate statistical methodology;

c) there are two sources of information and two primary statistical tools for dealing with them:

1) natural variation: use SPC, a listening tool;

2) induced variation: use DoE, a conversational tool.

10.4 Over-control, under-control and control of processes

A process monitoring system may give rise to the following situations:

a) over-control: action is taken when it should not be;

b) under-control: action is not taken when it should be;

c) control: action is taken when it should be and not taken when it should not be. A process is said to be under a state of (statistical) control when no special causes of variation are present. Variation can then be attributed purely to "common causes". Control is not a natural state but it is an achievement, arrived at by elimination, one by one, by determined effort, of special cause variation. To achieve that, it is essential to use SPC charts that set out to provide a signal when a special cause of variation is present, and to avoid giving false signals when a special cause is not present.

Sometimes "assignable cause" is taken to be synonomous with "special cause". However, a distinction should be recognized. In practice, not all special causes are assignable. A state of control does not imply that the common cause variation is large or small, within or outside of specification, but rather that it is predictable using statistical techniques.

Scenario 1: Operator reacts to each individual sample giving rise to process over-control

Suppose a particular preform extrusion process has a stable variation about the target weight value of 45 as shown in Figure 44.

The operator takes one measurement at intervals and decides, from each particular observation, whether or not to adjust.

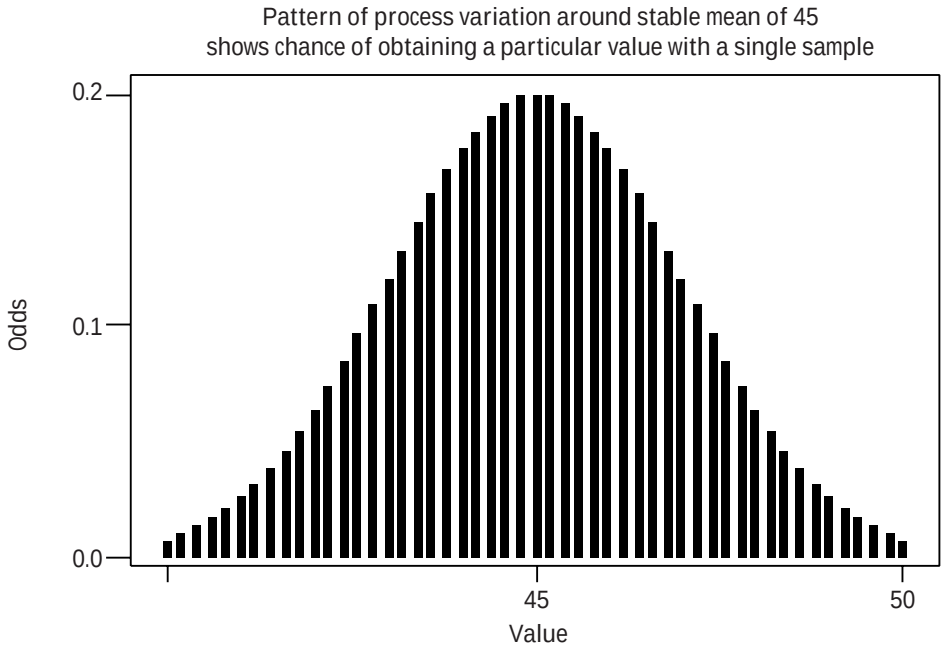


Figure 44 — Likelihood of setter/operator observing a single weight value when mean = 45

The setter/operator takes one weight measurement at 20 minute intervals and compares the result with the preferred, target or reference value of 45. Weight is controlled by adjustment of the speed feed. Adjustment is in steps of 1 so an appropriate adjustment is made if the result differs from 45 by 1 or more. Table 20 shows what may be expected in a process whose *actual* level is initially at the preferred level and which is also stable throughout with respect to variation about the various actual process levels experienced. A typical result from this monitoring plan is shown in Table 20.

Table 20 — Control plan: (take one measurement at intervals and adjust or do not adjust)

Time	Measurement value	Subsequent adjustment made	Actual process level
0800	46	−1	45
0820	42	+3	44
0840	46	−1	47
0900	48	−3	46
0920	44	+1	43
0940	43	+2	44
1000	47	−2	46
1020	44	+1	44
1040	47	−2	45
1100 etc.	44	+1	43

NOTE Measured values were obtained by taking values at random from a process with constant variation about actual process levels. This simulates a real life situation.

At 0800 the setter/operator sees 46 and increases feed speed to decrease weight thus over-controlling and bringing the actual mean weight down to 44. At 0820 the setter/operator measures 42 and decreases feed speed to increase weight by 3 units. The weight then overshoots to a mean of 47; again over-control. And so on.

The consequence of this monitoring plan is to increase overall variation[2)] from 10 units of weight (45 ± 5 in Figure 45) to 17 units of weight [$(43 - 5)$ to $(47 + 5)$] in Table 20.

This is an example of *process over-control*. Here the penalty of over adjustment is some 40 % increase in variation over the short time period considered. The general conclusion is that continual adjustment of a stable process will increase variation.

Scenario 2: Operator monitors a process using a run chart giving rise to haphazard control

Suppose a process is being monitored using a run chart as shown in Figure 45.

[2)] To the same variation criteria.

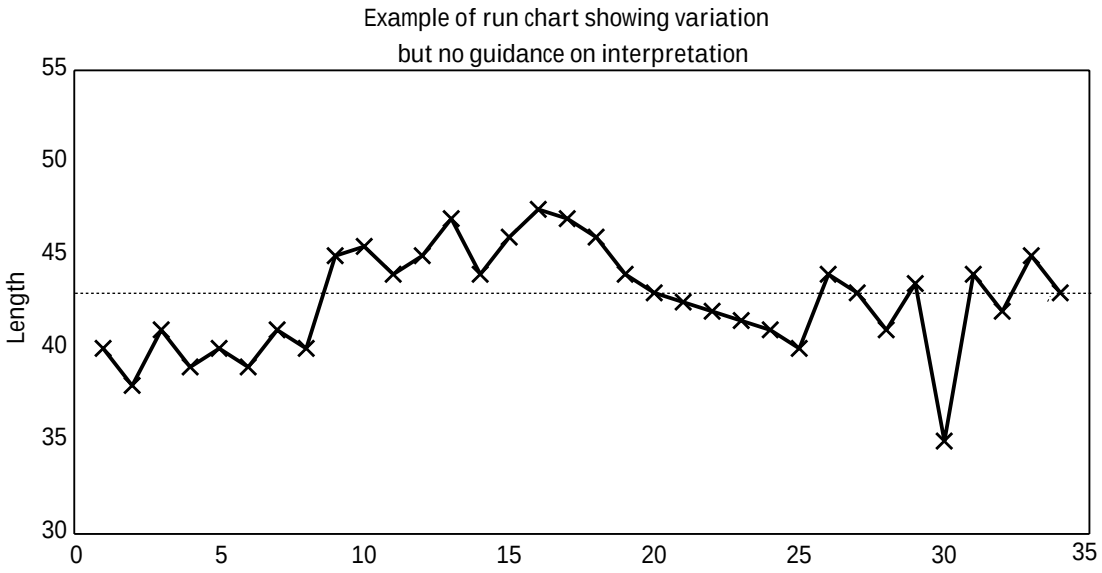


Figure 45 — Process run chart with no guidance on how to deal with variation

Whether or not reaction is made to changes in the results monitored will depend solely on the operators. Control is thus not likely to be effective. Under-control and/or over-control could thus be expected to arise. Control here is likely to be inconsistent and capricious as no guidance is given on how to interpret the variability.

Scenario 3: Monitoring using SPC chart with a potential for effective control

Here, *under-control* is the result if improper use is made of the control chart such as:

— "out-of-control" signals are not reacted to, as they arise, and a completed SPC chart is analysed purely on a retrospective basis;

— the data used for plotting do not represent process reality; for example, data is selected to make the process "look good".

Figure 46 shows an example of the use of an SPC chart with the data of Figure 45. Four "out-of-control" situations are flagged on the chart. If these are reacted to positively at the time they are signalled, then the process will be effectively controlled.

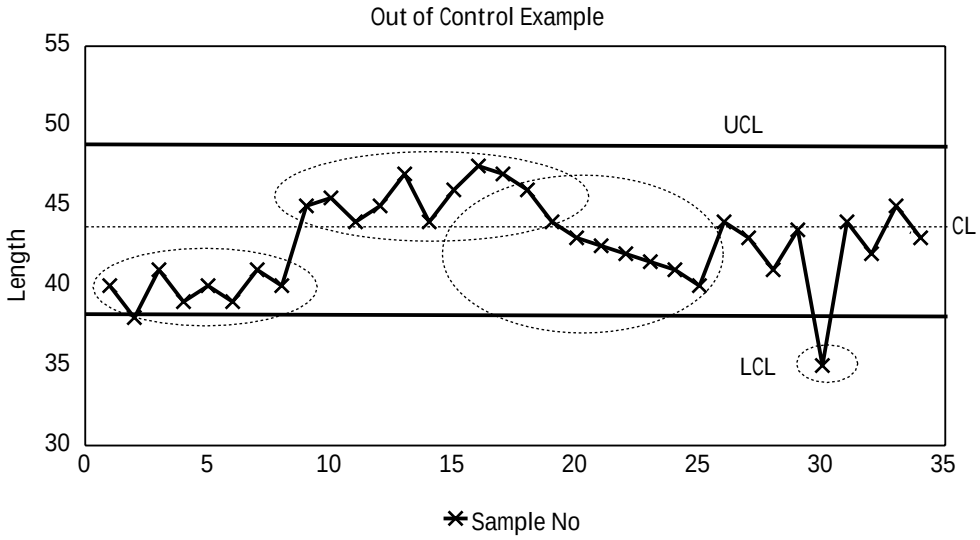


Figure 46 — Process control chart with criteria for "out-of-control" signals

Typical criteria for "out-of-control" include the following:

1) any point outside of the control limits (upper and lower: UCL and LCL);
2) any run of 7 consecutive points above or below the centre-line (CL);
3) any run of 7 consecutive points up or down;
4) any obvious non-random patterns (based on technical and operational knowledge of the process).

10.5 Key statistical steps in establishing a standard performance based control chart

10.5.1 *General*

Having identified the process parameter or product characteristic to be observed, it is first necessary to decide on a monitoring strategy, (such as how to constitute a sample or subgroup, how many observations to take, and how frequently) followed by the setting up and interpreting of the control chart (such as how to set control limits and establish out-of-control criteria). These are now discussed.

10.5.2 *Monitoring strategy*

10.5.2.1 *Subgroup constitution*

In constituting a subgroup, a number of factors need consideration.

The concept of subgrouping is that the variation *within* a subgroup is made up only of common causes, with all special causes of variation occurring *between* subgroups. As the primary role of a control chart is to distinguish between common and special cause variation, the choice of rational subgroup has a considerable bearing on the usefulness of a control chart for a given purpose. For instance, if a subgroup is made up of, say, the diameter of 3 consecutive parts on a high precision honing operation, the common cause variation within the subgroup may be miniscule. However, if the subgroup is made up of 3 parts, each of which is selected from consecutive wheel dressings, the common cause variation will be much greater. There will be far less homogeneity in the subgroup. This will have considerable impact on control limits. Hence the constitution of a subgroup will depend on the primary purpose of the control chart and a thorough knowledge of the process.

Frequently the term rational subgroup is used. This highlights the need for further care in subgroup constitution. Consider a multi-headed machine which is to be sampled at the rate of 1 per 15 minutes to make up a subgroup of 4. It would not be rational to take 1 measurement on head 1 at 0800, one from head 2 at 0815, one from head 1 at 0830 and one from head 3 at 0845 as it would be difficult to separate out within-head, between-heads and between-times variation.

Summarizing, the basic mean ($\overline{X}$) and range (R) control chart can be looked upon as a two-factor nested experimental design which separates out within-subgroup (common cause) variation from between-subgroup (special cause) variation. This is shown diagrammatically in Figure 47.

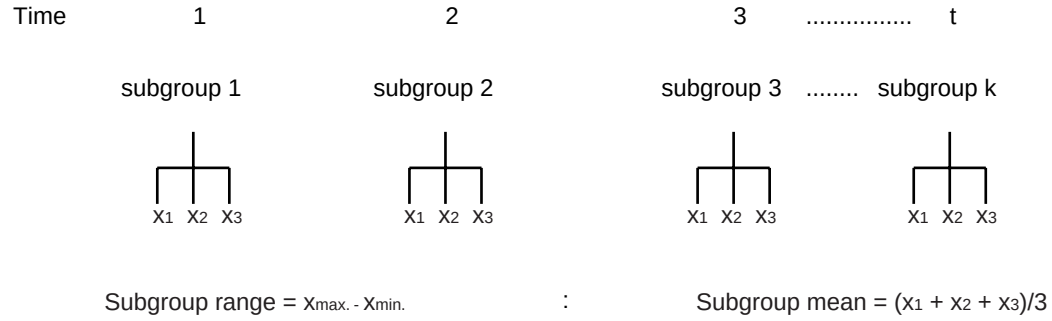


Figure 47 — A two factor nested design is the basis of an $\overline{X}$.R chart (illustrated with a subgroup size of 3)

The mean and range for each subgroup are calculated. These are then plotted in time sequence. The R chart evaluates the variation *within* a subgroup. The $\overline{X}$ chart assesses the variation between subgroups.

It is often said that the measurements in a subgroup should be independent of one another. However, in practice, this is frequently not achieved in a real life process. A measurement in a subgroup is often influenced by another to some degree. Hence data for control charts often exhibit serial correlation (autocorrelation). What impact does this have? A consensus view is that:

a) for most situations "significant" autocorrelation will have minimal impact upon control chart limits;

b) whilst severe autocorrelation may contaminate the control limits, the control chart may usually be safely interpreted at face value.

This indicates that one need not be overly concerned about the effects of autocorrelation on control chart interpretation in most situations. Hence, recourse to complex techniques, such as the use of variograms and correlograms, to distinguish between random, cyclic, trend and correlated variations, as expounded in ISO 11648 is usually not required.

10.5.2.2 *Subgroup size*

Sometimes the sample or *subgroup size* may be dictated by circumstances. If the measurement or test is destructive or expensive, or on a process parameter, such as curing time or flow rate, the subgroup size may be necessarily small, for example, $n = 1$ or $n = 2$. However, larger subgroups have certain technical advantages:

a) even if the individual measurements are not normally distributed, the distribution of the mean of the subgroups tends to normality as the subgroup size increases (central limit theorem). A sample size of 5 is generally adequate to achieve this;

b) the larger the subgroup size, the greater the ability of the control chart to detect changes in the mean.

This is depicted graphically in Figure 48.

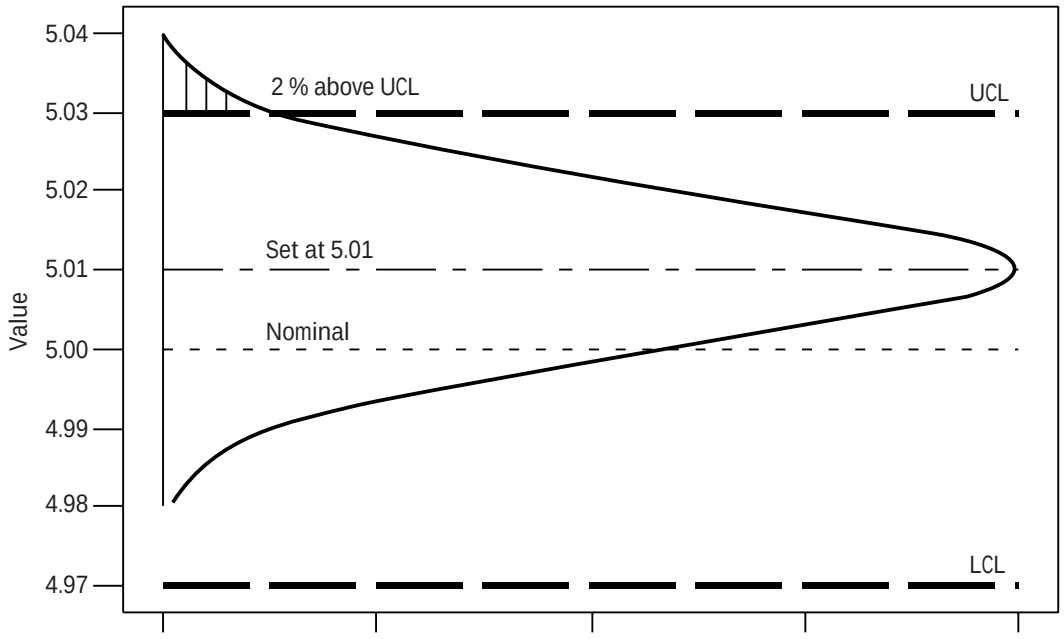


Figure 48a) — Control limits for individuals (n = 1) (setting = +0.01 on nominal)

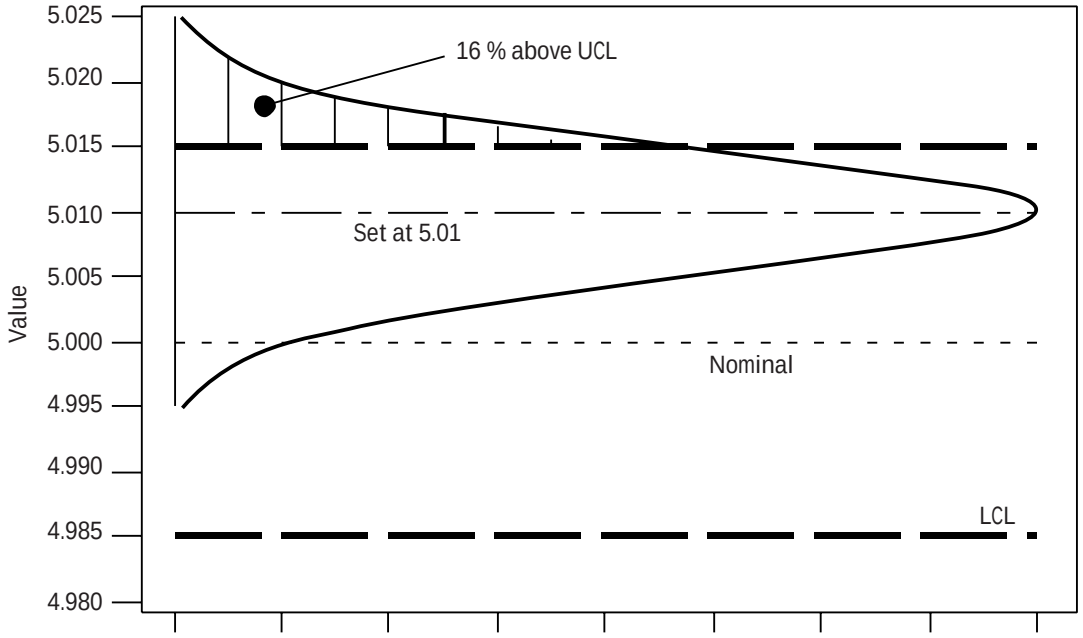


Figure 48b) — Control limits for averages (n = 4) (setting = +0.01 on nominal)

Figure 48 — Effect of subgroup size on ability to detect changes in process mean (process nominal = 5.00, process standard deviation = 0.01)

Figure 48a) shows that with a subgroup size of n = 1, a shift in mean of 0.01 will only be expected to be detected some 2 % of the time if the control limits are the only criteria for control. By contrast Figure 48b) shows that if the subgroup size is increased to n = 4, the same shift in mean is expected to be detected almost 16 % of the time.

This is because the variation of means is less than the variation of individuals according to the following relationship:

$$\text{Standard deviation of mean} = \frac{\text{Standard deviation of individual}}{\sqrt{\text{Subgroup size}}}$$

10.5.2.3 *Frequency of sampling*

The frequency of sampling is a compromise between sampling cost and value of the timely detection of process changes. A useful guide is to consider taking about six subgroups between anticipated changes in a process.

10.5.3 ***Construction of a standard control chart***

10.5.3.1 *Common features*

The generic control chart for both measured data and attributes (count and classified data) shares similar features. Typically it consists essentially of five lines and a series of plotted points:

a) a vertical scale of values of a chosen statistic, X, say, (e.g. mean, range, standard deviation, number of nonconformities) of the subject characteristic;

b) a horizontal scale depicting subgroup sequence numbers;

c) a centre-line (CL) , where CL = mean of $X = \overline{X}$;

d) an upper control limit: $UCL_s = \overline{X} + 3.s_s$ (where s_s = standard deviation of the statistic plotted);

e) a lower control limit: $LCL_s = \overline{X} - 3.s_s$;

f) plotted points representing the calculated values of the statistic, X, of rational subgroups sequentially formed from measurements of the chosen characteristic.

Standard formulae and tabled constants are available for the calculation of standard limits. These are given in annex A.

10.5.3.2 *Example of typical mean and range control chart for measured data*

Unlike attribute charts which are formed from a single statistic (see Figure 42), standard control charts for measured data are made up of two statistics; the mean or median, to monitor changes in the level of a characteristic between subgroups; and the standard deviation or range, to monitor variability within a subgroup.

An example of a mean and range ($\overline{X}$ and R) control chart for measured data is shown in Figure 49 for the weights of standard specimens of fabric given in Table 2.

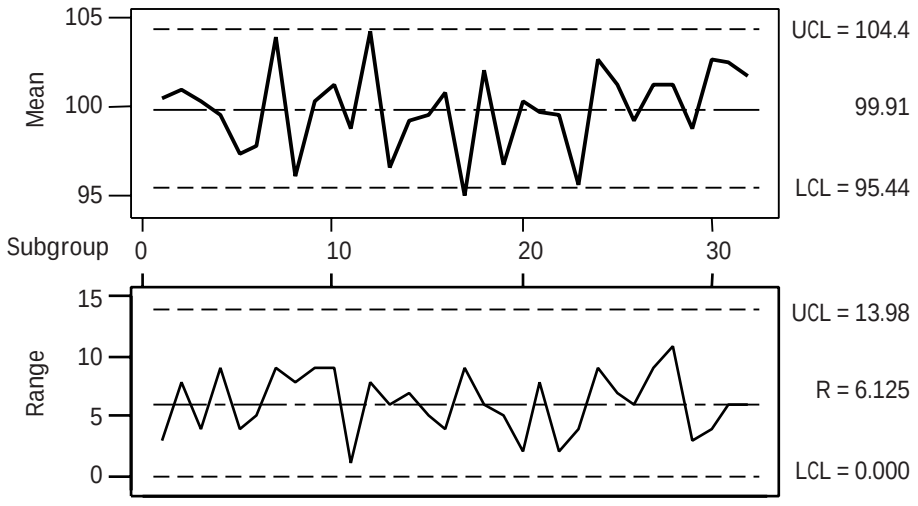


Figure 49 — Mean and range chart for weights of standard specimens of fabric

The calculations required for such a chart are as follows:

— first plotting point of Xbar chart is given by: $\overline{X}_1 = \frac{(101 + 99 + 100 + 102)}{4} = 100.5$;

— first plotting point of R chart is given by: $R_1 = 102 - 99 = 3$;

— average range = $R_{bar} = \overline{R} = \frac{(3 + 8 + 4 + ...)}{32} = 6.112$;

— average of $X_{bars} = \overline{\overline{X}} = \frac{(100.5 + 101 + 100.25 + ...)}{32} = 99.92$;

— UCLrange = $D_4 \times \overline{R} = 2.282 \times 6.112 = 13.95$ (where D_4 = constant for $n = 4$, see annex A);

— LCLrange = $D_3 \times \overline{R} = 0$;

— UCLmean = $\overline{\overline{X}} + (A_2 \times \overline{R}) = 99.92 + (0.729 \times 6.112) = 104.4$ (where A_2 = constant for $n = 4$, see annex A);

— LCLmean = $\overline{\overline{X}} - (A_2 \times \overline{R}) = 99.92 - (0.729 \times 6.112) = 95.5$.

10.5.3.3 *Rationale for control limits*

Traditionally there are two distinct approaches to the setting of control limits for performance-based control charts. One approach is due to Walter Shewhart, who chose control limits formed by adding to and subtracting from the "expected" value, 3 (2 for warning limits) times the standard deviation. This was based on his experience that this was an "acceptable economic value". At about the same time control limits were formed by adding to and subtracting from the "expected" value, 3.09 (1.96 for warning limits) times the standard deviation. The reason for this difference is brought out in Table 21. One approach focused on rounded odds and the other on rounded multiples of standard deviations.

Table 21 — The two traditional systems for calculating control limits

Control limits (action)		Warning limits (sometimes used)	
Equation[a]	Probability of being above or below[b]	Equation[a]	Probability of being above or below[b]
mean ± 3 standard deviations	0.135 %	mean ± 2 standard deviations	2.28 %
mean ± 3.09 standard deviations	0.1 % (1/1 000)	mean ± 1.96 standard deviations	2.5 % (1/40)

[a] The mean and standard deviation used for control limits are derived from prior process knowledge or a trial run of sufficient duration for the major sources of variation to manifest themselves. As the control chart is a model, or exemplar, of common cause variation, data arising due to special cause variation should not be used in the calculation of control limits. Once calculated there is no logic in routinely recalculating centre-lines and control limits as is sometimes common practice. Recalculation is only required when there has been a significant change in nominal value or common cause variation.

[b] Probabilities were obtained from the standard normal distribution (see Table 7).

Table 21 shows that the difference is trivial. The International Standards Organization has adopted the Shewhart system as the world standard (ISO 8258:1991).

From a rational viewpoint the use of ±3 standard deviations for action control limits can be argued, for a normal distribution, as striking a reasonable balance between:

— looking for trouble when it does not exist; and

— not looking for trouble when it does exist.

A normal distribution can usually be expected for a means chart based on subgroups of 5 or more, even if the distributions of individual values is non-normal. However, for smaller subgroup sizes in a means chart, and also for charts of individuals, ranges, standard deviations and attributes, the distribution of the plotted statistic can be decidedly non-normal. In such cases, limits based on probabilities for the representative distribution (e.g. skew) are sometimes used.

10.6 Interpretation of standard Shewhart type control charts

The lines in a control chart reflect common cause variation of the statistic. If the plotted points do not adhere to that model, the presence of special causes is indicated. To test for an out-of-control situation, certain guidelines are provided. Typical such guidelines for a normal distribution of the statistic plotted are shown in Table 22.

Table 22 — Probabilities associated with different decision criteria

Rule	Description	Probability	Odds
1	Point outside of upper or lower control limit (action)	0.001 35	1/741
2	Seven consecutive points above or below the centre-line	0.007 81	1/128
3	Seven consecutive points increasing or decreasing (including the first and last)	0.000 20	1/5 040
4	Any obvious non-random variation[a]		

[a] Based on process technical or operational knowledge rather than statistical probability

NOTE By way of illustration take Rule 1. If a value exceeds the upper control limit, say, either the process is:

— in control; in which case one has witnessed an extremely unusual phenomena, an event that has a very remote chance of occurrence, namely, 1 chance in 740; or it is

— out-of control because of the presence of a special cause which needs investigation with a view to elimination.

Using rules 1, 2 and 3 for the example shown in Figure 49, an out-of-control situation is indicated for the mean at subgroup 17. The cause should be sought and eliminated as soon as such an indication is seen on the control chart. The control limits should be recalculated, after discarding the data from this subgroup, to form the basis for ongoing control.

10.7 Selection of an appropriate control chart for a particular use

10.7.1 *Overview*

There are many classes and types of Shewhart type control charts available. In addition the cumulative sum (CUSUM) chart is becoming more widely recognized as a very useful diagnostic and control tool.

Standard Shewhart type SPC control charts normally require a different chart for each process parameter or product characteristic. This was recognized by Bothe [39] who has developed universal charts to apply to situations where continuity of charting is required in short run situations and across parameters and characteristics having different nominal or target and/or average range values.

The cumulative sum chart is, in many ways, superior to conventional Shewhart methods. It is appropriate for examining all forms of numerical data relative to a reference value, on a retrospective or current basis. It has three main uses: control, diagnosis and prediction.

10.7.2 *Shewhart type control charts*

Principal kinds of Shewhart type measured data charts are shown in Table 23a) and attribute data charts in Table 23b).

Table 23a) — Principal Shewhart type measured data control charts

	Subgroup size (n)					
	$n = 1$		$n < 10$		$n > 1$	
Chart	X chart	MR chart	$\overline{X}$ chart	R chart	$\overline{X}$ chart	s chart
Plot point	X	*moving R*	$\overline{X}$	R	$\overline{X}$	s
Centre-line	$\overline{X}$	$\overline{R}$	$\overline{\overline{X}}$	$\overline{R}$	$\overline{\overline{X}}$	$\overline{s}$
UCL	$\overline{X} + E_2.\overline{R}$	$D_4\overline{R}$	$\overline{\overline{X}} + A_2.\overline{R}$	$D_4\overline{R}$	$\overline{\overline{X}} + A_3.\overline{s}$	$B_4.\overline{s}$
LCL	$\overline{X} - E_2.\overline{R}$	$D_3\overline{R}$	$\overline{\overline{X}} - A_2.\overline{R}$	$D_3\overline{R}$	$\overline{\overline{X}} + A_3.\overline{s}$	$B_3.\overline{s}$

NOTE 1 The *standard individual and moving range (X and MR)* chart would be suitable in those situations where it is only practicable, or desirable, to take a single measurement at a time. Examples are process parameters such as temperature or pressure and where destructive testing is involved. Moving ranges are constructed from progressive sets of individuals, for example, of size two or greater. The constants, E_2, D_3 and D_4, are based on the size of the set which constitutes the range. An alternative is to use a moving average and moving range chart. Prior to calculating the control limits the resulting distribution should be checked for normality. The lesser sensitivity of the individuals chart, compared with the average chart, for detecting shifts in the level of the process should be noted (see Figure 48). An example of a standard individuals chart is shown for % silicon in Figure 53.

NOTE 2 The *standard average and range* chart is recommended for its simplicity, where manual charting is concerned, for subgroup sizes up to about 10. However, it should be borne in mind that the range is based only on the two extreme values in a subgroup and its efficiency falls off, in comparison with the standard deviation, as the subgroup size increases. An example of a standard average and range chart is shown for fabric weight specimens in Figure 49.

NOTE 3 The *standard average and standard deviation* chart can be used instead of the average and range for all subgroup sizes greater than 1.

The A, B and D constants depend on the subgroup size. They are tabulated in annex A. In the case of the moving range chart, the equivalent subgroup size is the number of individuals making up each successively plotted range.

Table 23b) — Principal Shewhart type attribute data control charts

	Events: nonconformities		Nonconforming units	
Chart	Constant sample size: "c" chart	Variable sample size: "u" chart	Constant sample size: "np" chart	Constant sample size: "p" chart
Plot point	c	u	$n.p$	p
Centre-line	$\overline{c}$	$\overline{u}$	$n.\overline{p}$	$\overline{p}$
UCL	$\overline{c} + 3\sqrt{\overline{c}}$	$\overline{u} + 3\sqrt{\frac{\overline{u}}{n}}$	$n.\overline{p} + 3\sqrt{n.\overline{p}(1 - \overline{p})}$	$\overline{p} + 3\sqrt{\frac{\overline{p}(1 - \overline{p})}{n}}$
LCL	$\overline{c} - 3\sqrt{\overline{c}}$	$\overline{u} - 3\sqrt{\frac{\overline{u}}{n}}$	$n.\overline{p} - 3\sqrt{n.\overline{p}(1 - \overline{p})}$	$\overline{p} - 3\sqrt{\frac{\overline{p}(1 - \overline{p})}{n}}$

NOTE 1 There are four types of *standard* attribute charts:

— c: number of incidences, events or nonconformities in a sample which is of constant size;

— u: number of incidences, events or nonconformities per unit in a sample which is of variable size;

— n.p: number of nonconforming units in a sample which is of constant size;

— p: proportion of nonconforming units in a sample which is of variable size.

The choice of which to use depends on whether the sample size (n) is constant or variable and whether incidences/nonconformities or nonconforming units are involved. It is advisable, from the point of view of simplicity, to keep sample sizes constant if possible. Incidences/nonconformities charts frequently provide more technical information than nonconforming units ones; however, certain logistics information may be lost. For example, if one has 14 nonconformities in a sample of 50 units it would not be known, from the chart, how many units were affected. On the other hand, if 8 units were involved, some with multiple nonconformities, diagnostic information would be lost on some nonconformities.

NOTE 2 The measured data chart is preferred whenever possible. An example would be a diameter, say, which could be checked with either a go: no-go gauge or measured with a micrometer. Another illustration is where subjective judgements made on a particular characteristic, are converted into a rating scale of, say 1 to 10. This permits the use of a measured data chart rather than an attribute one. An example is a scale of 1 to 5 for degree of fabric pilling.

NOTE 3 This preference is for two principal reasons: one, the measured data chart provides more information, and two, in the quality field the attribute chart often requires nonconformities or nonconforming units to happen before plotting can take place.

NOTE 4 Having made the decision of which type of attribute chart to use, a second choice is either to use a *single characteristic* chart (see Figure 42) or a *multiple characteristic* chart. The multiple characteristic chart facilitates prioritizing the sources of variation and diagnosis with a view to improving process capability.

NOTE 5 The *capability* of a standard attribute chart is given by the overall average (centre-line) value; nearly 10 % initially in Figure 42 and ultimately under 1 % after the improvement initiative. When plotting nonconformities or nonconforming units, the ultimate or preferred average value is zero.

10.7.3 *Cumulative sum (cusum) charts*

10.7.3.1 *Principal features of cusum charts*

A cusum is essentially a running summation of deviations from some pre-selected reference value. The mean of any group of consecutive values is represented visually by the current slope. Its principal features are:

a) its greater sensitivity than the Shewhart type chart in detecting small changes in the mean;

b) any changes in the mean, and the extent of the change, are indicated visually by a change in slope of the graph:

— horizontal graph – on target or reference value;

— downwards slope – mean less than the reference or target value: the greater the slope the bigger the difference;

— upwards slope – mean more than the reference or target value: the greater the slope the bigger the difference;

c) it can be used retrospectively for investigative purposes, on a running basis for process control, and for prediction of process performance in the immediate future.

10.7.3.2 *Construction of cusum charts*

The steps in setting up a cusum control chart are simply as follows.

Step 1: Choose a *reference*, target, control or preferred *value*, RV. The average of past results will give the best statistical discrimination.

Step 2: Tabulate the results in a meaningful (e.g. chronological) sequence. These results may, for example, be individual values or the average of subgroups. Subtract the reference value from each such result.

Step 3: Progressively sum the values obtained in step 2. These sums are then plotted as a cusum chart.

Step 4: To obtain the best visual effect set up a horizontal scale no wider than about 2.5 mm between plotting points.

Step 5: For reasonable discrimination without undue sensitivity:

a) choose a vertical scale (relative to horizontal) = (short term spread of results)/3. Round off as appropriate;

b) alternatively, where it is required to detect a known change, say C, choose a vertical scale such that:

$$C < \frac{\text{length of one unit on horizontal scale}}{\text{length of one unit on vertical scale}} < 2C.$$

Round off as appropriate.

10.7.3.3 *Application: fractional horse-power motor voltage*

Voltages are taken on fractional horse-power motors at an early stage of production for process control processes. 40 such results taken in chronological order are shown in Table 24. The reference, or nominal, value is 10 volts. A standard Shewhart type control chart for the data is shown in Figure 50a). As this chart was not very revealing, a cusum chart was plotted for the same data. This is shown, for comparison, in Figure 50b).

Table 24 — Voltages for fractional horse power motors

Voltage	Voltage − 10	Cusum (sum of column 2)
9	−1	−1
16	+6	+5
11	+1	+6
12	+2	+8
16	+6	+14
7	−3	+11
13	+3	+14
12	+2	+16
13	+3	+19
11	+1	+20
12	+2	+22
8	−2	+20
8	−2	+18
11	+1	+19
14	+4	+23
8	−2	+21
6	−4	+17
14	+4	+21
4	−6	+15
13	+3	+18
3	−7	+11
9	−1	+10
7	−3	+7
14	+4	+11
2	−8	+3
6	−4	−1
4	−6	−7
12	+2	−5
8	−2	−7
8	−2	−9
12	+2	−7
6	−4	−11
14	+4	−7
13	+3	−4
12	+2	−2
14	+4	+2
13	+3	+5
10	0	+5
13	+3	+8
13	+3	+11

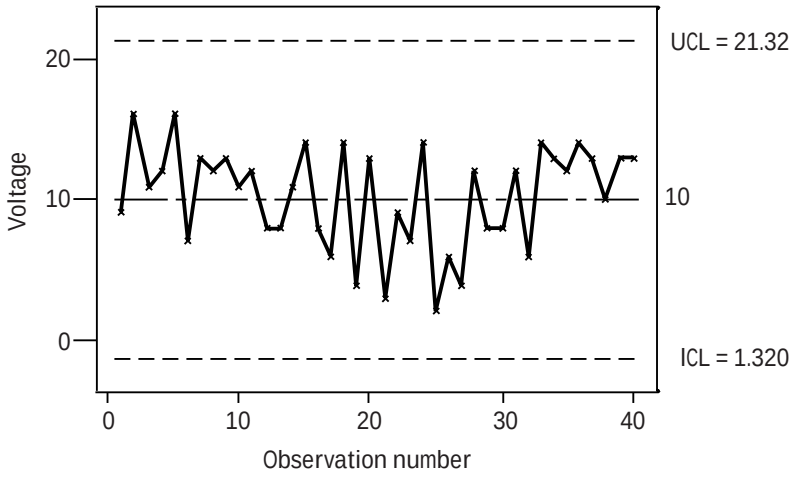


Figure 50a) — Standard Shewhart individuals control chart for fhp motor voltage

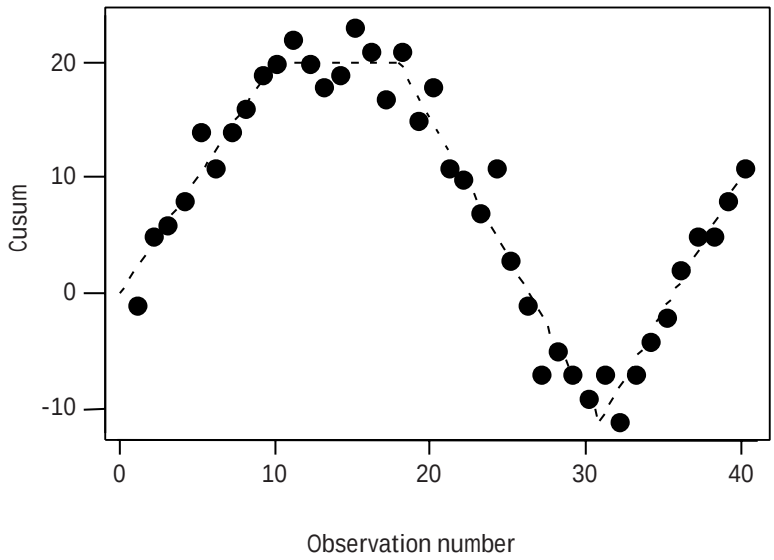


Figure 50b) — Cusum chart for the same fhp motor voltage data

Whereas it is not intuitively obvious from the standard Shewhart type control chart where any significant changes in voltage occurred, the cusum chart clearly indicates three changes in process level, at the 10^{th}, 18^{th} and 31^{st} motor.

The cusum chart shows:

i) a constant process level up to motor number 10 at a level higher than 10. The estimated level is given by the slope thus:

$$\text{Reference value} + \frac{\text{Cusum value at end of line} - \text{cusum value at start of line}}{\text{Number of observation intervals}}$$

therefore:

$$\text{Initial average value} = 10 + \frac{20 - 0}{10} = 12 \text{ volts.}$$

Similarly:

ii) a constant process level from motor 11 to 18 at the reference value of 10 volts;

iii) a constant process level from motor 19 to 31 at about 7.6 volts;

iv) a constant process level from motor 32 to 40 at about 12.4 volts.

This information may now be used to pinpoint root causes of the deviation from the reference value of 10 volts.

A new chart, based on these cusum results, may also be drawn. This takes all the noise out of the original plot shown in Figure 50a). This is termed a Manahattan chart. It is shown in Figure 50c).

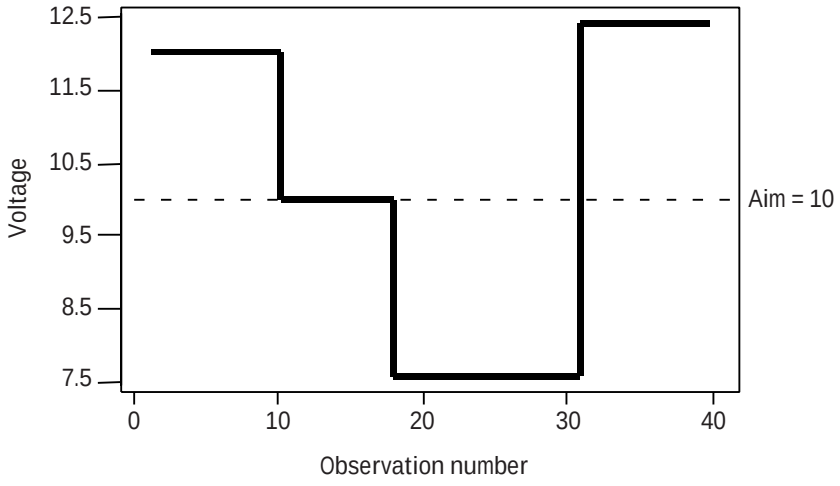


Figure 50c) — Manhattan chart for fhp motor voltages

This application shows the use of a cusum chart in investigative mode.

Cusum charts may also be used in control and predictive mode. Details are given in BS 5703.

11 Process capability

11.1 Overview

In clause **10** the performance based control chart was discussed purely in terms of process control. No regard was made to the acceptability, or otherwise, of the characteristic in respect to an imposed standard of performance. A further important technical role of the performance based control chart is the provision of the basis for the assessment of process capability against the requirements of a specification and for the formulation of standardized benchmarks of performance. The four states of any process are shown in Table 25.

Table 25 — The four possible states of any process

The four states		Control (stability)	
		not ok	ok
Capability (performance)	not ok	eliminate special causes reduce common causes	reduce common causes
	ok	eliminate special causes	ideal situation: monitor at low level

The performance based control chart provides answers to the following three significant business questions.

— Question 1: *Is the process in control?* This is directed, primarily, at operational people working in the process. If not in control, there is a need to seek out and eliminate detrimental special cause variation.

— Question 2: *What is the process capability in relation to the specified requirement or customer expectation?* This is directed, primarily, at technical people responsible for the process. If this is not at an acceptable level, there is a need to make fundamental changes to the process to reduce common cause variation, to use a more capable process or to relax the specification.

— Question 3: *Is there evidence of improvement?* This will be signalled as follows:

— in a measured data chart, by an "out-of-control" movement in the mean towards the preferred value and/or an "out-of-control" reduction in the within-subgroup variation indicated by the range or standard deviation chart;

— in an attribute chart by an "out-of-control" change in the mean towards the preferred value: seven consecutive points below the centre-line (Rule 2) in the case of the plot of nonconformities in Figure 42.

This is directed, primarily, at management who, in a "best practice" organization, are responsible for the "continual improvement" of processes.

11.2 Process performance v process capability

ISO 3534-2:2000 distinguishes between process performance and process capability as follows:

a) process performance and its related P_p ($Performance_{process}$) indices relate to the statistical estimate of the outcome of a characteristic from a process which may *not* have been demonstrated to be in a state of statistical control in relation to that characteristic; whereas

b) process capability and its related C_p ($Capability_{process}$) indices have an identical definition, with the exception that here the process has been demonstrated to be *in* a state of statistical control.

Arising from these international definitions, process performance measures are preliminary indicators confined to early development activities in developing the potential of new processes or more mature processes which are not in a state of statistical control. They are thus unrelated to the quality of product, process or service offered to the ultimate customer. Concentration here is therefore focused on process capability measures that are based on prior demonstration of process stability. Having stated that the calculations associated with process capability are identical to those of process performance, the significant difference is the stability of the data used in the calculations and the reliability of subsequent predictions.

11.3 Process capability for measured data

11.3.1 *General*

Process capability is calculated for a particular process parameter or product characteristic *only* after process stability has been confirmed *and* the distribution pattern of individual values has been determined. The confirmation of process stability is first established from a control chart. Only then may reliable predictions be made about and the distribution pattern determined by recourse to tally charts, histograms, probability papers or computer-based distribution techniques.

11.3.2 *Estimation of process capability (normally distributed data)*

For normally distributed data from a stable process, an estimate of the capability of a particular characteristic is given by the equation:

$\overline{\overline{X}} \pm z.s$

where

$\overline{\overline{X}}$ is the overall mean;

z is the chosen constant, often = 4;

s is the estimated standard deviation of individuals.

Example

If data taken from a stable process exhibits a normal pattern of variation and $\overline{\overline{X}} = 10.1$, $z = 4$ and $s = 0.01$, then the estimated capability is quoted as:

10.01 ± 0.04

within which (from Table 7) nearly 99.994 % of values are predicted to lie (as 2×32 parts per million are expected outside of this range of values).

If, on the other hand, the capability standard is less stringent and z is taken to be 3, then the estimated capability is quoted as:

10.01 ± 0.03

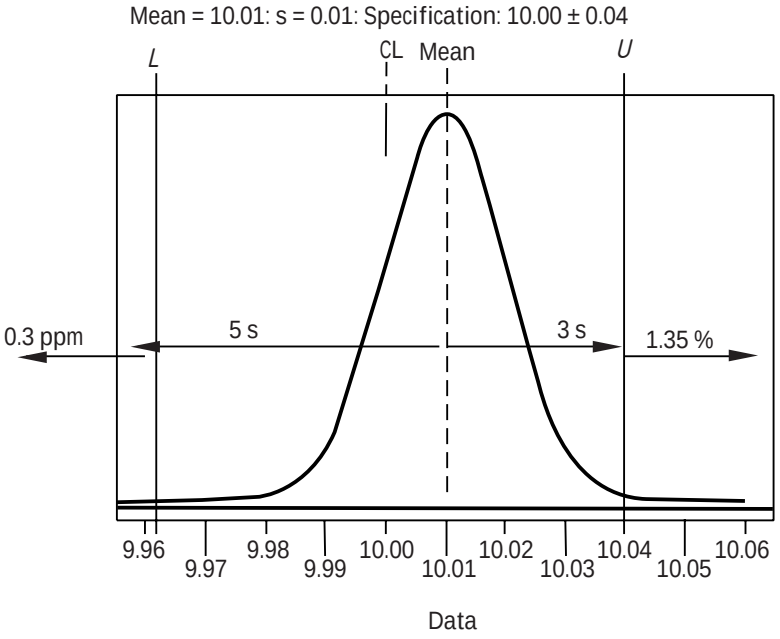
within which 99.73 % of values are expected to lie.

Capability can then be referenced to any imposed specification limits and the proportion expected outside of these limits estimated using Table 7. For instance, if the specification is 10.00 ± 0.04:

— enter Table 7 at $z = \dfrac{(U - \overline{\overline{X}})}{s} = (10.04 - 10.01)/0.01 = 3$, to give 0.135 % above the upper specification limit;

— enter Table 7 at $z = \dfrac{(\overline{\overline{X}} - L)}{s} = (10.01 - 9.96)/0.01 = 5$, to give 0.3 ppm below the lower specification limit.

The pictorial expression of this is shown in Figure 51.



NOTE CL = centre-line = mid-distance between upper specification limit (U) and lower specification limit (L)

Figure 51 — Graphical comparison of process capability with specified tolerance

It can be seen that the choice of z does not affect the proportion of values predicted to lie outside of the specified tolerance. However if $z = 3$, rather than $z = 4$ is mandated as a minimum standard, the implication is that, provided the capability expressed by 10.01 ± 0.03 lies within the specified tolerance band, the process capability is acceptable. For $z = 4$ this becomes 10.01 ± 0.04. In other words with $z = 3$, up to 0.135 % is tolerated outside each specification limit as opposed to 32 parts-per-million with $z = 4$. Thence in the example:

— if $z = 3$ is the minimum reference standard, the process is deemed *capable*;

— if $z = 4$ is the minimum reference standard, the process is deemed *incapable*. It can be made capable by either reducing the standard deviation from 0.01 to 0.007 5, by adjusting the mean from 10.01 to 10.00 or by changing the specification.

11.3.3 ***Estimation of process capability (non-normally distributed data)***

From the central limit theorem it is known that averages of subgroups tend to normality as the subgroup size increases. However, many processes quite naturally produce patterns of variation for individuals that are non-normal. For example, dimensions with a natural zero such as eccentricity, parallelism and taper are likely to be skewed. So are such things as times to pay, arrival times and length of time to resolve a query. It is essential that any statistical statement of capability be based upon the pattern of variation exhibited by the process.

Statistical expressions for capability of non-normal distributions are best expressed in probability terms rather than in terms of standard deviations.

A typical expression of capability for a skew distribution, equivalent to the ±3 standard deviations for the normal distribution, would be:

$$\text{mean} \begin{matrix} + \text{ range of values between the mean and upper 0.135 distribution percentile} \\ - \text{ range of values between the mean and lower 0.135 distribution percentile} \end{matrix}$$

This is indicated graphically in Figure 52 and put to use in “Solution 1: Process capability” later in this sub-clause.

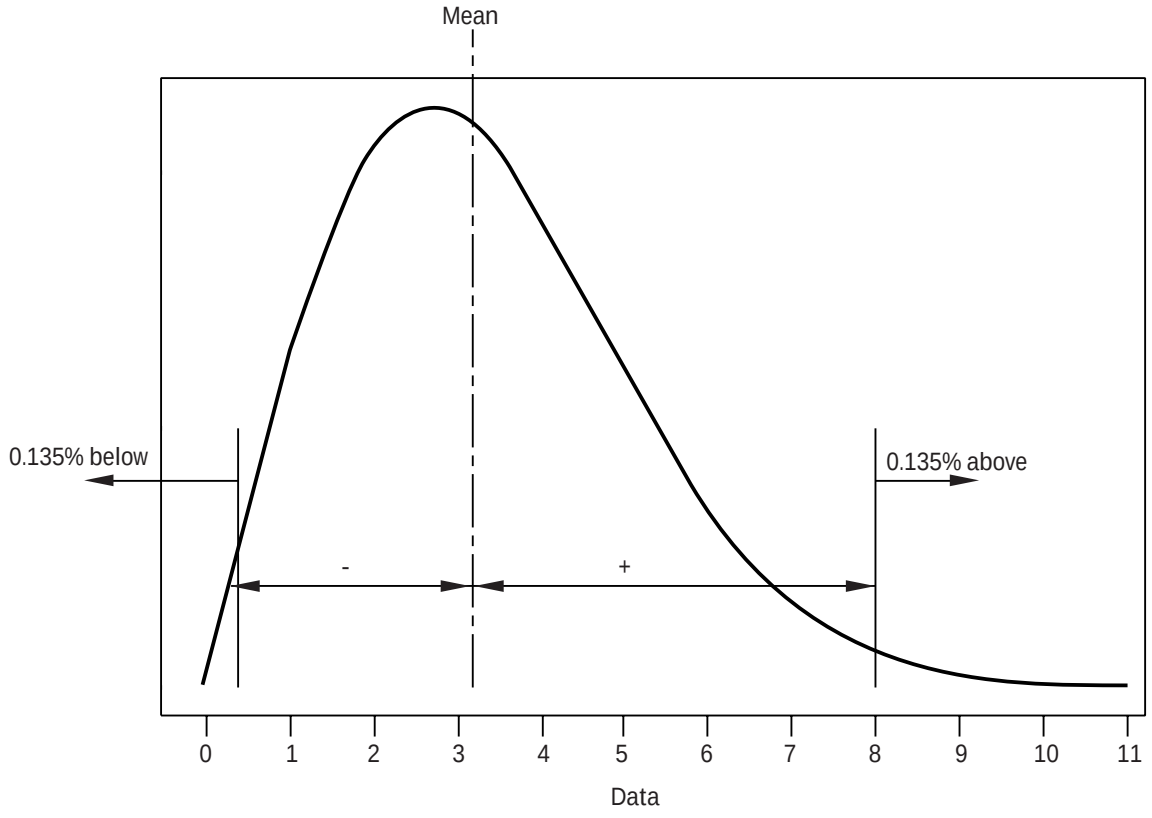


Figure 52 — Illustration of the estimation of capability with a skew distribution (equivalent to a range of $\pm 3\sigma$ in a normal distribution)

A simple graphical procedure is the use of an appropriate probability paper. An example is shown in Figure 55. Alternatively, good SPC computer programs use distribution fitting routines.

Example of assessing process capability of a skew distribution using probability paper

The data of Table 26, relating to the measurement of silicon through the taphole of a blast furnace, is used to illustrate the probability paper method.

The taphole of a blast furnace is opened at 3 hour intervals and the % silicon is measured and recorded. 90 of these values, taken in sequence, are shown in the table. (Data reference: ISO 11648.)

Table 26 — % silicon values taken in sequence from a blast furnace

0.13	0.15	0.19	0.22	0.20	0.20	0.18	0.26	0.40
0.10	0.22	0.29	0.18	0.13	0.16	0.28	0.34	0.20
0.19	0.21	0.28	0.25	0.15	0.22	0.23	0.32	0.20
0.22	0.24	0.21	0.19	0.12	0.31	0.24	0.30	0.42
0.45	0.19	0.29	0.22	0.21	0.18	0.18	0.31	0.31
0.25	0.22	0.15	0.17	0.22	0.16	0.22	0.31	0.36
0.13	0.15	0.32	0.15	0.23	0.14	0.31	0.27	0.27
0.14	0.17	0.20	0.18	0.34	0.16	0.25	0.12	0.36
0.25	0.22	0.30	0.15	0.32	0.19	0.31	0.24	0.27
0.23	0.25	0.19	0.11	0.18	0.34	0.45	0.40	0.21

NOTE Sequence of readings: first read downwards in column 1, then downwards in column 2, etc.

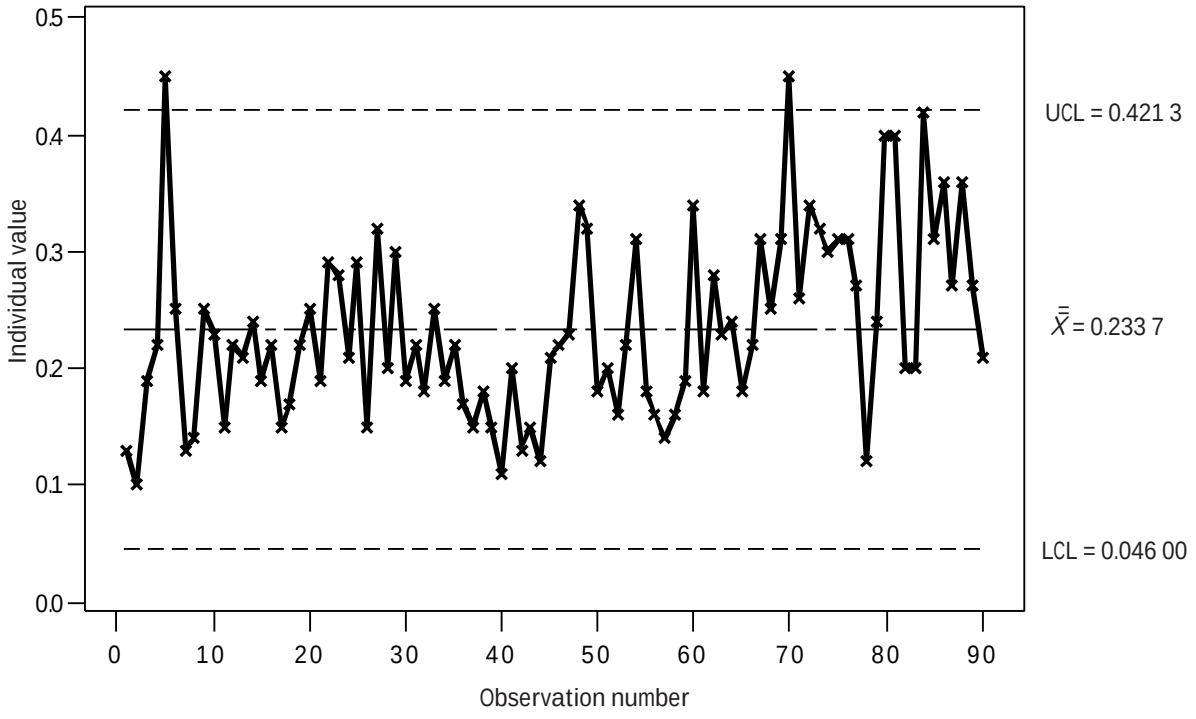


Figure 53 — Standard Shewhart type control chart for blast furnace % silicon *(*with control limits based on normality of data*)*

The control chart of Figure 53 indicates out-of-control situations:

— due to points above the control limits at observations number 5 and 70;

— runs below and above the centre-line.

However, the control limits are based on normality of the data, and as individuals are plotted rather than means (which are protected by the central limit theorem) it is necessary, prior to plotting such a control chart, to check on the pattern of variation of the observations. A histogram for the data is shown in Figure 54.

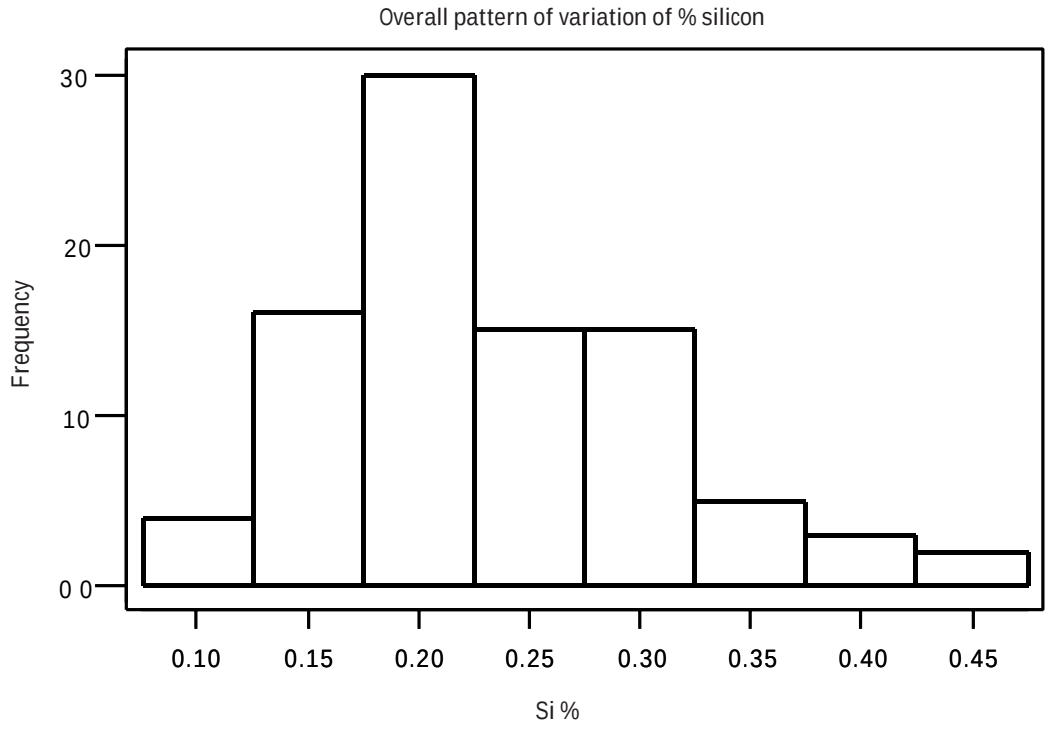


Figure 54 — Histogram for % silicon data

Figure 54 indicates that the pattern of variation of the % silicon data is skewed rather than normal. In such a case a simple graphical resolution of two concerns is now possible, that of:

— determination of control limits;

— estimation of capability of the characteristic;

based on the skew distribution.

This solution is provided by a skew distribution worksheet based on “extreme value” probability paper. Figure 55 shows the application of such a worksheet to the solution of both concerns.

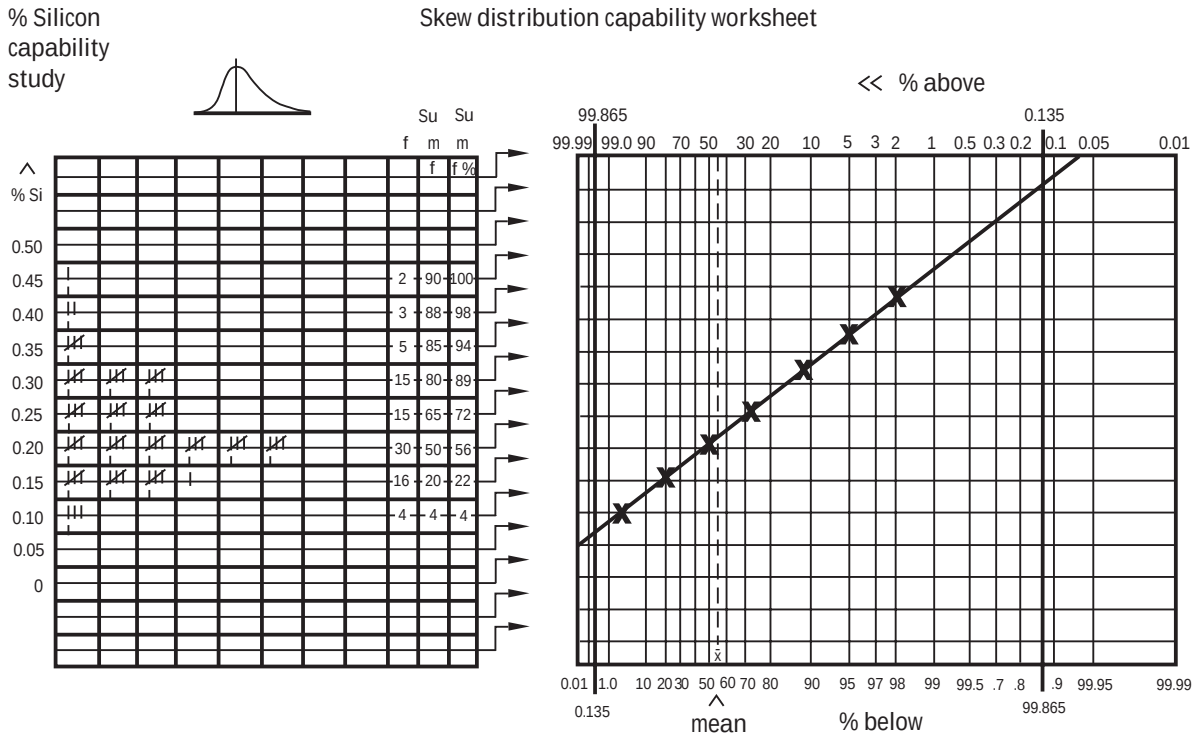


Figure 55 — Application of Skew distribution probability paper to the derivation of control limits and process capability for % silicon

Probability worksheet method

Figure 55 shows a more convenient and practical way of handling a modest skew distribution than the usual log v normal probability paper. It uses a linear vertical scale which facilitates tally charting and an "Extreme Value" probability horizontal scale, whereas the log v normal probability paper uses a logarithmic vertical scale for the measurement and a normal probability horizontal scale.

Using the worksheet shown, the data is first tallied using "five bar gates" as indicated in Figure 55. This indicates a modestly skewed distribution which, if plotted on standard linear v normal probability paper, yields a concave upwards curve which is difficult to extrapolate into the regions of interest at the tails of the distribution. This indicates an incorrect initial choice, as the primary objective of probability paper is to transform the cumulative frequency distribution into a straight line reference standard. This is why the linear v extreme value probability paper was chosen here.

The numerical frequencies within each class interval are entered in column "f". The cumulative frequencies are put into column "$\sum$f %". The cumulative frequencies are then transformed into percentages in column "$\sum$f %".

The cumulative percentages are then plotted in line with the top of the appropriate measurement interval class, as denoted by the arrows at the appropriate percentage on the bottom horizontal scale of the worksheet.

If the appropriate horizontal scale has been chosen, the plotted points will follow a straight line as indicated. This facilitates extrapolation beyond the sample values experienced into the tails of the distribution, which is often of greater interest.

Solution 1: Process capability

The equivalent $\overline{\overline{X}} \pm 3s$ capability for the normal distribution is given here in terms of the mean and the 0.135 % above and 0.135 % below intercepts with the measurement scale thus:

$$0.23^{+\ 0.37}_{-0.13}$$

As the process was not in statistical control during the period in which the data was taken, the results given may not provide a reliable prediction of the capability of this characteristic. However, the prime intention here is to provide an illustration of a simple graphical method for deriving process capability of non-normally distributed data.

The second use of the worksheet is that of derivation of control limits for a control chart for individuals when the distribution of individual values is skewed.

Solution 2: Control limits

As the control chart (Figure 53) is for individuals, these *also* represent its upper and lower control limits rather than the usual ± 3 standard deviations, namely:

UCL = 0.60; LCL = 0.10; not

UCL = 0.42 and LCL = 0.05.

Of course, having set the control limits and observed special causes of variation, these should be investigated and the control limits then modified to reflect only common causes of variation.

11.4 Process capability indices

11.4.1 *General*

Process capability indices provide simple standardized metrics, in world-wide use, which assess the capability of measured characteristics in relation to specified requirements. The application of these indices is growing rapidly with an increasing number of customers requiring documented proof of first time quality through:

— the achievement and demonstration of appropriate control chart stability, together with;

— the realization and confirmation of minimum value capability indices for significant process parameters and product characteristics.

Of equal consequence is the use of SPC and capability indices to provide suppliers themselves with the means to use "first time quality health profiling" as a business tool within their organization and those of their sub-suppliers.

Originally, capability indices were intended for use with normally distributed data. Unfortunately, there are those who still calculate and declare indices based on normality even when the distribution is patently non-normal. This has arisen, to a large degree, by the equations for the indices often appearing to be generic when indeed they are specific to the normal distribution.

Discussion of specific capability indices is confined to standardized ones that are in general use.

11.4.2 *The Cp index*

The Cp process capability index relates a standardized process spread to the specified tolerance interval. It does not take the location (e.g. mean) of the distribution into account. Generically, for a process in control, it is given by:

$$\text{Cp} = \frac{\text{Permissible range of values}}{\text{Actual standardized range of values}} = \frac{\text{Specified tolerance}}{(99.865 \text{ percentile} - 0.135 \text{ percentile})^*}$$

NOTE* The Cp index is referenced worldwide, quite arbitrarily, against the probability equivalent to 6 standard deviations for the normal distribution. Figure 52 indicates the significance of the 99.865 and 0.135 percentiles.

For a normal distribution this expression reduces to:

$$\text{Cp} = \frac{\text{Specified tolerance}}{6 \text{ standard deviations}} = \frac{U - L}{6 \text{ standard deviations}}$$

As Cp does not take the location of the distribution into account it provides a value for the relative capability of a centred process. For a non-centred process it represents the potential capability of the process parameter or product characteristic. Hence Cp should always be used in conjunction with other indices that do take location into account.

The minimum acceptable value of Cp will depend on the appropriate customer contractual requirement or benchmark set internally by the supplier for a given application. In some business sectors Cp $\geq$ 1.33. For a centred process having a normal distribution a Cp of 1.33 can be expected to give rise to 32 ppm above and 32 ppm below specification limits. Substitution in the equation:

$$\text{Cp} = \frac{\text{Specified tolerance}}{6 \text{ standard deviations}} = \frac{U - L}{6 \text{ standard deviations}}$$

indicates that:

— a Cp of 1.33 equates to: $U - L = 8$ standard deviations. If centred and normal, from Table 7, this can be expected to give rise to 32 ppm above and 32 ppm below specification limits; namely, nearly 99.994 % conforming to specification;

— a Cp of 1.00 equates to: $U - L = 6$ standard deviations. If centred and normal this will give rise to 0.135 % above and 0.135 % below specification limits; namely 99.73 % conforming to specification.

Cp is an estimate. It is thus subject to sampling variation. Strictly speaking, confidence intervals should be computed to provide a range of Cps which include the true Cp with a given probability. A centred process is then deemed capable if Cp $\geq$ lower confidence limit. In practice, it is the exception rather than the rule to use such confidence limits. Table 27 (due to Li, Owen and Borrego) [40] provides values with which to factor the estimated Cp to obtain the 95 % lower confidence limit for a range of subgroup sizes in terms of number of subgroups.

Table 27 — Multiply the tabulated value by Cp estimate to obtain Cp_{min} at the 95 % confidence level

Subgroup size	Number of subgroups				
	1	5	10	20	30
3	0.255	0.631	0.735	0.811	0.845
4	0.369	0.697	0.783	0.845	0.873
5	0.443	0.735	0.811	0.865	0.890
6	0.494	0.760	0.829	0.879	0.901
7	0.533	0.779	0.843	0.888	0.908
8	0.562	0.793	0.853	0.895	0.914
9	0.586	0.804	0.861	0.901	0.919
10	0.605	0.813	0.867	0.906	0.923

Example

Cp has been calculated as 1.60, based on the average subgroup range of 20 subgroups of 5, for a stable process having a normal distribution.

From the table, Cp_{min} (at the 95 % confidence level) = 0.865 × 1.60 = 1.38.

11.4.3 *The Cpk family of indices*

There are three indices from the *Cpk* family in general use. These are:

$$Cpk_U = \frac{U - \text{mean}}{\text{Range between mean and upper 0.135 distribution percentile}}$$

$$Cpk_L = \frac{\text{Mean} - L}{\text{Range between mean and lower 0.135 distribution percentile}}$$

Cpk = Minimum of Cpk_U and Cpk_L

For a normal distribution Cpk_U and Cpk_L reduce to:

$$Cpk_U = \frac{U - \text{mean}}{3 \text{ standard deviations}}$$

$$Cpk_L = \frac{\text{Mean} - L}{3 \text{ standard deviations}}$$

The *Cpk* family of indices relates both the process variability and the location (setting) of the process in relation to specification limits.

Cpk_U is an index which relates process variability and location to the upper specification limit; whereas Cpk_L relates process variability and location to the lower specification limit. This is shown graphically in Figure 56.

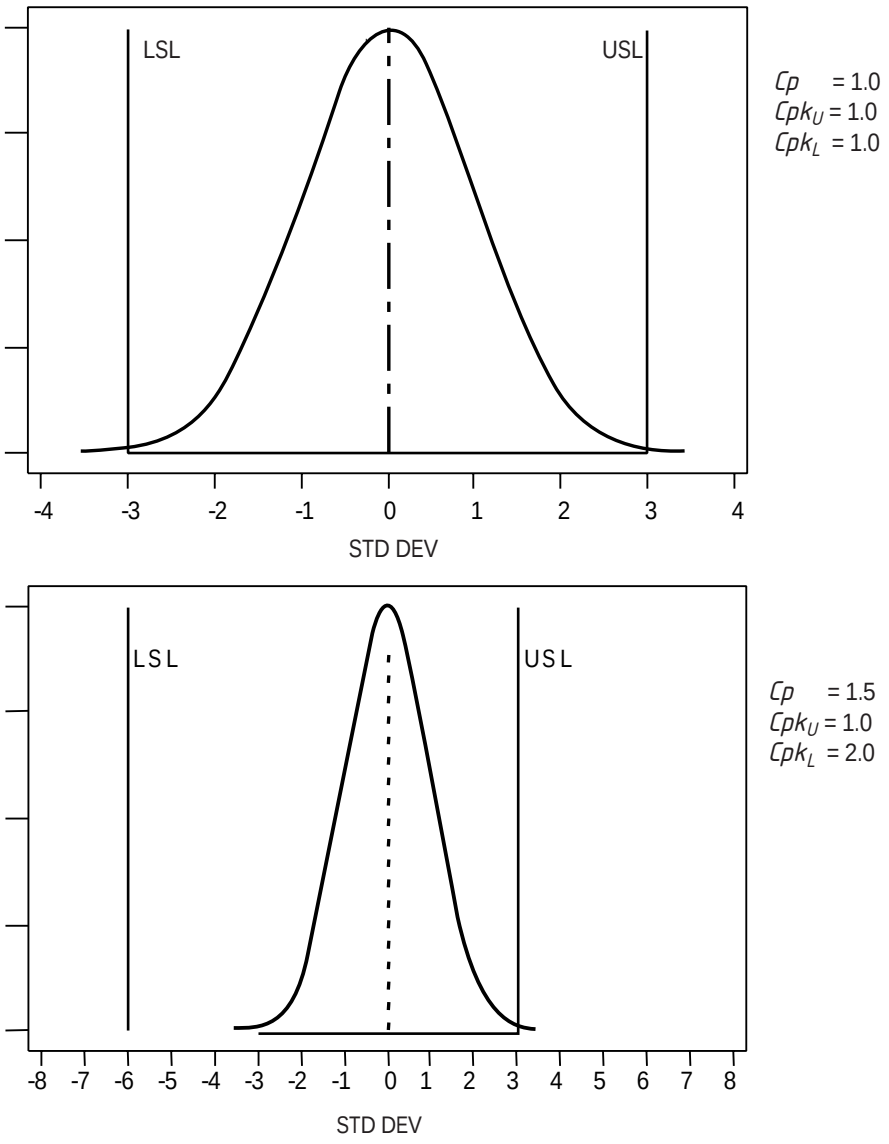


Figure 56 — Relationship between *Cp* and Cpk_U and Cpk_L

For a single-sided specification limit, only one of these indices can be calculated. Knowing Cpk_U and/or Cpk_L the proportion lying outside of a specification limit can be determined using Table 28.

Cpk, the lowest of Cpk_U and Cpk_L, is sometimes quoted alone as a minimum standard in contractual requirements. Typical such values are:

$Cpk \geq 1.33$

However, it should be borne in mind that *Cpk*, on its own, gives no indication of the direction in which the process is biased, if at all; the location of the distribution; or the extent of the variation. It thus degrades the information conveyed. This is particularly relevant to a supplier if the penalty of transgressing one limit is different from transgressing the other. Such a situation could arise, for example, with a characteristic such as a length, too short could give rise to scrap and too long to less expensive rework. Neither is the preferred value always on nominal: for example, if minimum is best then one would aim for the minimum acceptable Cpk_L whilst, at the same time, maximizing *Cp* and Cpk_U.

In practice the minimum standard for Cpk_U and Cpk_L is often taken to be 1.33. However, this will depend on contractual requirements or self imposed benchmarks currently in place in a given sector or organization.

These indices are becoming more widely used:

— by customers for supplier process certification/accreditation in certain industrial sectors. An example is the automotive sector, which specifies:

— by suppliers to provide quality health profiles for their organizations. A typical example of such a profile is shown in Table 28.

To avoid confusion, or worse, it is recommended that any statement of capability using these indices should contain at least five items of information, viz. *Cp*, Cpk_U, Cpk_L, distribution shape and an indication of the preferred value, namely maximum, minimum or nominal is best. Table 28 provides such information.

Table 28 — Steel works quality health profile for selected process characteristics

Characteristic	Aim	In control	Distribution	Capability		
				Cpk_L	*Cp*	Cpk_U
Silicon	nominal	yes	skew	1.3	1.0	0.9
Aluminium	nominal	yes	normal	1.4	1.5	1.6
Teeming temperature	nominal	yes	normal	1.3	1.3	1.3
Teeming time	nominal	yes	normal	1.6	1.7	1.8
Injuries per workforce per week	minimum	yes	attribute	0.73 %		
Cobbles	minimum	yes	attribute	0.14 %		
Billet rhomboidity	minimum	yes	normal	2.4	—	—
Time to charge	minimum	no	bi-modal	disparity between steelman[a]		

[a] Subject of investigation.

As *Cpk* is an estimate, from a statistical viewpoint, it should be qualified by confidence bounds between which the true value is expected to lie. In such cases, frequently the lower confidence limit only is quoted. Table 29, due to Chou, Owen and Borrega [41] provides such limits for a range of *Cpk*s and sample sizes.

Table 29 — Approximate 95 % lower confidence limits for *Cpk* (in terms of estimated *Cpk* and sample size for a normal distribution)

Cpk_{est}	Sample size, *n*						
	20	30	40	50	100	200	300
1.0	0.66	0.72	0.76	0.79	0.85	0.89	0.91
1.2	0.81	0.88	0.93	0.95	1.03	1.08	1.10
1.4	0.95	1.04	1.09	1.12	1.20	1.26	1.29
1.6	1.10	1.20	1.25	1.29	1.38	1.45	1.47
1.8	1.25	1.35	1.41	1.46	1.56	1.63	1.66
2.0	1.39	1.51	1.58	1.62	1.74	1.81	1.85

11.4.4 *The Cpm index*

11.4.4.1 *Current specification practice v optimal design values*

The point value of a measured characteristic expressed in a design specification is intended to reflect preferred value. This focus has been diffused by two practices:

— the setting of acceptable tolerances around the preferred value to reflect the presence of some variation in the realization process, for example, 20.0 ± 0.1 mm;

— the quoting of the range of permissible values, for example, 20 Nm to 80 Nm, without any reference to a preferred value. This leaves it open as to whether nominal, minimum or maximum is best.

These practices can give rise to two types of response:

— no emphasis or regard is placed on achieving preferred value. The "goalpost mentality" prevails; namely, anything within the specified tolerance represents acceptable, or even, A1, quality;

— aiming at a value which is most cost effective from the supplier's point of view, often to the detriment of the customer. If this is coupled with a drive to minimize variation in the process to maximize the gain to the supplier, then this can be even more detrimental to the customer. An example would be to offset the aim of the process towards the specification limit which provides the greatest saving in material. As an illustration, a garment manufacturer who buys wool by the kilogram could knit more jumpers or cardigans per kilogram if the wool is on the thinner side (higher "count" wool) of the specified tolerance. Whilst this would be cost advantageous to the manufacturer, it would be to the detriment of the retailer and of the ultimate wearer.

11.4.4.2 *Expression for Cpm index*

An index, *Cpm*, has been devised to provide a single quantitative measure of diminished utility which can arise in terms of process offset from preferred value and the extent of process variability.

$$Cpm = \frac{U - L}{6\sqrt{s^2 + (\overline{X} - T)^2}}$$

where

s is the standard deviation;

$\overline{X}$ is the process mean;

T is the target value.

When, $\overline{X} = T$, $Cpm = Cp$.

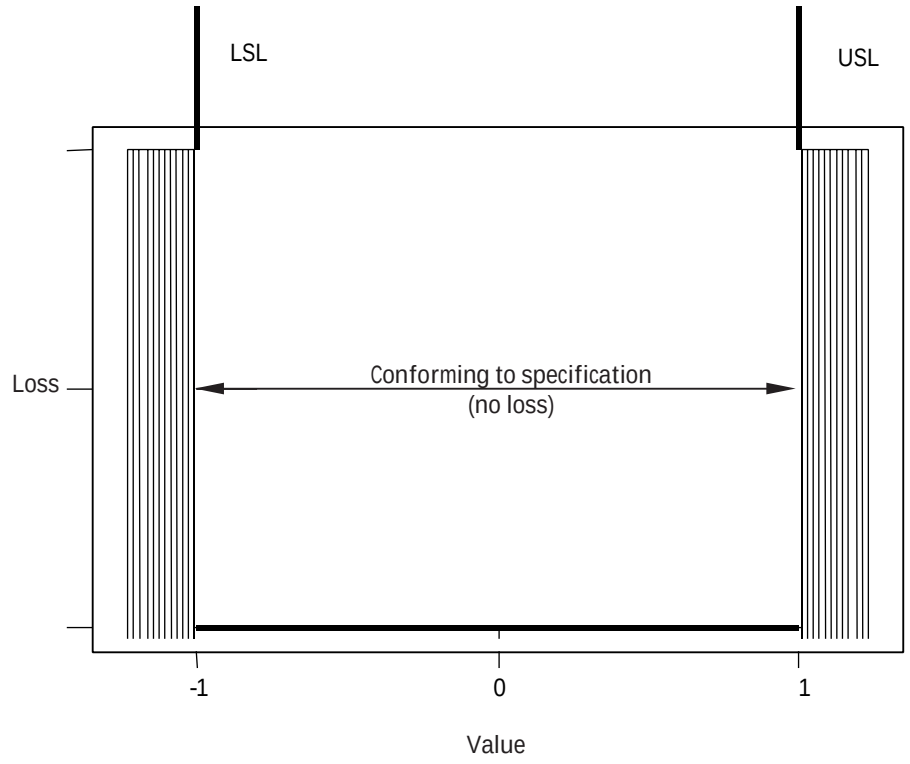
As the mean drifts from the target and/or the standard deviation increases, the *Cpm* value declines.

Thus *Cpm* is a measure of both process spread and level in relation to the target value. The use of *Cpm* refocuses on the targeting of optimal values rather than a degraded minimum requirement of conformance to specified tolerances.

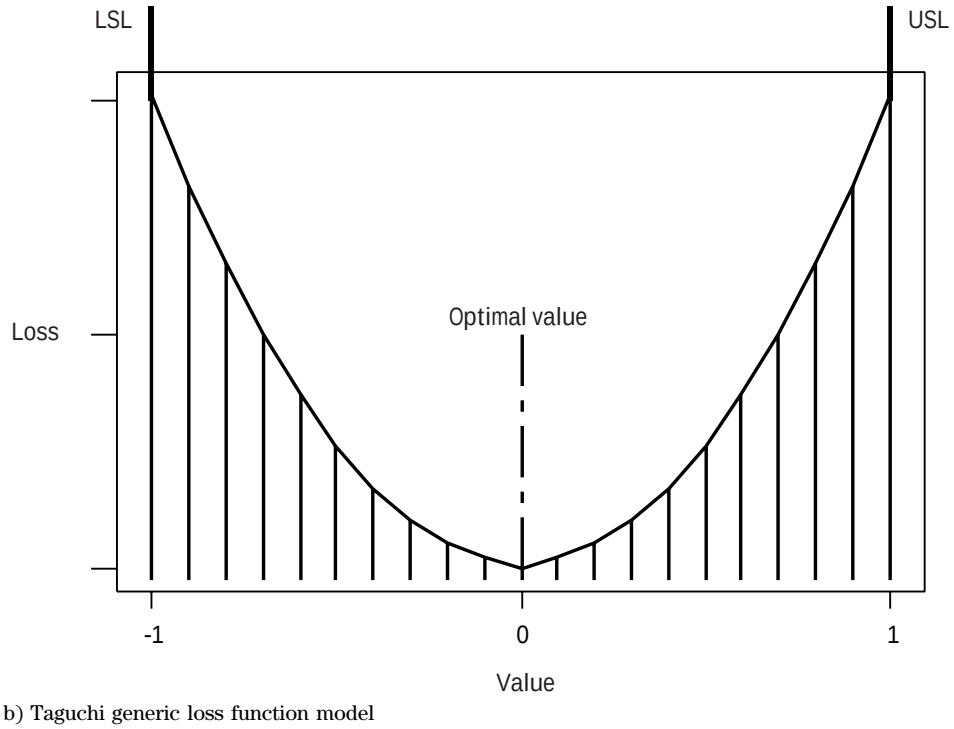
The *Cpm* index is based on some fundamental loss concepts as illustrated in Figure 57.

11.4.4.3 *Basis of Cpm index*

Quality is frequently defined as "conformance to specification". Traditionally, such specifications for measured characteristics embrace an allowable tolerance band. This widely practised approach is based on the "*goal post*" mentality, as indicated in Figure 57.



a) "Goal post" loss function model



b) Taguchi generic loss function model

Figure 57 — Comparison of conformance to toleranced specification with optimal value approach

In terms of loss function it assumes, in the model of Figure 57a), that there is no loss for values of a characteristic anywhere within a specified tolerance band but there is an incremental loss for those beyond the specification limits.

The implications of this mind set are that:

— all characteristic values within the specified tolerance range are equally acceptable so there is no or little incentive to aim at an optimal design value; namely, it produces a mindset which inhibits quality improvement. However, it does enable clear cut decisions to be made about conformity;

— exploitation of a) is acceptable to the detriment of the customer. One example of deliberately *offsetting* the process to achieve gain to the supplier at the expense of the customer has already been given in **11.4.3.1** for the wool garment manufacturer. Another illustration of this exploitation of relatively low process variation relative to specified tolerance is the practice of permitting the process mean to *drift* across the specified range. This can arise fortuitously due to lack of statistical control of a process or deliberately in situations, for instance, involving physical tool wear or progressive diminishing in the strength of a solution. In such cases this will result in marginally acceptable characteristics at the start and end of each cycle of tool replacement or topping up. Such practices frequently give rise to a decrease in utility to the customer.

This goal post mentality model is contrasted with the Taguchi [42] economic loss model shown in Figure 57b). To obviate the need for extensive calculations for each and every design characteristic, Taguchi advocates the use of a generic quadratic loss function. The resulting parabola has its minimum point at zero at the optimal design value and rises on either side in proportion to the square of the distance from the preferred or target value. This quadratic loss function can be conveniently split into two elements:

— the process variance (s^2) around its own mean;

— the square of the offset of the process mean from the target.

Thence,

average loss = $ks^2 + k(\overline{X} - T)^2$

where

k is the loss parameter.

This gives rise to the function for *Cpm*.

Complications arise in the use of *Cpm* if the optimal value is not the mid-point between specification limits, if the distribution is non-normal or if the process is not under statistical control.

Although the generic quadratic loss function may be difficult to quantify in specific instances, one should not lose sight of the very important message it conveys. That is that:

"*quality as perceived by a customer is not a go: no-go situation. There is an optimum or target value. As a characteristic varies from this point, the perception of quality progressively deteriorates until at some point, possibly a specification limit, the condition becomes untenable.*"

A simple example of this is ambient temperature. Although in an industrial situation there may be statutory maximum and minimum values laid down, any deviation from the perceived ideal value may cause a degree of discomfort depending on the extent of that deviation.

11.5 Process capability for attribute data

The capability of an attribute process is obtained simply from the centre-line of the attribute control chart of a stable process. It is typically expressed as:

a) average nonconformities or faults per unit for c and u charts;

b) average proportion of units nonconforming for p and $n.p$ charts.

Removing special causes of variation from an attribute process, through elimination of *sporadic* causes of variation, restores the status quo. It does not improve it.

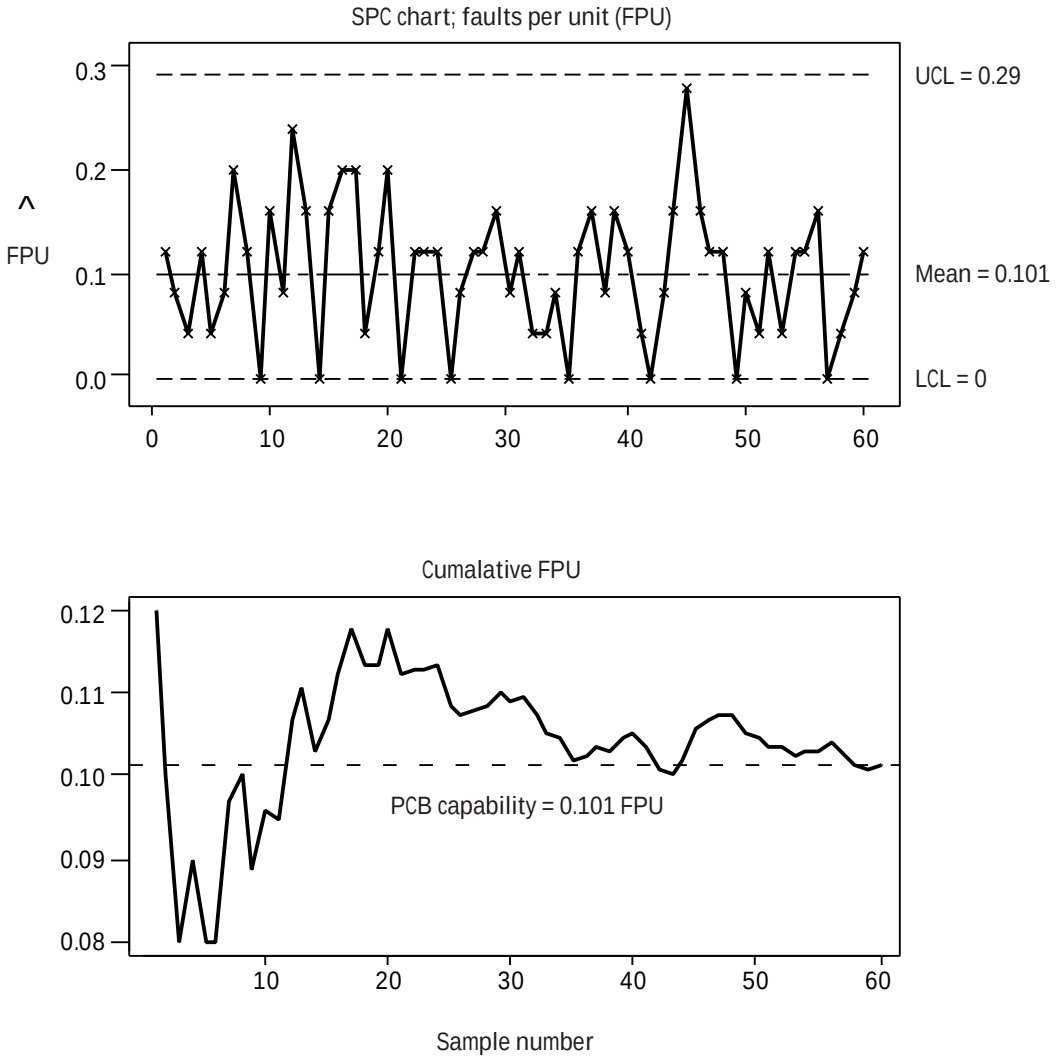
A process in statistical control reflects *systemic* causes of variation, the extent of which is indicated by the status quo or on-going level of performance, namely the capability of the process. Reduction of systemic causes demands fundamental changes in approach to that adopted for the removal of sporadic causes. An example of this is shown in Figure 42. Changing the capability to a more favourable value improves attribute process performance. This is seen to be achieved when the level of performance moves significantly (shown by an "out-of-control" control chart) towards the preferred or targeted value of capability. This preferred value is frequently zero where nonconformities or nonconforming items are concerned. However, the targeted value should be realistic and reflect the specific diagnostic and improvement programme established to achieve it.

Example

Printed circuit boards of a particular type are assembled in batches of 25. These are 100 % inspected in assembly sequence and the number of nonconformities/faults per batch recorded. The results of 60 such batches are shown in Table 30. Determine the capability of the printed circuit board assembly process.

Table 30 — Faults per batch on printed circuit boards (60 batches of 25)

Batch no.	Faults	Batch no.	Faults	Batch no.	Faults
1	3	21	0	41	1
2	2	22	3	42	0
3	1	23	3	43	2
4	3	24	3	44	4
5	1	25	0	45	7
6	2	26	2	46	4
7	5	27	3	47	4
8	3	28	3	48	3
9	0	29	4	49	0
10	4	30	2	50	2
11	2	31	3	51	1
12	6	32	1	52	3
13	4	33	1	53	1
14	0	34	2	54	3
15	4	35	0	55	3
16	5	36	3	56	4
17	5	37	4	57	0
18	1	38	2	58	1
19	3	39	4	59	2
20	5	40	3	60	3



NOTE 1 The SPC chart is a plot of number of faults/batch size, namely faults per unit or faults per printed circuit board here. Hence the first plot is 3/25 = 0.12, the second plot 2/25 = 0.08 and so on. The upper control limit is given by:

$$\text{UCL} = mean + 3\sqrt{\frac{mean}{sample\ size}} = 0.101 + 3\sqrt{\frac{0.101}{25}} = 0.101 + 0.191 = 0.29$$

NOTE 2 The faults per unit (FPU) chart is a plot of the cumulative faults per unit.

Hence the first plot is 3/25 = 0.12, the second one is (3 + 2)/(25 + 25) = 5/50 = 0.10, the third one is (3 + 2 + 1)/(25 + 25 +25) = 6/75 = 0.08, and so on.

Figure 58 — Printed circuit board faults SPC chart and cumulative faults per unit (FPU) chart

The attributes control chart for this data shown in Figure 58 is seen to be in statistical control. Hence capability may be calculated from the control chart mean. This is given as 0.101 faults per unit.

A plot is also shown in Figure 58 of cumulative faults per unit. In practice it is recommended that this figure be plotted as well as the control chart to determine whether or not the capability value has stabilized. It is seen that it starts to stabilize at about the 35^{th} batch in this particular case. Any prediction of capability prior to this could be unreliable.

In summary:

1) the control chart confirms that the process is in a state of statistical control and provides a value of the overall mean;

2) the cumulative chart indicates when enough data has been collected to provide a stable estimate of the process capability.

12 Statistical experimentation and standards

12.1 Basic concepts

12.1.1 *What is involved in experimentation?*

An experiment involves changing things that are believed to have an effect on the performance of the process, product or service. By changing from one set of conditions to another, to a pre-determined pattern, the actual effect can be estimated. In an experiment:

a) the things that are changed are called *factors*;

b) the conditions to which the factors are changed/set are known as *levels*[3)];

c) the value of the performance characteristic outcome is called the *response*;

d) the change in the response as a result of a change in factor level is termed an *effect*.

12.1.2 *Why experiment?*

Experimentation has many practical uses. It enables one to determine how standards of performance, dependability, acceptability and affordability of products and services, processes, materials and mixtures are influenced by:

a) features of products and services (e.g. tolerances, nominal values);

b) parameters of processes (e.g. temperature, pressure);

c) properties of materials (e.g. hardness, machinability);

d) formulations of mixtures (e.g. of alloys, fuels, concrete, cloth).

Whilst experimentation plays a major role in problem solving, there is a need to progressively shift the emphasis to its integration in the mainstream activities of design and development. Genichi Taguchi [43] has proposed a two-step approach, which uses experimentation to "tune in" a basic prototype design, which he terms "parameter" design and "tolerance" design.

Parameter design is concerned with the identification and exploitation of three types of design factor:

— control factors: those which affect the variability of the response;

— signal factors: those which affect only the level of the response;

— null factors: those which do not materially affect either the variability or level of response.

Firstly, control factors are identified and adjusted to achieve design "robustness". A robust design is one which is insensitive to, so called, noise factors which are impossible, inconvenient or impractical to manage. Examples of noise factors are: environmental, ambient temperature, humidity, vibration, supply voltage and dust; deterioration, wear, drift and fatigue; and imperfections in manufacture, delivery or use, deviations from nominal.

Secondly, signal factors are adjusted to bring the response on target.

Thirdly, the null factors are adjusted to the most economic level.

The overall effect in identifying and setting nominal values of design factors in this way is to achieve optimal performance over a wide range of conditions with economy.

Tolerance design is concerned with specifying the most liberal tolerances and controls to meet a given performance. This is achieved by experimentation which seeks to take advantage of any non-linear relationship between factors and responses.

12.1.3 *Where does statistics come in?*

Today's statistical experimental designs emanate from R.A. Fisher's [44] work in England in the 1920s. Prior to this it was deemed scientifically sound to conduct a multi-factor experiment by varying the level of one factor at a time, keeping the levels of all other factors constant. Fisher introduced the concept of a *factorial experimental design* in which all factors are varied simultaneously. The principal motivators for using such statistically designed experiments include:

a) increase in information for a given number of experimental runs, including the separation of main effects, interactions and experimental "noise";

b) potential for cost and time savings through the reduction in the number of experimental runs required for a given effectiveness and the ability to plan and execute tests more efficiently;

c) ability to predict optimal combinations of factor levels even when they do not form part of the actual experimental plan;

d) ability to adopt a sequential rather than a one-shot approach;

e) relative ease of analysis and interpretation of the results.

[3)] The term "level" is normally associated with a quantitative characteristic such as temperature, in which case differing experimental levels could be 200 °C and 220 °C, say. In experimentation it also serves as the term describing the setting of a qualitative characteristic, for example, the absence or presence of a catalyst, compound A or compound B and matt or gloss ink base.

12.1.4 *What types of standard experimental designs are there and how does one make a choice of which to use?*

12.1.4.1 *Full factorial experiments*

Full factorial experiments in the form of orthogonal (balanced) arrays are well suited for determining the extent to which the effect on the response of a change in level of a factor differs at different levels of other factors.

However, when the number of factors and/or their levels become large the size of a full factorial experiment can become prohibitively large. For example, to test all combinations of 6 factors each at 4 levels would require a minimum of $4^6 = 4\,096$ experimental runs. Additional runs would still be required to investigate variation in the response at each combination and to estimate experimental noise. In such an event *fractional factorial designs* often provide an economic solution which is technically adequate particularly in situations where higher order interactions or non linearity can be safely ignored.

12.1.4.2 *Fractional factorial experiments*

Fractional factorial designs stem from the work of Tippett, Finney and Rao in the 1930s and 1940s. More recently they have been popularized by Taguchi [43]. A number of orthogonal arrays are available together with simple, mainly pictorial, instructions for selection, application and analysis. The versatility of the most popular basic two level orthogonal array is shown in Table 31. It is seen that if technical considerations indicate that some interactions are not likely to be important, then considerable economy in experimental effort is possible. At least a three level design is required to investigate non-linearity.

The L8 design of Table 31 is a standard orthogonal (balanced) array with seven columns and eight rows. Factors A, B, C, etc. can be assigned to the columns. Factor levels are indicated by a 1 or a 2. In some texts, minus and plus signs are used instead. Each row indicates a combination of factor levels to run in the experiment. The design is such that four independent estimates can be made of the effect of each factor on the response, at each level, under different operating conditions of other factors. These four estimates can then be averaged for each factor level. This is illustrated in the design validation and development example given later within this sub-clause.

In using these factorial designs a number of features need to be considered:

a) the statistical desirability of *randomizing* the run sequence to protect against bias due to factors not included in the experiment. For example, without randomization, take the situation if the first two runs of L8 were performed on Saturday morning, the next two on Saturday afternoon with the further four runs done similarly on Sunday. It would not be possible to separate out the day to day effect present in column 1 from the factor A effect. Statisticians would say the effects are confounded. Neither would it be possible to separate out the morning to afternoon effect in column 2 from the factor B effect. On the other hand operational interests would prefer to retain the order given in Table 31 if some factor levels are more difficult to change than others. The most difficult factor to change would be put into column 1 which has the minimum number of changes and the easiest factor to change in column 4 which has the maximum number of changes. Hence the actual run order will often be based on a trade-off between statistical and operational considerations;

b) *replication/repetition* of the experiment for each specified combination of factor levels. This is desirable for two principal reasons: one, to estimate the value of any noise or error and; two, to provide a measure of the variability of the response at each combination. The latter is required if the aim of the experiment is to optimize the response with minimum variation;

c) *sequential experimentation*, as opposed to one-shot experiments. This is possible with the L8 design. This flexibility facilitates the building of knowledge as the experiment progresses in order to meet the experimental objectives with the minimum of effort and cost. For instance, if the L8 does not yield the information required, say, with four or seven factors, it may be extended into an L16, which has 15 factor columns and 16 runs.

Table 31 — Alternative useful designs with the "Taguchi" (Lattice) L8 two level array

L8 Lattice		Column for factors						
		1	2	3	4	5	6	7
Run no.	1	1	1	1	1	1	1	1
	2	1	1	1	2	2	2	2
	3	1	2	2	1	1	2	2
	4	1	2	2	2	2	1	1
	5	2	1	2	1	2	1	2
	6	2	1	2	2	1	2	1
	7	2	2	1	1	2	2	1
	8	2	2	1	2	1	1	2
Design 1: full factorial 3 factor design with all interactions isolated		**A**	**B**		**C**			
				AB		AC	BC	ABC
Design 2: 4 factor design with main effects clear of all 2 factor interactions		**A**	**B**		**C**			**D**
				AB		AC	BC	
				CD		BD	AD	
Design 3: 7 factor design with each factor confounded[a] with 3 two factor interactions (only 2 factor shown)		**A**	**B**	**C**	**D**	**E**	**F**	**G**
			AC	AB	AE	AD	AG	AF
		BC			BF	BG	BD	BE
					CG	CF	CE	CD
		DE	DF	DG				
			EG	EF				
		FE						

[a] A factor si said to be confounded with another factor, or factors, when their separate effects cannot be isolated.

Design validation and development example

This example shows an application of experimental design 2 of Table 31. It has two roles; one, as a design validation tool to determine the suitability of a sintered part for a particular application and two, as a development tool in the sense of searching for preferred operating conditions. Four design factors were investigated each at two levels as indicated in Table 32a).

Table 32a) — Sintered part design factors and their levels

Design factor	Level 1	Level 2
A. Surface finish	fine turned	microlled
B. Lubrication	yes — no 2 oil	no
C. Speed	low	high
D. Density	6.5	6.8

The experimental layout chosen uses columns 1, 2, 4 and 7 of a standard L8 array. Strength of fit, in kN, at minimum interference conditions was recorded for each part subjected to each experimental combination.

Three parts were used for each run in order to separate out means from variation in order to permit a search for design factors which would enhance mean strength (signal factors) and those which would reduce variation (control factors). Variation is expressed in terms of standard deviation. The results are shown in Table 32b).

Table 32b) — Results from the runs on sintered part experiment

Run	A	B	C	D	Result 1	Result 2	Result 3	Mean	Std. dev.
1	1	1	1	1	12.70	7.27	9.74	9.90	2.72
2	1	1	2	2	9.41	8.52	7.29	8.41	1.06
3	1	2	1	2	12.61	15.19	14.11	13.97	1.30
4	1	2	2	1	13.99	7.65	8.10	9.91	3.54
5	2	1	1	2	10.36	10.45	9.05	9.95	0.78
6	2	1	2	1	7.45	8.90	10.02	8.79	1.29
7	2	2	1	1	16.80	14.76	13.92	15.16	1.48
8	2	2	2	2	11.52	13.92	10.33	10.33	0.74

Analysis is carried out as follows.

An effect on the mean is given by:

$\overline{A}_1 - \overline{A}_2 = (9.90 + 8.41 + 13.97 + 9.91)/4 - (9.95 + 8.79 + 15.16 + 10.33)/4 = 10.55 - 11.06 = -0.51$

This indicates that the estimated effect of surface finish change from microlled to fine turned is to reduce the mean push-off strength by 0.51 kN.

Such analysis continues for the other effects in terms of both the mean and standard deviation.

Conclusions are shown graphically in Figure 59.

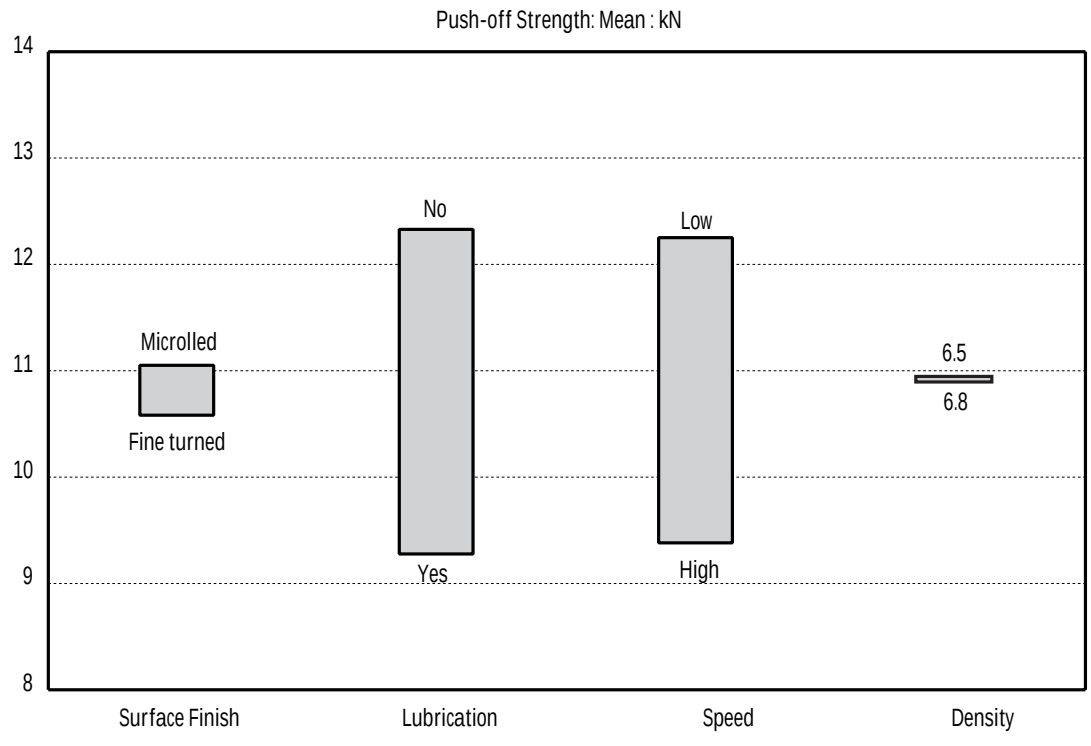


Figure 59a) — Effect of lubrication and speed on mean push-off strength

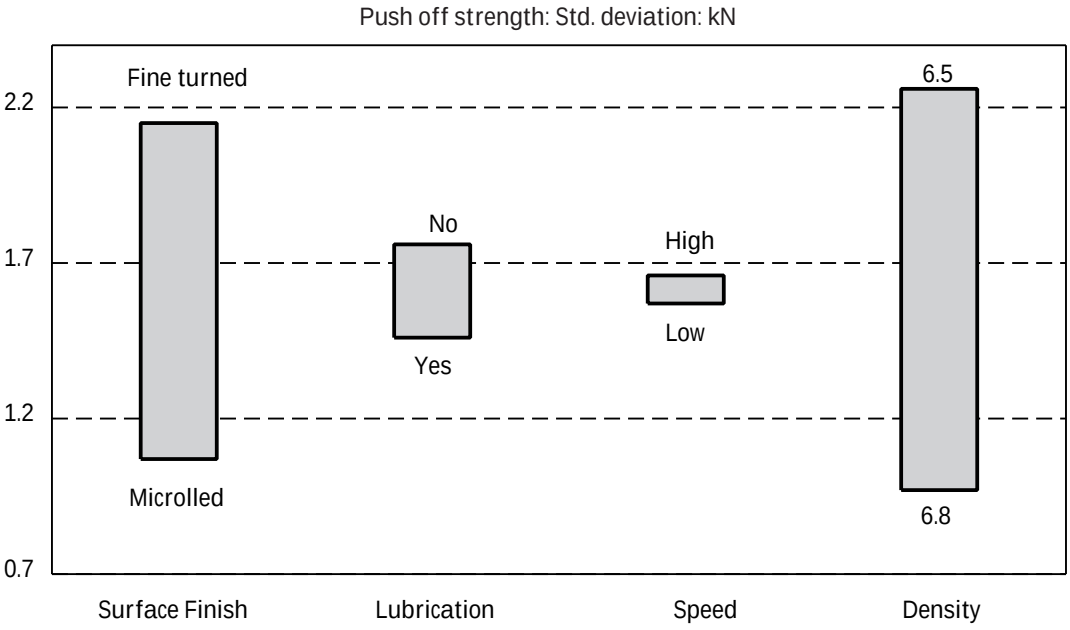


Figure 59b) — Effect of surface finish and density on standard deviation of push-off strength

Figure 59 illustrates that:

a) mean push-off strengths are influenced primarily by lubrication and speed: no lubrication and low speed producing the highest mean results;

b) variation is influenced primarily by surface finish and density: microlled and high density giving the lowest standard deviation.

The preferred conditions are seen, relative to average experimental results, to increase the mean strength by 34 % and reduce the variation by 53 %.

Problem solving example

This example illustrates the application of experimentation to problem solving where it is not possible to measure results, but simply to make a count of good/bad or record the yield or percentage nonconforming. It also illustrates how interaction effects can be assessed and presented and indicates how a process may be made more "robust".

On silk screen printing visual blemishes, termed "trail-marks", are being experienced. Squeegee speed, ink viscosity and dwell time were investigated. Factors and levels are shown in Table 33a).

Table 33a) — Silk screen printing design factors and levels

Design factor	Level 1	Level 2
Squeegee speed	45	80
Viscosity	700 cp	2 200 cp
Dwell time	auto	4.5

Ten blank sheets of polyester material were taken from each run of a standard L8 array (design 1 of Table 31) and a defect count was taken with the aid of a 200 square matrix. The total number of squares affected by trail-marks was counted for each run. This response was then recorded as a percentage. Results are shown in Table 33b).

Table 33b) — Results from the runs on the silk screen printing experiment

Run	Squeegee speed	Ink viscosity	Dwell time	% Trail marks
1	1	1	1	0.00
2	1	1	2	0.20
3	1	2	1	5.40
4	1	2	2	5.85
5	2	1	1	1.05
6	2	1	2	0.05
7	2	2	1	0.05
8	2	2	2	0.00

The significant interaction between squeegee speed and ink viscosity can be assessed and is shown visually in Table 33 c).

Table 33c) — Interaction matrix of squeegee speed and ink viscosity

		Ink viscosity	
		1	2
Squeegee speed	1	0/0.2	5.4/5.8
	2	1.05/.05	.05/0

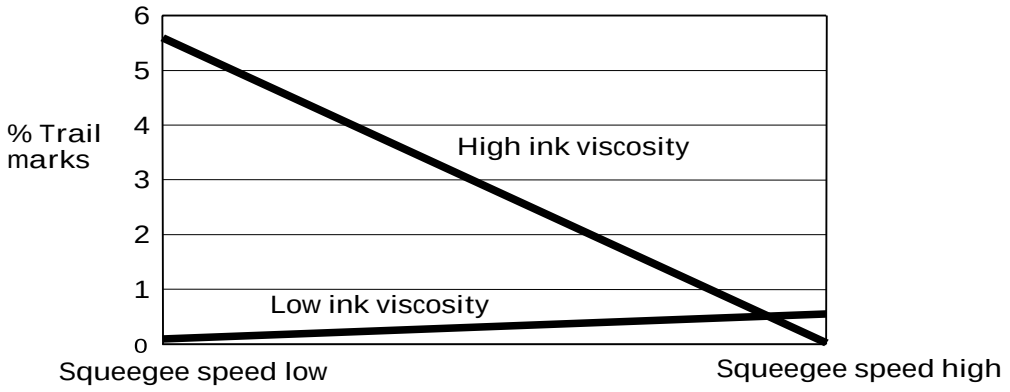


Figure 60 — Interaction between squeegee speed and ink viscosity

Figure 60 shows that at low ink viscosity the process is robust to changes in squeegee speed whereas at high ink viscosity it is not. This information enabled the screen printer to adjust process parameters and run with a consistently higher yield process having enhanced productivity.

12.1.4.3 *Nested or hierarchical design*

Another common experimental design is the *nested* or *hierarchical design*. In such a design each level of a given factor appears in only a single level of any other factor. An example is shown in Figure 47.

12.1.4.4 *Composite response surface designs*

Another important class of experiment is the *composite design*. These are used in the development of response surfaces to find optimal factor level combinations in the presence of interactions and non-linearity, where only first order non-linearity and two factor interactions are considered. One such design consists of three types of points: 2^k factorial points, $2k$ axial points and n centre points, where k is the number of factors. The number of tests required for a complete central composite design is thus:

$2^k + 2k + n$

A complete central composite design for three factors each at three levels is shown pictorially in Figure 61. The corresponding tabular arrangement is shown in the example. The axial points equate to the varying of one factor only with other factors kept constant at their nominal levels. The factorial points allow two factor interactions to be estimated. The axial and centre points together allow linear and quadratic (curvature) terms to be assessed. Although there are variations on the central composite design, it provides a fairly even coverage of the design space with economy.

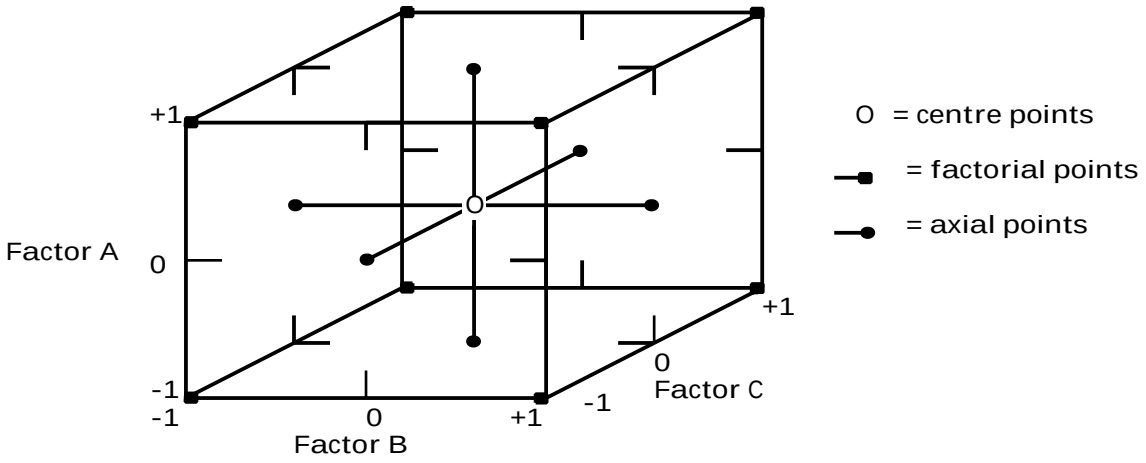


Figure 61 — Central composite design for 3 factors each at 3 levels: −1, 0 and +1

Example

The application of a central composite design to an etching process is now shown. Technical and operational considerations indicate that three factors, gas ratio, power and pulse may influence oxide uniformity. Non-linearity and interactions are expected so each factor was investigated at three levels using an 18 run central composite design as shown figuratively in Figure 61. The three levels of each factor were coded −1, 0 and 1 for convenience. Results were as shown in Table 34.

Table 34 — Results from experiment runs on etching process

Gas ratio	Pulse	Power	Oxide uniformity
0	0	0	29.4
0	0	0	32.1
0	0	0	31.5
0	0	0	30.9
−1	−1	−1	16.9
−1	1	−1	17.2
−1	1	1	22.7
−1	−1	1	52.4
1	−1	−1	10.7
1	1	−1	22.6
1	1	1	23.8
1	−1	1	43.5
0	−1	0	32.7
0	0	−1	16.4
0	1	0	24.1
0	0	1	37.5
1	0	0	27.6
−1	0	0	31.8

Using a standard computer-based technique, termed stepwise multiple regression, a statistical model was established which, when reduced to significant terms, became:

Oxide uniformity = 28.0 − 4.6 × pulse + 9.6 × power − 7.7 × pulse × power

The resulting contour plot is shown in Figure 62.

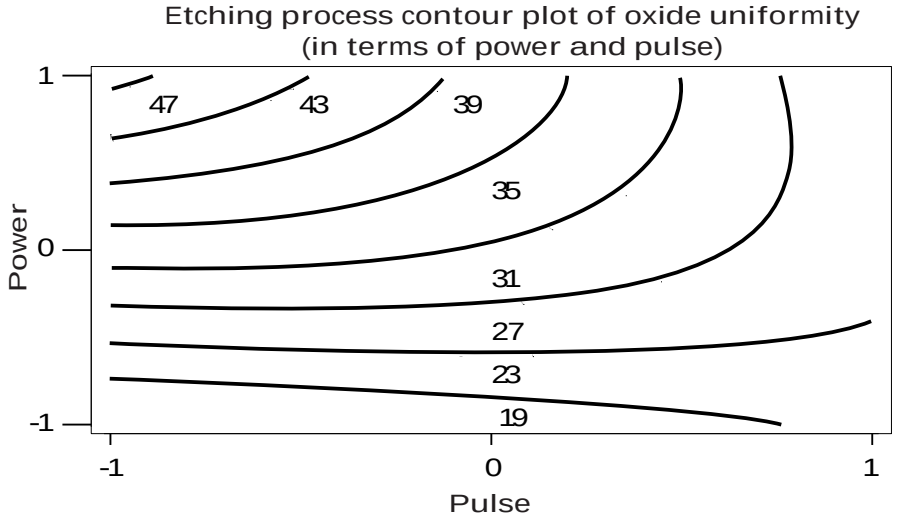


Figure 62 — Computer generated contour plot for oxide uniformity in terms of pulse and power

12.1.4.5 *Mixture designs*

A *mixture design* is a special class of experiment in which the response depends only on the relative proportions of the factors (ingredients) and not on their absolute amounts. They are applied to products comprising a blend of two or more ingredients in order to optimize the performance of various blends and mixtures. Such experimental designs are Simplex in form. For a 2 component blend the mixture space is a straight line, for a 3 component blend the mixture space is an equilateral triangle, and for a 4 component blend a regular tetrahedron. This arises because of the "constrained mixture design region" compared with the "unconstrained factor design region" of the central composite design of Figure 63. Figures 64 and 65 illustrate this feature.

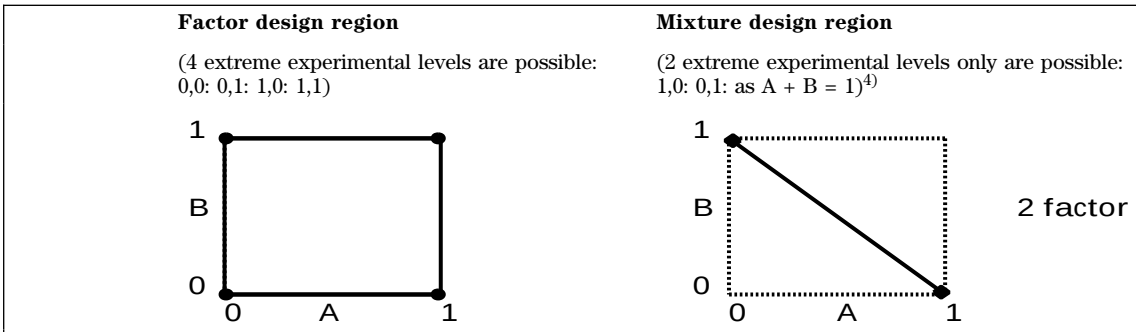


Figure 63 — Illustration of the fundamental difference in designs for 2 independent factors as compared with a 2 component mixture

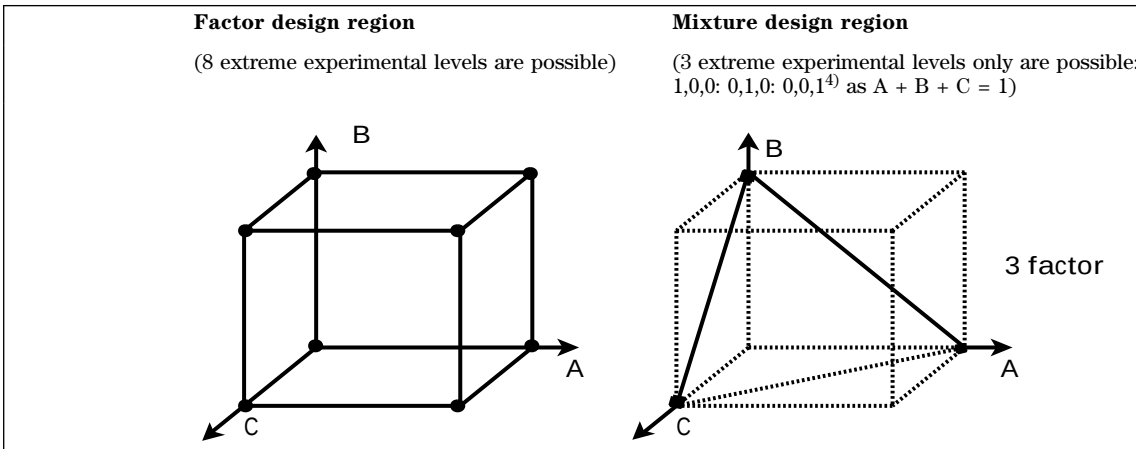


Figure 64 — Illustration of the fundamental difference in designs for 3 independent factors as compared with a 3 component mixture

Various mixture designs are available. A comprehensive treatment of mixture designs is due to Cornell [45]. A typical design for a 3 component blend, the *augmented simplex centroid*, is shown pictorially in Figure 65. The corresponding tabular arrangement is shown in the example. The augmented simplex centroid design is a popular mixture design with 6 points spaced on the perimeter, a further 3 interior axial points located midway between the centroid and the vertices and 1 point at the centroid.

4) Note to Figures 63 and 64: As with factor designs, mixture design levels are expressed in standardized format. For example: 1,0,0 does not indicate that the mixture is pure A with no B or C content. To determine the actual mix proportions it is necessary to refer to the lower proportion constraints together with either the corresponding upper proportion constraints or the total proportion of the mixture made up by the sum of the experimental components. For example suppose the fuel in a propellant mix has the following proportion constraints: $0.2 \leq$ fuel ≤ 0.4, then standardized 1 = actual 0.4, standardized 0 = actual 0.2 and standardized 0.5 = actual 0.3.

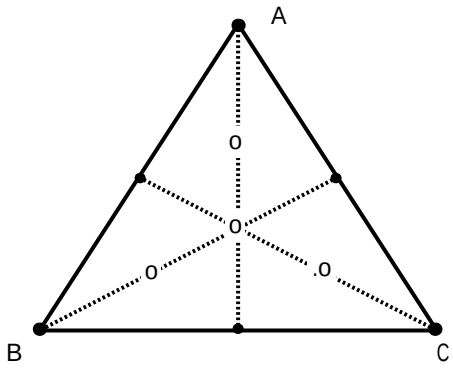


Figure 65 — Ten point augmented simplex centroid 3 component design

Example (due to Montgomery and Voth [46])

The utility of mixture, or blending, designs in product development is now illustrated. Suppose a blend of three components, fuel, oxidizer and binder, make up a propellant used in aircrew escape systems. Mixture constraints are:

- $0.30 \leq$ fuel ≤ 0.50;
- $0.20 \leq$ oxidizer ≤ 0.40;
- $0.20 \leq$ binder ≤ 0.40;
- fuel + oxidizer + binder = 0.9.

A response of particular interest is level and variation in burn rate. The design specification calls for a formulation that satisfies the following:

— mean burning rate >95 cm/sec;

—burning rate standard deviation <4.5 cm/sec.

In order to establish a statistical model relating fuel, oxidizer and binder concentrations to burn rate a standard ten point augmented simplex centroid experimental design was chosen as shown in Figure 65. This 10 point design was augmented by replicating the vertices twice and the centroid three times. This formed a 15 run design as shown in Table 35. Note that the factor levels are standardized, not actual, proportions as indicated in the note to Figure 64. A number of runs were made with each blend to enable calculation of both mean burning rate and burning rate standard deviation. The results are also shown in Table 35.

Table 35 — Propellant formulation experiment design and results

Fuel: X1	Oxidizer: X2	Binder: X3	Burn rate: mean	Burn rate: std. dev.
1	0	0	32.5	4.1
1	0	0	37.9	3.7
1/2	1/2	0	44.0	6.8
1/2	0	1/2	63.2	4.7
0	1	0	54.5	8.9
0	1	0	32.5	9.2
0	1/2	1/2	94.0	4.5
0	0	1	64.0	14.0
0	0	1	78.5	13.0
2/3	1/6	1/6	67.1	3.5
1/6	2/3	1/6	73.0	5.2
1/6	1/6	2/3	87.5	7.0
1/3	1/3	1/3	112.5	4.6
1/3	1/3	1/3	98.5	3.5
1/3	1/3	1/3	103.6	3.0

Computer generated propellant response contours for burn rate mean and burn rate standard deviation are given in Figures 66 and 67.

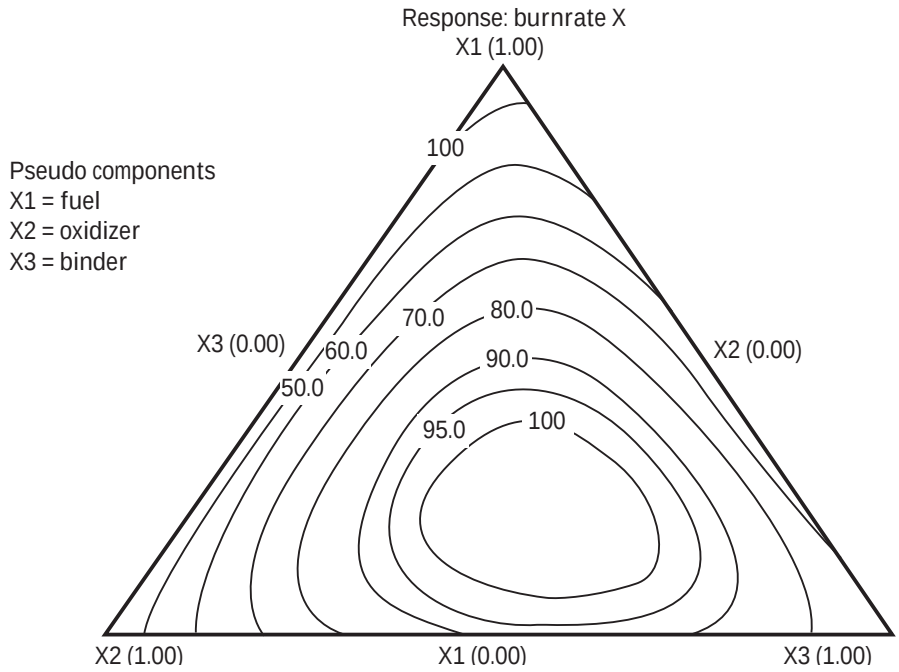


Figure 66 — Response surface contours for *mean* burn rate in terms of fuel (X1), oxidizer (X2) and binder (X3) blend components

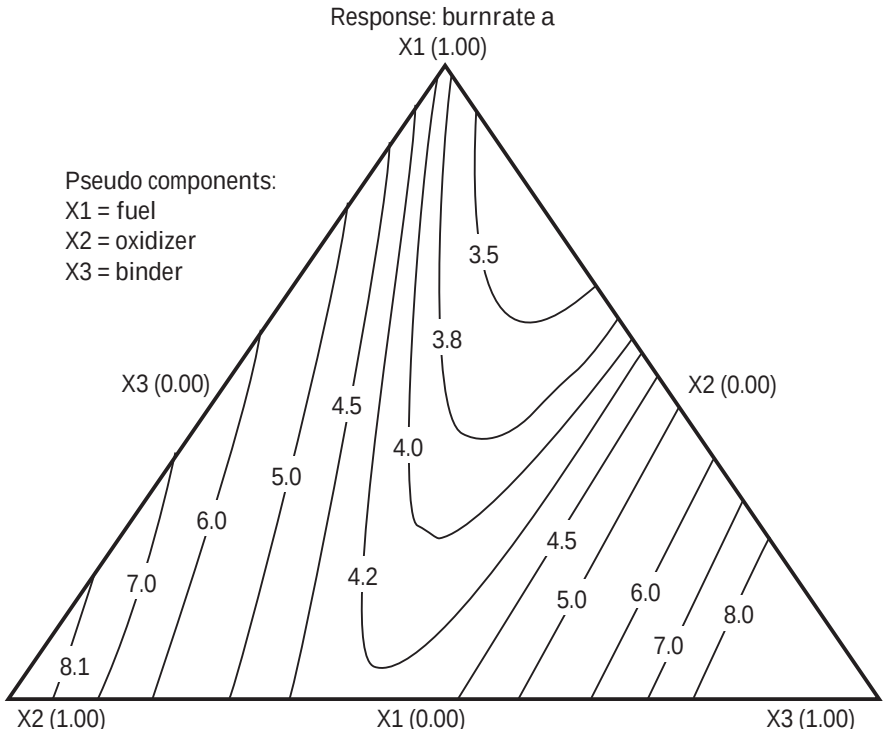


Figure 67 — Response surface contours for *standard deviation* of burn rate in terms of fuel (X1), oxidizer (X2) and binder (X3) blend components

These figures indicate technically feasible blends which satisfy the product formulation performance constraints of >95 cm/sec mean burn rate and <4.5 cm/sec burn rate standard deviation. This is achieved, most simply, by superposition of one figure on the other.

The actual choice of formulation of fuel, oxidizer and binder can now be made, within the technical feasible zones, based on other appropriate criteria such as material cost and process productivity.

12.1.4.6 *Evolutionary operation (EVOP) designs*

Evolutionary operation (EVOP) comes into the experimentation scenario in two specific ways; one, as an extremely simple experimental design; and, two, as a very complex computer intensive numerical search technique, typically using genetic algorithms, to determine optimum solutions. The first is described here.

George Box [47] has compared the development of a process, product or service to evolution in nature. Living things advance by means of two mechanisms, mutation and natural selection. New species are produced by mutation. Similarly a major change in product design (e.g. reciprocating engine to gas turbine) or fundamental transformation in a manufacturing process (e.g. from manual to robots) constitute a mutation. Whereas in nature variants occur spontaneously, in industry variants need to be introduced through changes in level of factors. The consequence is a process of natural selection in that unpromising combinations of factor levels are neglected in favour of promising ones. This is the essence of EVOP. Basically the EVOP type experiment calls for replacing conventional in-line operation of a process, namely setting predetermined values for operation of significant process parameters and keeping the process in statistical control.

It requires the making of small changes (nudges or perturbations) in factor levels, noting the effect and then progressively making adjustments to improve performance. EVOP is an optimizing technique for routine use on industrial processes. One popular type of EVOP design, Box EVOP (due to Box) [47] is shown in Figure 68. The basic design combinations of the two factors, A and B, for the maximization experiment shown in Figure 68, are the corners of a square with the addition of a centre point representing their nominal values.

Suppose it is wished to improve the yield of a batch related process. There are two process parameters which are thought to have a major influence on the yield, A and B. Currently the process is run under fixed operating conditions which are considered standard. Both process parameters are currently set at their nominal values to give a process yield of 68 % as shown in Figure 68a).

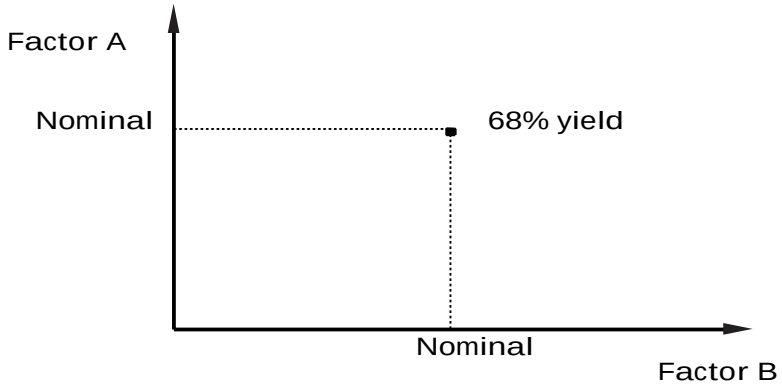


Figure 68a) — Factors A and B set at nominal to give a process yield of 68 %

Figure 68b) shows the first stage of a Box EVOP optimization process. Factors A and B are nudged or tweaked simultaneously to a simple predetermined factorial experimental design. Such a two factor design is represented by the corners of a square together with a centre point. In applying these designs in production it should be appreciated that runs are repeated at each of the experimental combinations according to a statistical methodology that provides for testing for significant differences between the results achieved. Such methodology indicates when to make a decision at that particular iteration and the preferable course of action, namely, to continue in a preferred direction or to stop because a local optimum has been reached.

Average yields under a first stage iteration are shown in Figure 68b).

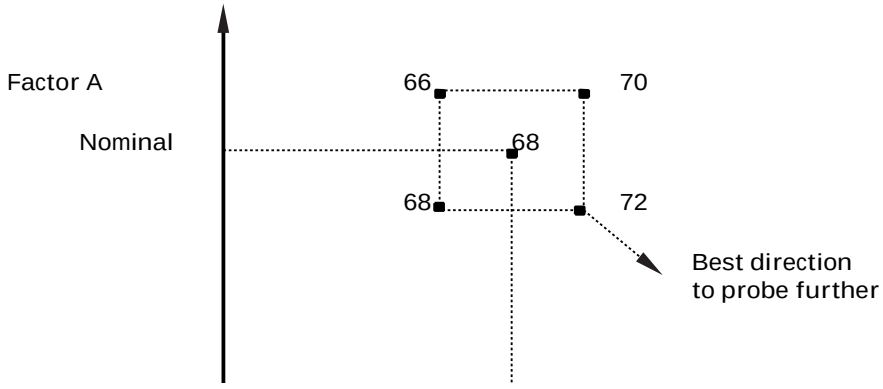


Figure 68b) — First stage optimization using Box EVOP

Figure 68b) shows a statistically significant improvement of 72 % in the lower right hand corner. This indicates the best direction to probe further in a second stage iteration.

This process continues until the yield at the centre point is significantly higher than that at the corners at which stage a local optimum is indicated as shown in Figure 68 c).

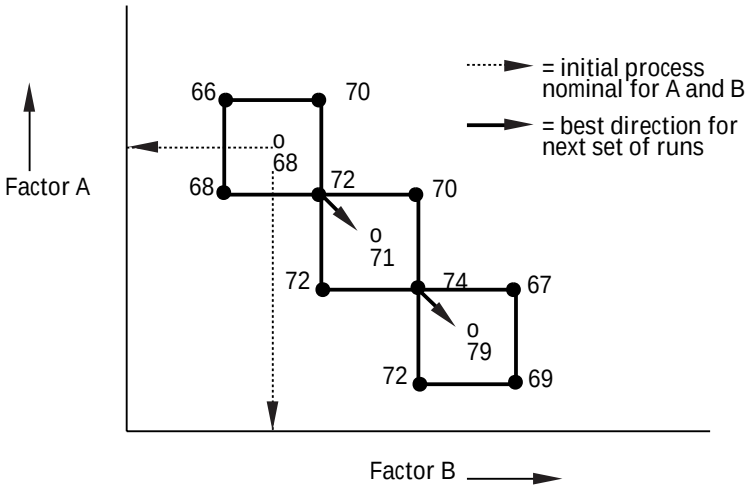


Figure 68c) — Final stage Box EVOP as local optimum has been found

In this case it is seen that the improvement in process yield is substantial, rising from 68 % to 79 %.

However, the consequences of possible temporary degradation of the process during such experimentation needs prior consideration. Such risks can usually be contained within EVOP by nudging the process, in small steps, towards an optimum operating level.

Alternative designs are available to Box EVOP, such as Simplex EVOP (due to Spendley) [48]. A simplex is the most elemental geometric figure. The Simplex EVOP experimental configuration for:

a) two factor designs is an equilateral triangle;

b) three factor designs is a regular tetrahedron.

Thus with Simplex EVOP, after the first iteration with three experimental points, only one further experimental point is required for each subsequent run. It is thus more responsive and simpler to apply than Box EVOP. This has considerable merit in a production situation.

After each run (with repeats as necessary) the point in the simplex which exhibits the poorest response is replaced by its reflection. This forms a new simplex.

Example

A Simplex EVOP maximization example with two factors is shown in Figure 69.

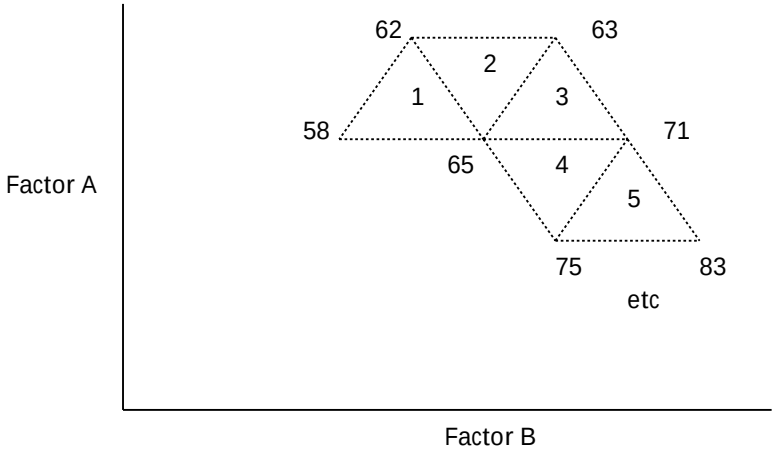


Figure 69 — Incomplete 5 stage simplex maximization experiment for two factors in terms of yield

It is seen that the first set of experimental runs are with factors A and B, set at the vertexes of the two factor simplex (an equilateral triangle) labelled 1. This first run gave yields of 58 %, 62 % and 65 %, as shown in the appropriate vertex.

The second run is made at a vertex formed by a reflection of the vertex that gives the worst response on the previous set of runs. The vertex, so formed, in the new simplex (labelled 2), creates a new pair of values for the factors A and B being investigated. This run gives a yield of 63 %.

Similarly a third run yields 71 %, a fourth run 75 % and a fifth run 83 %, and so on.

13 Measurement systems

13.1 Measurements and standards

Measurement involves comparison with an accepted standard. From time immemorial such standards have existed. The standard for length, the inch, was once the "width of a man's thumb" and later "three barley corns, dry and round, laid end to end". More recently there was a prototype metre, a platinum iridium alloy bar with lines inscribed one metre apart. Accurate but not true replicas were distributed as reference standards. These were subject to deterioration. This led to the present day definition of a metre in unchanging fundamental terms, rather than artefacts, as "the length of the path travelled by light in a vacuum during the time interval of 1/299 792 458 of a second". Today, measurements are largely conducted within the framework of the International System of Units (SI) adopted and recommended by the General Conference on Weights and Measures.

In order to achieve and maintain measurement integrity it is standard practice today that measuring equipment be *calibrated* through a hierarchy of intermediate measurement standards traceable to the accepted primary standard. Where such a standard does not exist, *traceability* is established to other measurement standards, such as to reference materials. An example is sapphire as a heat capacity calibrant in calorimetry.

The overall *traceability* process typically consists of several stages, descending from the international standard firstly through a national standard to a reference standard and then to a working standard. At each of these stages of measurement transfer it is recommended that at least a 4:1 uncertainty ratio and preferably a 10:1 uncertainty ratio be maintained. This means, for example, that the dimension of a plastic part having a tolerance (or process spread) of $\pm 0.01''$ would need to be measured by calipers calibrated to $\pm 0.001''$, which, in turn, would be calibrated against gauge blocks calibrated to $\pm 0.000\,1''$, and so on.

13.2 Measurements, quality and statistics

The aim of any measurement is to determine the true value of a characteristic. This however is achieved only with an *ideal* measurement system: one that rarely, if ever, exists. *Actual* measurement systems exhibit less desirable statistical properties.

Certain properties of a measurement system affect the quality of the result and hence need *metrological confirmation* (i.e. a set of operations required to ensure that an item of measuring equipment is in a state of compliance with the requirements for its intended use [ISO 10012]) These include:

a) *accuracy*: closeness of the agreement between a test result and the accepted reference value (ISO 3534-2). Accuracy involves a combination of random components and a systematic, bias, component;

b) *resolution*: smallest difference between indications of a displaying device that can be meaningfully distinguished (VIM 5.12) [49];

c) *precision*: the closeness of agreement between independent test results obtained under stipulated conditions (ISO 3534-2). Precision depends only on the distribution of random errors. It does *not* relate to the true value;

d) *stability*: ability of a measuring instrument to maintain constant its metrological characteristics with time (VIM 5.14) [49];

e) *uncertainty*: parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the subject of measurement (VIM 3.9) [49].Uncertainty includes all components of variation in the measurement system including uncertainty due to sampling and those arising from systematic effects.

Having established what basic properties a measurement system should have, it is now necessary to clarify what is meant by *intended use* in order to establish acceptable numerical values for these properties for specific uses.

The ISO 9000 series quality management systems supporting standard ISO 10012 states that it is applicable "only to measuring equipment used as part of a measurement process *used to demonstrate compliance with specified requirements*".

The automotive sector QS 9000 supporting standard "*Measurement systems analysis*", on the other hand, makes the valid point that "intended use" includes that of quality monitoring and statistical process control and improvement. This application can put much more onerous demands on the measuring system, particularly where process capabilities are high relative to the specified tolerances for process parameters and product characteristics monitored. It brings out the need for measuring systems to be shown to:

1) be in a state of statistical control (so that no special cause variation is present and hence measurement system variation is due to common causes alone);

2) have a measurement system variability small compared with process variability (to enable the process to be controlled);

3) have a measurement system variability small compared with specification limits (to enable the correct decision to be made on acceptance for conformity or not);

4) possess small increments of measure (one-tenth or better) relative to the lesser of the process variability or the specified tolerance.

Subsequent analyses of measurement system properties are based on the more comprehensive intended use outlined, which includes the requirements for in-line quality monitoring and statistical process control. These are dealt with using simple graphical statistical methods as far as possible and are intended for practical application at working levels to meet the requirements of prescriptive quality and process performance based standards.

13.3 Examples of statistical methods to ensure quality of measured data

13.3.1 *Example 1: Resolution*

For a measuring system whose sole purpose is to demonstrate compliance with specified requirements it is recommended that resolution be of the order of one-tenth of specified tolerance. However, if it is used to improve understanding of process variation and for process control and needs to detect the presence of special causes, then resolution should preferably be of the order of one-tenth of process spread (based on 6 standard deviations). See Figure 70.

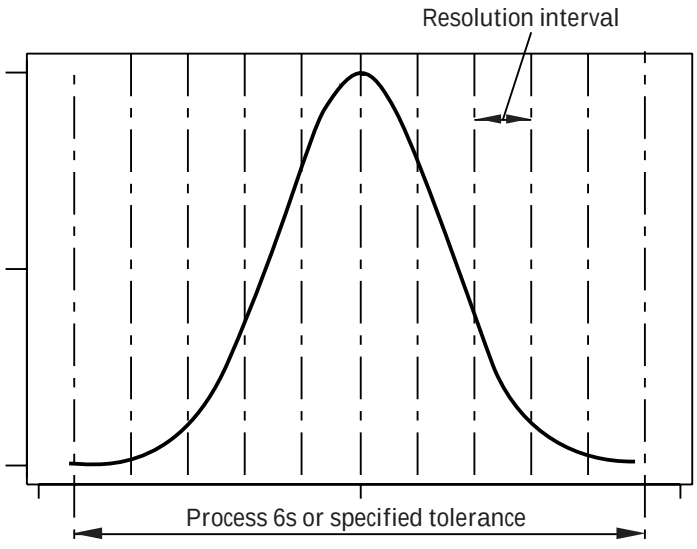


Figure 70 — Recommended resolution for process control and determination of compliance with specified tolerance

The "range" process control chart is a very good and simple indicator of the degree of resolution of a measuring system. Inadequate resolution is shown by a range chart with:

a) only 1, 2 or 3 possible values for the range within the control limits;

b) 4 possible values within the control limits with more than 25 % of the ranges showing zero.

This is illustrated in Figure 71.

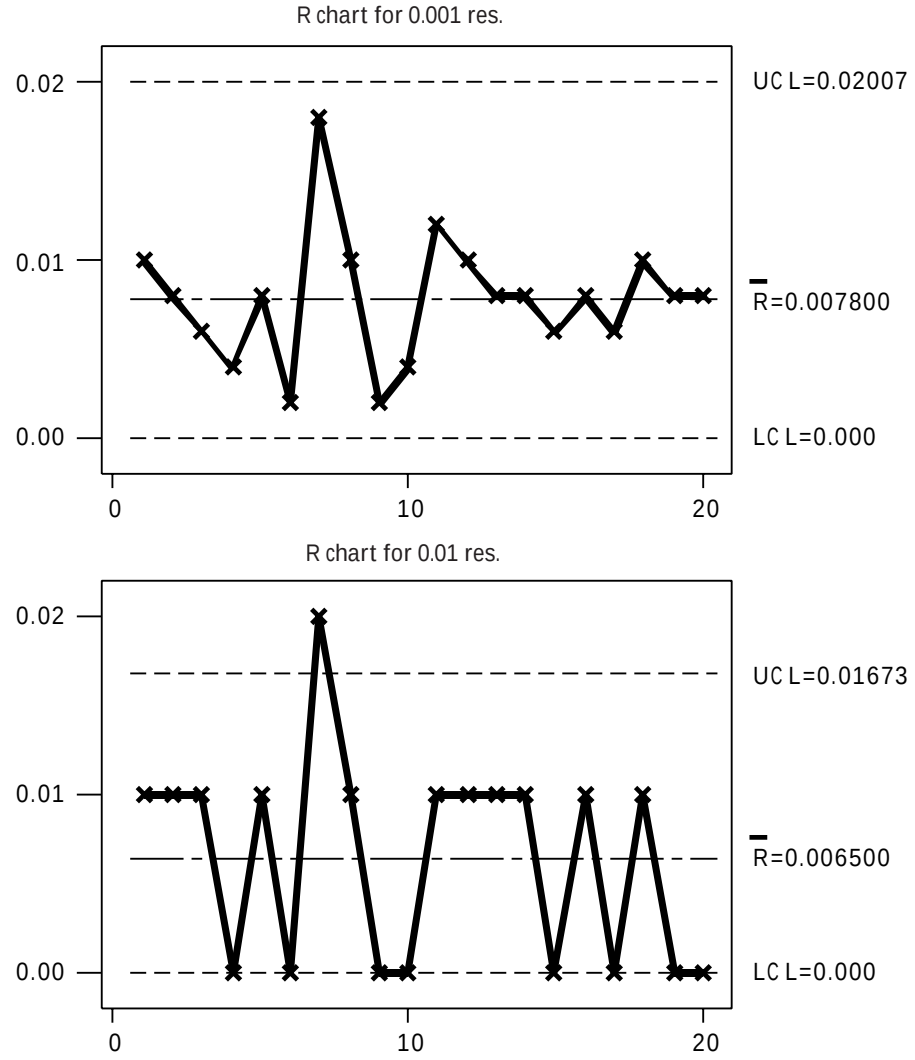


Figure 71 — Range chart showing adequate and inadequate resolutions

13.3.2 *Example 2: Bias and precision*

Aspects that need consideration in the assessment of bias and precision are now discussed.

a) Random and systematic errors

A measuring instrument may have one of two reasons for giving a reading that is inaccurate:

1) the instrument is out of calibration. Namely a series of readings made on a single unit gives an average which differs from the true value by an amount greater than that specified. This is a measure of the systematic error, termed *bias*;

2) irrespective of the state of calibration, the instrument will not give identical values when making a series of readings on a single unit. This is a measure of the random error, termed *precision*.

The difference between bias and precision is illustrated visually in Figure 72.

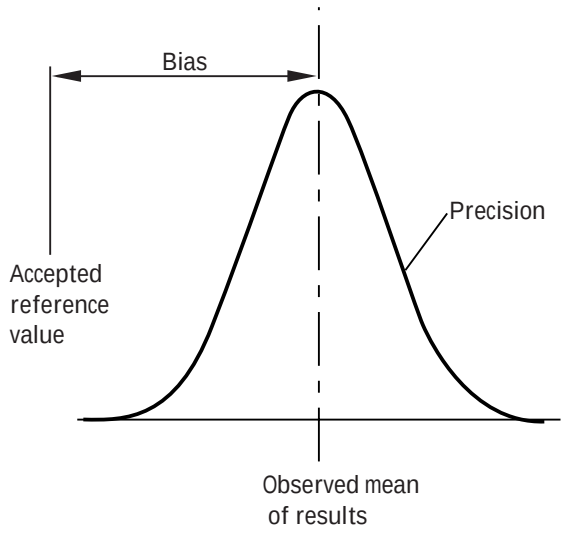


Figure 72 — Bias and precision

b) Bias

Bias is the difference between the expectation of a test result (observed mean of several measurements) and the accepted reference value.

c) Precision

Precision is the closeness of agreement between independent test results obtained under stipulated conditions.

d) Uncertainty

All measurements involve *uncertainty* to some degree.

The status of a measurement is quantified and qualified by its degree of uncertainty. There is a need to distinguish between precision and uncertainty. Uncertainty is a statement of the limits of the range within which the *true value* is expected to lie in relation to the measured result. Precision, on the other hand, relates only to the distribution of random errors and not to the true value, for example, systematic errors such as bias. Precision is a random component of uncertainty. The effect of uncertainty on a product compliance decision is illustrated in Figure 73.

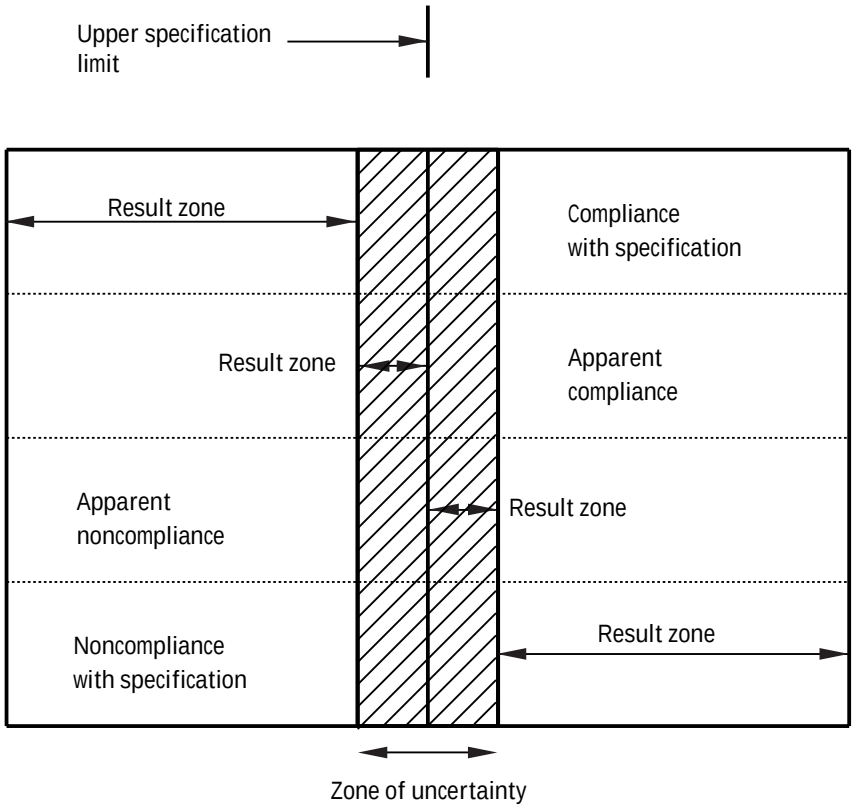


Figure 73 — Effect of measuring system uncertainty on compliance decision

Figure 73 illustrates that if the difference between a test result and a specification limit lies within the zone of measurement system uncertainty, the consequence is that nonconforming product might be accepted and conforming product rejected.

Example

A statutory requirement for a foodstuff is that it shall not contain more than 20 $\mu g \cdot g^{-1}$ of a particular constituent.

Suppose one manufacturer analyses a batch and obtains a result of 19 $\mu g \cdot g^{-1}$ for that constituent. His measurement system uncertainty is ± 0.5 $\mu g \cdot g^{-1}$. The true result is thus expected to lie within the range 18.5 to 19.5. Then it can be said that there is "compliance with the specification", the legal limit is not exceeded.

However if another manufacturer achieves an identical result but his measurement system uncertainty is ± 1.5 $\mu g \cdot g^{-1}$ then there is no such assurance of legal conformance as the true value is expected to lie within the range 17.5 to 20.5.

This shows the need for understanding of the contribution of uncertainty to product (and/or process) variation in relation to both compliance decision-making and in the setting of test specification limits in terms of customer or statutory requirements.

e) Simple graphical determination of bias and precision

If a measurement process is repeated under ostensibly constant conditions, a number of possibly different readings may be obtained. If one wishes to avoid extensive calculations and, at the same time, do both a visual check for normality and for any peculiarities in the readings, graphical analysis using normal probability paper is recommended. The method is illustrated by example.

Example

A pressure gauge is subject to an input pressure of 20.000 ± 0.001 in a constant environment. Repeated observations of the measuring instrument yielded the following indicated pressures:

20.30; 19.96; 20.14; 19.99; 20.24; 20.18; 20.20; 20.04; 20.11; 20.07.

Determine the status of the measuring instrument in terms of bias and 6 standard deviation (im)precision.

The outline solution is shown in Figure 74. In practice, one would need to similarly check at other key points of the operating range of the measuring instrument.

Bias, once determined, typically is corrected either by modifications to the measuring instrument itself or by making appropriate adjustment to each measuring instrument reading. Precision usually is expressed in terms of imprecision and computed as a multiple (often 6) of the standard deviation of the results. Less precision is reflected by a larger standard deviation. When such a measuring instrument is used for product acceptance purposes, the value of the measuring instrument six standard deviation (im)precision needs to be small in relation to the specified tolerance of the product characteristic that is subject to measurement. Where an instrument is used for process monitoring, the value of the instrument six standard deviation (im)precision needs to be small in relation to the six standard deviation (im)precision of the process characteristic subject to measurement.

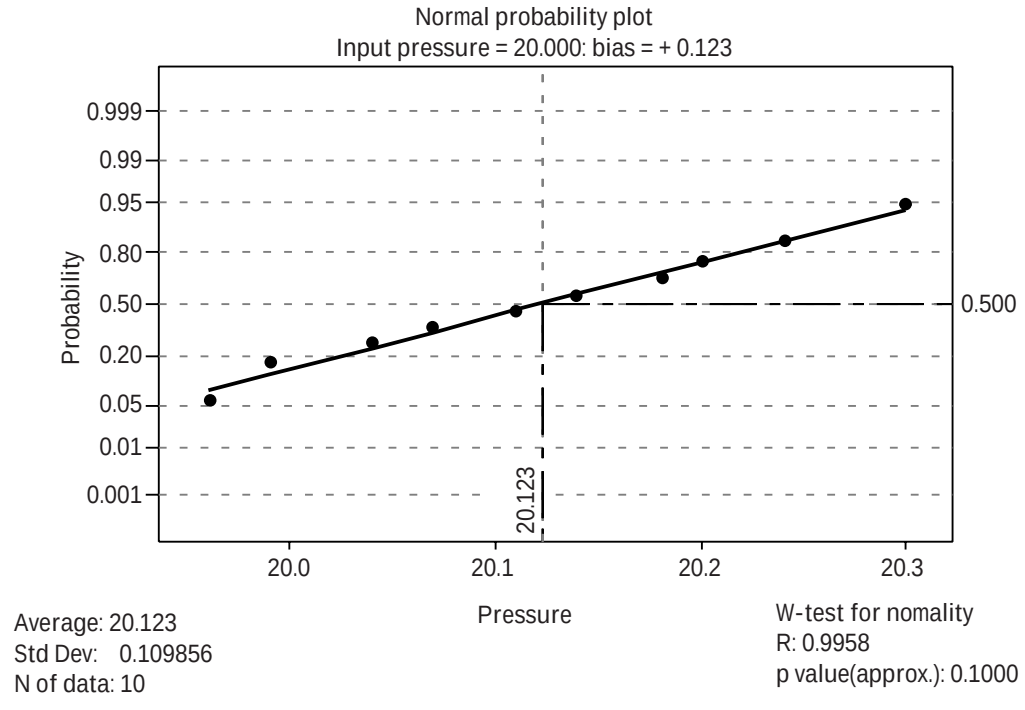


Figure 74 — Establishing bias and precision for a pressure gauge

13.3.3 *Precision — repeatability*

Repeatability is defined as precision under repeatability conditions. Repeatability condition is "observation condition where independent test results are obtained with the same method on identical test items in the same test facility by the same operator using the same equipment within short intervals of time" (BS ISO 3534-2).

The previous pressure gauge study would be looked upon as an analysis of repeatability. Of course, one should always check for data stability, for example, using an individuals and moving range chart as shown in Figure 75.

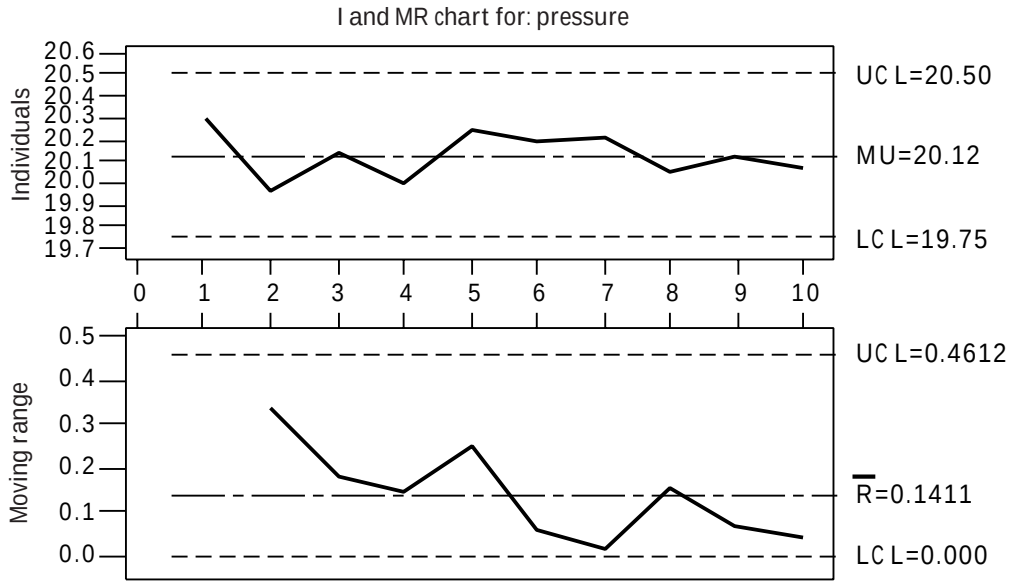


Figure 75 — Check for stability of results prior to performing a bias and precision test

The moving range chart in Figure 75 indicates a progressive downwards trend, which shows the possibility of weaknesses in experimental protocol in estimating repeatability.

13.3.4 ***Precision — reproducibility***

Reproducibility is precision under reproducibility conditions. Reproducibility condition is "observation condition where independent test results are obtained with the same method on identical test items in different test facilities with different operators using different equipment" (BS ISO 3534-2).

Often, precision is established under intermediate precision conditions, namely "a condition where test results are obtained with the same method, on identical test items in the same test facility, under some different operating condition". The different operating condition may take the form of time, calibration, operator and equipment. Precision, so determined, is called "intermediate precision measure".

Example of repeatability and reproducibility study

There are standard forms available for facilitating repeatability and reproducibility analysis at shop-floor level. One such type is now illustrated. They are so simple to use that one has to be careful not to switch off mentally when applying the various criteria to arrive at repeatability and reproducibility estimates. This is particularly important as there is usually a wealth of interesting aspects of variations in measurement systems that have not hitherto been revealed. These can provide a basis for significant improvements in measurement system precision. This will also be considered.

Three appraisers are involved, each performing three tests on each of ten specimens. The experimental protocol was to:

a) identify ten specimens for measurement representing the range of process variation. Number the parts 1–10, the markings not being visible to the appraisers;

b) appraiser A to measure each part (A1) and record the results, followed by appraiser B (B1) and then appraiser C (C1) using the same calibrated instrument;

c) repeat the cycle of measurements twice more until each appraiser has measured each part three times.

The values given in Table 36 are measured values, in microns from nominal: specification 56.388 ± 0.038.

Table 36 — Values obtained on repeatability and reproducibility study

Part	A1	A2	A3	R_A	B1	B2	B3	R_B	C1	C2	C3	R_C
1	18	12	8	10	18	10	6	12	12	8	8	4
2	18	14	10	8	18	8	8	10	12	8	10	4
3	16	12	10	6	16	8	8	8	14	8	10	6
4	12	8	10	4	16	8	8	8	12	8	10	4
5	16	8	10	8	14	8	8	6	10	8	10	2
6	12	10	10	2	16	6	8	8	10	8	10	2
7	16	10	8	8	12	8	6	6	12	8	8	4
8	18	10	8	10	16	8	6	10	12	8	10	4
9	12	12	10	2	14	8	6	8	8	8	10	2
10	14	12	10	4	14	6	8	8	8	8	10	2
Total	152	108	94	62	154	78	72	84	110	80	96	34
*$\overline{R}_A$ = 6.2; #$\overline{X}_A$ = 11.8					$\overline{R}_B$ = 8.4; $\overline{X}_B$ = 10.13				$\overline{R}_C$ = 3.4; $\overline{X}_C$ = 9.53			
$\overline{R} = (\overline{R}_A + \overline{R}_B + \overline{R}_C)/3 = 6.0$					$\overline{X}_{diff} = \overline{X}_{\max} - \overline{X}_{\min} = 2.27$							

* $\overline{R}_A$ = average R_A = (152 + 108 + 94)/30.

$\overline{X}_A$ = average of all A readings.

A1 = appraiser A trial 1; A2 = appraiser A trial 2; A3 = appraiser A trial 3; etc.

R_A = difference between largest and smallest reading over each three trials for each part for observer A, etc.

a) Test data for stability

Upper control limit for range = $D_4 \cdot \overline{R}$ = 2.58 × 6.0 = 15.48.

If any individual range exceeds this limit, the measurement should be reviewed, repeated, corrected or discarded, as appropriate and new averages and ranges calculated. If no individual range exceeds this limit it is all right to proceed with the calculations of repeatability and reproducibility.

Here, maximum range = 12, therefore it is all right to proceed.

b) Repeatability

Repeatability = $k_1 \cdot \overline{R}$ [for details of derivation of k_1 see e)]

= 3.54 × 6.0 = 21.24.

% repeatability = (100 × repeatability)/total specified tolerance

= 100 × 21.14/76 = 27.8 %

c) Reproducibility

Reproducibility $= \sqrt{(\overline{X}_{\text{diff}} \times k_2)^2 - \text{repeatability}^2/(nr)}$

$= \sqrt{[(2.27 \times 3.14)^2 - 21.24^2/10 \times 3]}$

= 5.98

where

n = number of parts;

r = number of trials.

% reproducibility = 100 × 5.98/76 = 7.9 %.

d) Repeatability and reproducibility

$$\text{Repeatability and reproducibility} = \sqrt{\text{repeatability}^2 + \text{reproducibility}^2}$$

$$= \sqrt{21.24^2 + 5.98^2}$$

$$= 22.1.$$

% repeatability and reproducibility = $100 \times 22.1/76 = 29$ %.

This overall % figure should now be compared with the specified requirements. For this particular application, measurement system requirements were expressed as:

0 to 20 %: good;

21 to 30 %: marginal;

over 30 %: unacceptable.

This measurement system is thus marginal, bordering on the unacceptable.

e) Derivation of k_1 and k_2

In statistical process control, the standard deviation is estimated from the average range using the formula:

$s = \overline{R}/d_2$

where d_2 is dependent on subgroup size.

This formula assumes that the ranges have been averaged over a large number of subgroups. This is usually the case when control charts are involved but rarely so with measurement system studies.

When only a few subgroups are involved, a better estimate of standard deviation is made using the formula:

$s = \overline{R}/d_2^*$

where d_2^* (pronounced dee-sub-two-star) is found from Table 37.

Table 37 — Tabulated values of d_2^* in terms of number in subgroup and number of subgroups

g	Number in subgroup													
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	1.41	1.91	2.24	2.48	2.67	2.83	2.96	3.08	3.18	3.27	3.35	3.42	3.49	3.55
2	1.28	1.81	2.15	2.40	2.60	2.77	2.91	3.02	3.13	3.22	3.30	3.38	3.45	3.51
3	1.23	1.77	2.12	2.38	2.58	2.75	2.89	3.01	3.11	3.21	3.29	3.37	3.43	3.50
4	1.21	1.75	2.11	2.37	2.57	2.74	2.88	3.00	3.10	3.20	3.28	3.36	3.43	3.49
5	1.19	1.74	2.10	2.36	2.56	2.73	2.87	2.99	3.10	3.19	3.28	3.35	3.42	3.49
6	1.18	1.73	2.09	2.35	2.56	2.73	2.87	2.99	3.10	3.19	3.27	3.35	3.42	3.49
7	1.17	1.73	2.09	2.35	2.55	2.72	2.87	2.99	3.10	3.19	3.27	3.35	3.42	3.48
8	1.17	1.72	2.08	2.35	2.55	2.72	2.87	2.98	3.09	3.19	3.27	3.35	3.42	3.48
9	1.16	1.72	2.08	2.34	2.55	2.72	2.86	2.98	3.09	3.18	3.27	3.35	3.42	3.48
10	1.16	1.72	2.08	2.34	2.55	2.72	2.86	2. 98	3.09	3.18	3.27	3.34	3.42	3.48
11	1.16	1.71	2.08	2.34	2.55	2.72	2.86	2. 98	3.09	3.18	3.27	3.34	3.41	3.48
12	1.15	1.71	2.07	2.34	2.55	2.72	2.85	2. 98	3.09	3.18	3.27	3.34	3.41	3.48
13	1.15	1.71	2.07	2.34	2.55	2.71	2.85	2. 98	3.09	3.18	3.27	3.34	3.41	3.48
14	1.15	1.71	2.07	2.34	2.54	2.71	2.85	2. 98	3.08	3.18	3.27	3.34	3.41	3.48
15	1.15	1.71	2.07	2.34	2.54	2.71	2.85	2. 98	3.08	3.18	3.26	3.34	3.41	3.48
>15	1.128	1.693	2.059	2.326	2.534	2.704	2.847	2.970	3.078	3.173	3.258	3.336	3.407	3.472
g = number of subgroups.														

Reference: Reproduced from Duncan [50].

Derivation of k

The first decision to be made in deriving k_1 and k_2 is what multiple of standard deviation is to be used? Typical values are 6 (embracing 99.73 %) and 5.15 (embracing 99 %). 6 will be used here.

The next decision is to determine the appropriate value of d_2* to be used.

k_1: repeatability

With 3 trials, the number within a subgroup (to work out R) is 3 and the number of subgroups is the number of appraisers times the number of specimens (= 30). Hence, from the table, d_2* = 1.693. Hence k_1 = 6/1.693 = 3.54.

k_2: reproducibility

With 3 appraisers the number within a subgroup (to work out $\overline{X}_{diff}$) is 3. The number of subgroups is simply one. Hence, from the table, d_2* = 1.91. Hence k_2 = 6/1.91 = 3.14.

f) Analysis and improvement

The largest contributor to variation in the measurement system is seen to be repeatability. The test for control in a) indicated that it was appropriate to treat the data as stable. However this is a relatively crude test. The ranges for the various appraisers in Figure 76 show patterns and considerable variation.

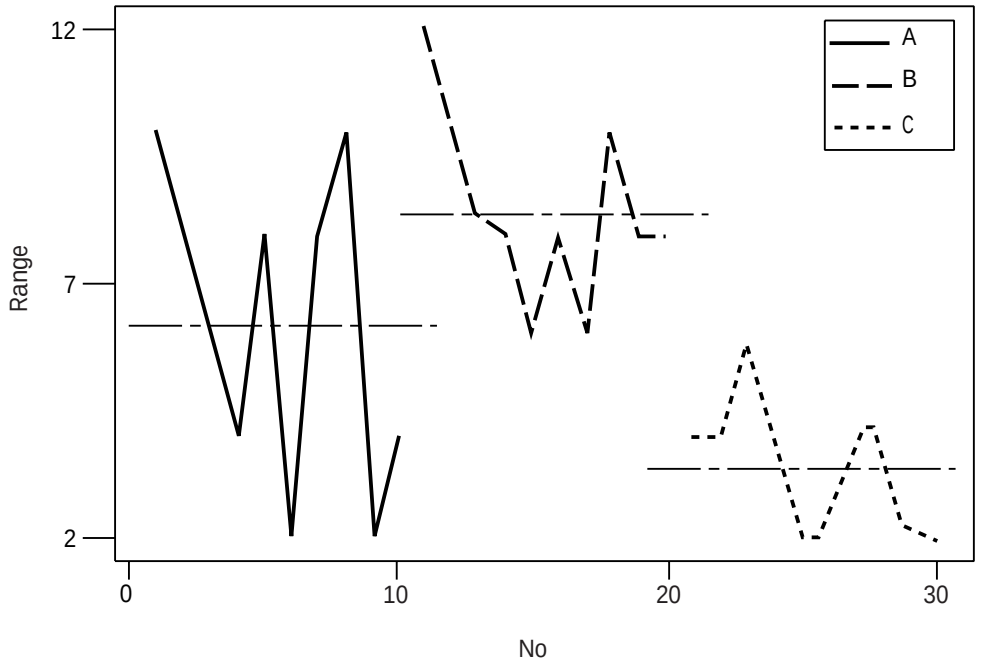


Figure 76 — Variation in repeatability within and between appraisers

Analysis and exploitation of these patterns and variation by seeking out and standardizing on best measuring practice used by the appraisers should yield significant improvements in measured system performance in this particular application.

Annex A (informative)
Measured data control charts: Formulae and constants

Table A.1 — Control limits constants in terms of subgroup size (n) for mean, median and individual control charts based on range

Subgroup size	Mean $\overline{X}$	Median $\tilde{X}$	Individual X	Range	
n	A_2	$\tilde{A}_2$	E_2	D_3	D_4
2	1.88	—	2.66	0	3.27
3	1.02	1.19	1.77	0	2.57
4	0.73	—	1.46	0	2.28
5	0.58	0.69	1.29	0	2.11
6	0.48	—	1.18	0	2.00
7	0.42	0.51	1.11	0.08	1.92
8	0.37	—	1.05	0.14	1.86
9	0.34	0.41	1.01	0.18	1.82
10	0.31	—	0.98	0.22	1.78

NOTE 1 Subgroup sizes greater than 10 are not recommended for range-based control charts due to the loss of information with using just the maximum and minimum in a subgroup.

NOTE 2 For individual and moving range control charts, where only one sample is taken at a time, the subgroup size is based on the number of ranges taken to constitute the moving range chart. This is often 2 or 3.

NOTE 3 Constants are not given for even numbers for the median chart due to the additional complexity for calculation operationally. With odd numbers the mid value is chosen.

NOTE 4 The appropriate formulae to use with these constants are given in Table A.2.

Table A.2 — Formulae for constructing control limits for mean, median and individual control charts based on range

Mean chart	$UCL_{\overline{X}} = \overline{\overline{X}} + A_2\overline{R}$
	$LCL_{\overline{X}} = \overline{\overline{X}} - A_2\overline{R}$
Median chart	$UCL_{\tilde{X}} = \overline{\tilde{X}} + \tilde{A}_2\overline{R}$
	$LCL_{\tilde{X}} = \overline{\tilde{X}} - \tilde{A}_2\overline{R}$
Individual chart	$UCL_X = \overline{X} + E_2\overline{R}$
	$LCL_X = \overline{X} - E_2\overline{R}$
Range chart	$UCL_R = D_4\overline{R}$
	$LCL_R = D_3\overline{R}$

Table A.3 — Control limit constants in terms of subgroup size (n) for mean charts based on standard deviations

Subgroup size	Mean $\bar{X}$	Standard deviation s	
n	A_3	B_3	B_4
2	2.66	0	3.27
3	1.95	0	2.57
4	1.63	0	2.27
5	1.43	0	2.09
6	1.29	0.03	1.97
7	1.18	0.12	1.88
8	1.10	0.19	1.82
9	1.03	0.24	1.76
10	0.98	0.28	1.72
11	0.93	0.32	1.68
12	0.89	0.35	1.65
13	0.85	0.38	1.62
14	0.82	0.41	1.59
15	0.79	0.43	1.57
16	0.76	0.45	1.55
17	0.74	0.47	1.53
18	0.72	0.48	1.52
19	0.70	0.50	1.50
20	0.68	0.51	1.49
21	0.66	0.52	1.48
22	0.65	0.53	1.47
23	0.63	0.55	1.46
24	0.62	0.56	1.45
25	0.61	0.57	1.44

NOTE The appropriate formulae to use with these constants are given in Table A.4.

Table A.4 — Formulae for constructing control limits for mean control charts based on the standard deviation

Mean chart	$UCL_{\bar{X}} = \bar{\bar{X}} + A_3\bar{s}$
	$LCL_{\bar{X}} = \bar{\bar{X}} - A_3\bar{s}$
Standard deviation chart	$UCL_s = B_4\bar{s}$
	$LCL_s = B_3\bar{s}$

NOTE $\bar{s}$ is the average of the individual subgroup sample standard deviations.

Table A.5 — Formulae for centre-lines for standard measured data control charts

Mean	$\overline{\overline{X}} = \frac{(\overline{X}_1 + \overline{X}_2 + \overline{X}_3 + ...\overline{X}_k)}{k}$
Median	$\overline{\tilde{X}} = \frac{(\tilde{X}_1 + \tilde{X}_2 + \tilde{X}_3 + ...\tilde{X}_k)}{k}$
Individual	$\overline{X} = \frac{(X_1 + X_2 + X_3 + ...X_m)}{m}$
Range	$\overline{R} = \frac{(R_1 + R_2 + R_3 + ...R_k)}{k}$
Standard deviation	$\overline{s} = \frac{(s_1 + s_2 + s_3 + ...s_k)}{k}$

NOTE 1 k is the number of subgroups.

NOTE 2 m is the number of individual values in an individuals chart.

Annex B (informative)
Percentage points of the *t*-distribution

v	Q = 0.25	Q = 0.1	Q = 0.05	Q = 0.025	Q = 0.01	Q = 0.005	Q = 0.002 5	Q = 0.001	Q = 0.000 5
	2Q = 0.5	2Q = 0.2	2Q = 0.1	2Q = 0.05	2Q = 0.02	2Q = 0.01	2Q = 0.005	2Q = 0.002	2Q = 0.001
1	1.000 0	3.077 7	6.313 8	12.706 2	31.820 5	63.656 7	127.321 3	318.308 8	636.619 2
2	0.816 5	1.885 6	2.920 0	4.302 7	6.964 6	9.924 8	14.089 0	22.327 1	31.599 1
3	0.764 9	1.637 7	2.353 4	3.182 4	4.540 7	5.840 9	7.453 3	10.214 5	12.924 0
4	0.740 7	1.533 2	2.131 8	2.776 4	3.746 9	4.604 1	5.597 6	7.173 2	8.610 3
5	0.726 7	1.475 9	2.015 0	2.570 6	3.364 9	4.032 1	4.773 3	5.893 4	6.868 8
6	0.717 6	1.439 8	1.943 2	2.446 9	3.142 7	3.707 4	4.316 8	5.207 6	5.958 8
7	0.711 1	1.414 9	1.894 6	2.364 6	2.998 0	3.499 5	4.029 3	4.785 3	5.407 9
8	0.706 4	1.396 8	1.859 5	2.306 0	2.896 5	3.355 4	3.832 5	4.500 8	5.041 3
9	0.702 7	1.383 0	1.833 1	2.262 2	2.821 4	3.249 8	3.689 7	4.296 8	4.780 9
10	0.699 8	1.372 2	1.812 5	2.228 1	2.763 8	3.169 3	3.581 4	4.143 7	4.586 9
11	0.697 4	1.363 4	1.795 9	2.201 0	2.718 1	3.105 8	3.496 6	4.024 7	4.437 0
12	0.695 5	1.356 2	1.782 3	2.178 8	2.681 0	3.054 5	3.428 4	3.929 6	4.317 8
13	0.693 8	1.350 2	1.770 9	2.160 4	2.650 3	3.012 3	3.372 5	3.852 0	4.220 8
14	0.692 4	1.345 0	1.761 3	2.144 8	2.624 5	2.976 8	3.325 7	3.787 4	4.140 5
15	0.691 2	1.340 6	1.753 1	2.131 4	2.602 5	2.946 7	3.286 0	3.732 8	4.072 8
16	0.690 1	1.336 8	1.745 9	2.119 9	2.583 5	2.920 8	3.252 0	3.686 2	4.015 0
17	0.689 2	1.333 4	1.739 6	2.109 8	2.566 9	2.898 2	3.222 4	3.645 8	3.965 1
18	0.688 4	1.330 4	1.734 1	2.100 9	2.552 4	2.878 4	3.196 6	3.610 5	3.921 6
19	0.687 6	1.327 7	1.729 1	2.093 0	2.539 5	2.860 9	3.173 7	3.579 4	3.883 4
20	0.687 0	1.325 3	1.724 7	2.086 0	2.528 0	2.845 3	3.153 4	3.551 8	3.849 5
21	0.686 4	1.323 2	1.720 7	2.079 6	2.517 6	2.831 4	3.135 2	3.527 2	3.819 3
22	0.685 8	1.321 2	1.717 1	2.073 9	2.508 3	2.818 8	3.118 8	3.505 0	3.792 1
23	0.685 3	1.319 5	1.713 9	2.068 7	2.499 9	2.807 3	3.104 0	3.485 0	3.767 6
24	0.684 8	1.317 8	1.710 9	2.063 9	2.492 2	2.796 9	3.090 5	3.466 8	3.745 4
25	0.684 4	1.316 3	1.708 1	2.059 5	2.485 1	2.787 4	3.078 2	3.450 2	3.725 1
26	0.684 0	1.315 0	1.705 6	2.055 5	2.478 6	2.778 7	3.066 9	3.435 0	3.706 6
27	0.683 7	1.313 7	1.703 3	2.051 8	2.472 7	2.770 7	3.056 5	3.421 0	3.689 6
28	0.683 4	1.312 5	1.701 1	2.048 4	2.467 1	2.763 3	3.046 9	3.408 2	3.673 9
29	0.683 0	1.311 4	1.699 1	2.045 2	2.462 0	2.756 4	3.038 0	3.396 2	3.659 4
30	0.682 8	1.310 4	1.697 3	2.042 3	2.457 3	2.750 0	3.029 8	3.385 2	3.646 0
40	0.680 7	1.303 1	1.683 9	2.021 1	2.423 3	2.704 5	2.971 2	3.306 9	3.551 0
60	0.678 6	1.295 8	1.670 6	2.000 3	2.390 1	2.660 3	2.914 6	3.231 7	3.460 2
120	0.676 5	1.288 6	1.657 7	1.979 9	2.357 8	2.617 4	2.859 9	3.159 5	3.373 5
∞	0.674 5	1.281 6	1.644 9	1.960 0	2.326 3	2.575 8	2.807 0	3.090 2	3.290 5

NOTE Q is the upper tail area of the distribution for v degrees of freedom, for use in a single-tailed test. For a two-tailed test, use 2Q.

Bibliography

Standards publications

BS 2846-1:1991, *Guide to statistical interpretation of data — Part 1: Routine analysis of quantitative data.*

BS 2846-2:1981, *Guide to statistical interpretation of data — Part 2: Estimation of the mean: confidence interval.*
[ISO 2602]

BS 2846-3:1975, *Guide to statistical interpretation of data — Part 3: Determination of a statistical tolerance interval.*
[ISO 3207]

BS 2846-4:1976, *Guide to statistical interpretation of data — Part 4: Techniques of estimation and tests relating to means and variances.*
[ISO 2854]

BS 2846-5:1977, *Guide to statistical interpretation of data — Part 5: Power of tests relating to means and variances.*
[ISO 3494]

BS 2846-6:1976, *Guide to statistical interpretation of data — Part 6: Comparison of two means in the case of paired observations.*
[ISO 3301]

BS 2846-7:1997, *Guide to statistical interpretation of data — Part 7: Tests for departure from normality.*
[ISO 5479]

BS 2846-8:1995, *Guide to statistical interpretation of data — Part 8: Estimation of a median.*
[ISO 8595]

BS 5660-2, *Methods of sampling iron ores — Part 2: Mechanical method of increment sampling and sample preparation.*
[ISO 3082]

BS 5662-1, *Evaluation of sampling methods for iron ores — Part 1: Experimental methods for evaluation of quality variation.*
[ISO 3084]

BS 5700, *Guide to process control using quality control chart methods and cusum techniques.*

BS 5701, *Guide to number-defective charts for quality control.*

BS 5702, *Guide to statistical process control charts for variables.*

BS 5703-1, *Guide to data analysis and quality control using cusum techniques — Part 1: Introduction to cusum charting.*

BS 5703-2, *Guide to data analysis and quality control using cusum techniques – Part 2: Decision rules and statistical tests for cusum charts and tabulations.*

BS 5703-3, *Guide to data analysis and quality control using cusum techniques – Part 3: Cusum methods for process/quality control by measurement.*

BS 5703-4, *Guide to data analysis and quality control using cusum techniques — Part 4: Cusums for counted/attributes data.*

BS 6001-0, *Sampling procedures for inspection by attributes — Part 0: Introduction to the BS 6001 attribute sampling system.*

BS 6001-1:1999, *Sampling procedures for inspection by attributes — Part 1: Specification for sampling plans indexed by acceptable quality level (AQL) for lot-by-lot inspection.*
[ISO 2859-1:1999]

BS 6001-2:1993, *Sampling procedures for inspection by attributes — Part 2: Specification for sampling plans indexed by limiting quality (LQ) for isolated lot inspection.*
[ISO 2859-2:1985]

BS 6001-3:1993, *Sampling procedures for inspection by attributes — Part 3: Specification for skip-lot procedures.* [ISO 2859-3:1991]

BS 6001-5:2000, *Sampling procedures for inspection by attributes — Part 5: Procedures for assessment of stated quality levels.*

BS 6002-1:1993, *Sampling procedures for inspection by variables — Part 1: Specification for single sampling plans indexed by acceptable quality level (AQL) for lot-by-lot inspection.* [ISO 3951:1989]

BS 6002-4.1, *Sampling procedures for inspection by variables — Part 4: Specification for sequential sampling plans for percent nonconforming — Section 4.1: Known standard deviation.* [ISO 8423]

BS 7782:1994, *Control charts — General guide and introduction.* [ISO 7870:1993]

BS 7783:1994, *Acceptance control charts.* [ISO 7966:1993]

BS 7785:1994, *Shewhart control charts.* [ISO 8258:1991]

BS EN 30012-1, *Quality assurance requirements for measuring equipment — Part 1: Metrological confirmation system for measuring equipment.*[ISO 10012-1]

BS EN ISO 9001, *Quality systems — Model for quality assurance in design, development, production, installation and servicing.*

BS EN ISO 9004-1, *Quality management and quality assurance standards — Part 1: Guidelines.*

BS ISO 3534-2, *Statistics. Vocabulary and symbols — Part 2: Statistical quality control.*

BS ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions.*

BS ISO 5725-2, *Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic methods for the determination of repeatability and reproducibility of a standard measurement method.*

BS ISO 5725-3, *Accuracy (trueness and precision) of measurement methods and results — Part 3: Intermediate measures of the precision of a standard measurement method.*

BS ISO 5725-4, *Accuracy (trueness and precision) of measurement methods and results — Part 4: Basic methods for the determination of the trueness of a standard measurement method.*

BS ISO 5725-5, *Accuracy (trueness and precision) of measurement methods and results — Part 5: Alternative methods for the determination of the precision of a standard measurement method.*

BS ISO 5725-6, *Accuracy (trueness and precision) of measurement methods and results — Part 6: Use in practice of accuracy values.*

BS ISO 10012-2, *Quality assurance for measuring equipment — Part 2: Guidelines for control of measurement processes.* [ISO 10012-2]

prEN 12326-1, *Slate and stone products for discontinuous roofing and cladding — Part 1: Product specification.*

prEN 12603, *Glass in building — Determination of the bending strength of glass — Procedures for goodness of fit and confidence levels for Weibull distributed data.*

IEC 61649:1997, *Goodness-of-fit tests, confidence intervals and lower confidence limits for Weibull distributed data.*

ISO 2859-4, *Sampling procedures for inspection by attributes — Part 4: Procedures for assessment of stated quality levels.*

ISO 7873, *Control charts for arithmetic average with warning limits.*

ISO 10725, *Acceptance sampling plans and procedures for inspection of bulk materials.* (under preparation)

ISO 11453, *Statistical interpretation of data — Tests and confidence intervals relating to proportions.*

ISO 11648-1, *Statistical aspects of sampling from bulk materials — Part 1: General introduction.*

ISO 11648-2, *Statistical aspects of sampling from bulk materials — Part 2: Sampling of particulate materials.*

ISO 11648, *Statistical aspects of sampling from bulk materials* (the following parts are in preparation):

Part 3: Sampling of liquids.

Part 4: Sampling of gases.

ISO 16269-7, *Statistical interpretation of data — Part 7: Median: Estimation and confidence intervals.*

ISO 16269, *Statistical interpretation of data* (the following parts are in preparation):

Part 3: Tests for departure from normality.

Part 4: Detection and treatment of outliers.

Part 5: Estimation and tests for means and variances for the normal distribution, with power functions for tests.

Part 6: Determination of statistical tolerance intervals.

ISO/TR 7871, *Cumulative sum charts — Guidance on quality control and data analysis using CUSUM techniques.*

ISO/TR 10014, *Guidelines for managing the economics of quality.*

ISO/TR 10017, *Guidance on statistical techniques for ISO 9001:1994.*

ISO/TR 13425, *Guide for the selection of statistical methods in standardization and specification.*

MIL-STD-1235A (1974), *Single and multi-level continuous sampling procedures and tables for inspection by attributes.* US Government Printing Office, Washington, D.C.

MIL-STD-1916 (1996), *DOD Preferred methods for acceptance of product.* US Department of Defense Test Method Standard. US Government Printing Office, Washington, D.C.

Other documents

[1] SHEWHART, W. A. *Economic control of manufactured products*. London: MacMillan Ltd., 1931.

[2] SHEWHART, W. A. *Economic control of quality of manufactured product*, Reinhold Company, Princeton, N.J., 1931.

[3] KANAGAWA, A., ARIZONO, I., and OHTA, H. Design of the $(\bar{x}, s)$ control chart based on Kullback-Leibler Information. *In:* Lenz, H.-J. et al, eds. *Frontiers in Statistical Quality Control* 5, 1997. Heidelberg, Physica-Verlag, 183–192.

[4] MOLENAAR, W. *Approximations to the Poisson, binomial and hypergeometric distribution functions*. Mathematisch Centrum, Amsterdam, 1973.

[5] BLYTH, C. R. Approximate binomial confidence limits. *Journal of the American Statistical Association*. 1986, **81**, 843–855.

[6] WALTERS, D.E. In defence of the arc sine approximation. *The Statistician*. 1979, **28**, 219–222.

[7] HAHN, G. J. Factors for calculating two-sided prediction intervals for samples from a normal distribution. *Journal of the American Statistical Association*. 1969, **64**, 878–888.

[8] HAHN, G. J. Additional factors for calculating prediction intervals for samples from a normal distribution. *Journal of the American Statistical Association*. 1970, **65**, 1668–1676.

[9] HAHN, G. J., and NELSON, W. A survey of prediction intervals and their applications. *Journal of Quality Technology*. 1973, **5**, 178–188.

[10] HAHN, G. J., and MEEKER, W. Q. *Statistical Intervals — A Guide for Practitioners*. New York: John Wiley & Sons Inc., 1991.

[11] ODEH, R. E. and OWEN, D. B. *Tables for Normal Tolerance Limits, Sampling Plans and Screening*. New York: Marcel Dekker Inc., 1980.

[12] MANN, N.R., SCHAFER, E., and SINGPURWALLA, N. *Methods for statistical analysis of reliability and lifetime data*. New York: J. Wiley and Sons, 1974.

[13] LAWLESS, J. F. *Statistical models and methods for lifetime data*. New York: J. Wiley and Sons, 1982.

[14] MOOD, A. M. On the dependence of sampling inspection plans upon population distributions. *Annals of Mathematical Statistics*. 1943, **14**, 415–425.

[15] WETHERILL, G. B. and CHIU, W. K. A review of acceptance sampling schemes with emphasis on the economic aspect. *International Statistical Review*. 1975, **43**, 191–209.

[16] VON COLLANI, E. *The economic design of control charts*. Stuttgart: B.B Teubner, 1989.

[17] WALD, A. *Sequential analysis*. New York: Wiley, 1947.

[18] BAILLIE, D. H. Sequential sampling plans for inspection by attributes with near-nominal risks. *Fourth Conference of the Asia Pacific Quality Control Organization* (APQCO). Kuala Lumpur, Malaysia. 1994.

[19] BAILLIE, D. H. Curtailed sequential sampling plans for inspection by attributes with near-nominal risks. *International Symposium on Statistical Methods for Quality and Reliability — Recent trends in theory and practice*. Indian Association for Productivity, Quality and Reliability, Calcutta. 1997.

[20] DODGE, H. F. A sampling inspection plan for continuous production. *Annals of Mathematical Statistics*. 1943, **14**, 264–279.

[21] BEATTIE, D. W. A continuous acceptance sampling procedure based upon a cumulative sum chart for the number of defectives. *Applied Statistics*. 1962, **11**, 137–147.

[22] READ, D. R., and BEATTIE, D. W. The variable lot-size acceptance sampling plan for continuous production. *Applied Statistics*. 1961, **10**, 147.

[23] WADSWORTH, H. M., and WASSERMAN, G. S. A modified cusum procedure as a continuous sampling scheme. *In:* LENZ, H.-J., et al, eds. *Frontiers in Statistical Quality Control* 3. 1987, Physica Verlag, Heidelberg, 179–195.

[24] WASSERMAN, G. S. *Design of a Beattie procedure for continuous sampling or process surveillance*. Doctoral thesis, Department of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia. 1986.

[25] DODGE, H. F. Skip-lot sampling plan. *Industrial Quality Control*. 1955, **11**, 3–5.

[26] DODGE, H. F., and PERRY, R. L. A system of skip-lot plans for lot by lot inspection. *American Society for Quality Control Technical Conference Transactions*. Chicago, Illinois, 1971, 469–477.

[27] LIEBESMAN, B. S. and SAPERSTEIN, B. A proposed attribute skip-lot sampling program. *Journal of Quality Technology*. 1983, **15**, 130–139.

[28] LIEBESMAN, B. S. The development of an attribute skip-lot sampling standard. *In:* LENZ, H.-J., et al, eds. *Frontiers in Statistical Quality Control* 3. 1987, Physica Verlag, Heidelberg, 3–23.

[29] SQUEGLIA, N. L. *Zero acceptance number sampling schemes*. (4^{th} edition.) Milwaukee, Wisconsin: ASQC Quality Press, 1994.

[30] KLAASSEN, C. A. J. A bonus-malus system in statistical sampling for testing jewelry on fineness. *In: Proc. IPMI Seminar on Precious Metal Testing and Analysis*, San Antonio, Texas, 1996. International Precious Metals Institute, Allentown, Pennsylvania.

[31] BAILLIE, D. H. (1997) Normal approximations to the distribution function of the symmetric beta distribution. *In:* LENZ, H.-J., et al, eds. *Frontiers in Statistical Quality Control* 5, 1997, Physica Verlag, Heidelberg, 52–65.

[32] HAMAKER, H. C. *The construction of double sampling plans for variables*. ISO/TC 69/SC 5/WG 3 document N 28, International Organization for Standardization, Geneva, 1982.

[33] BAILLIE, D. H. Double sampling plans for inspection by variables when the process standard deviation is unknown. *Asia Pacific Journal of Quality Management*. 1992, **1**.

[34] BAILLIE, D. H., and KLAASSEN, C. A. J. Credit-based accept-zero sampling schemes for the control of outgoing quality. *In:* LENZ, H.-J., et al, eds. *Frontiers in Statistical Quality Control* 6. (to be published). Physica Verlag, Heidelberg.

[35] BAILLIE, D. H. Multivariate acceptance sampling. *In:* LENZ, H.-J., et al, eds. *Frontiers in Statistical Quality Control* 3. 1987. Physica Verlag, Heidelberg, 83–115.

[36] BAILLIE, D. H. Attriables acceptance sampling plans. *In:* LENZ, H.-J., et al, eds. *Frontiers in Statistical Quality Control* 4. 1992. Physica Verlag, Heidelberg, 3–33.

[37] BROWN, LOWE and BENHAM. *Fundamental statistical process control — Reference manual*. Chrysler, Ford and General Motors Corporations.

[38] DOWN, LOWE and DAUGHERTY. *Measurement system analysis — Reference manual*. Chrysler, Ford and General Motors Corporations.

[39] BOTHE, DAVIS, R. *SPC for short production runs*. Northville, Michigan. International Quality Institute, Inc.

[40] LI, H., OWEN, D. B., and BORREGO, S.A. Lower confidence limits on process capability indices based on the range. *Communications in Statistics Simulation and Computation*, **19**(1), 1–24.

[41] CHOU, Y.-M., OWEN, D. B., and BORREGO, S. A. Lower confidence limits on process capability indices. *Journal of Quality Technology*, **22**(3), 223–229.

[42] TAGUCHI, Genichi. *Introduction to quality engineering*, *Designing quality into products and processes*. White Plains, NY: Unipub.

[43] TAGUCHI, Genichi. *System of experimental design*, *Volumes 1 and 2*. White Plains, NY: Unipub.

[44] FISHER, R. A. *The design of experiments*. Edinburgh: Oliver and Boyd.

[45] CORNELL, J. A. *Experiments with mixtures; designs, models and the analysis of mixture data*. New York: John Wiley & Sons.

[46] MONTGOMERY, Douglas C., and VOTH, Sheila R. Multicollinearity and leverage in mixture experiments. *Journal of Quality Technology, Vol. 26*, No 2.

[47] BOX G., HUNTER, W. and HUNTER, J. S. *Statistics for experimenters*. New York: John Wiley & Sons.

[48] SPENDLEY, HEXT and HIMSWORTH. Sequential application of simplex designs in optimisation and evolutionary operation. *Technometrics*. 1962, **4**(4), 441–461.

[49] ISO. International vocabulary of basic and general terms used in metrology [VIM].

[50] DUNCAN, A. J. *Quality control and industrial statistics*. Homewood, Illinois: Irwin.

Relevant tables

DODGE, H.F., and ROMIG, H. G. *Sampling inspection tables, single and double sampling*, 2nd ed. New York: John Wiley, 1959.

FISHER, R. A., and YATES, F. *Statistical tables for biological, agricultural and medical research*, 6th edition. London: Oliver and Boyd, 1963.

OWEN, D. B. *Factors for one-sided tolerance limits and for variables sampling plans*. Monograph R-607, Sandia Corporation, 1963.

PEARSON, E.S., and HARTLEY, H. O. *Biometrika tables for statisticians*, Volume 1. Cambridge University Press, 1954.

Other reading

BENARD, A., and BOS-LEVENBACH, E.C. The plotting of observations on probability paper. 1953, *Statistica*, **7**, 163.

BLOM, G. *Statistical estimates and transformed beta-variables*. New York: Wiley, 1958.

CROSBY, P. *Quality is free: the art of making quality certain*. New York: McGraw-Hill, 1985.

DEMING, W. E. *Out of the crisis*. Cambridge, Mass.: Massachusetts Institute of Technology, Center for Advanced Engineering Study, 1986.

DODGE, H.F., and ROMIG, H. G. A method of sampling inspection, *The Bell System Technical Journa*l. 1929, **8**(10), 612–631.

DUNCAN, A. J. *Quality control and industrial statistics*. Homewood, Illinois: Irwin, 1986.

GROVE, D. M., and DAVIS, T. P. *Engineering, Quality and Experimental Design*. Harlow: Longman Group, 1992.

HALD, A. *Statistical theory of sampling inspection*, Part 1: Institute of Mathematical Statistics, University of Copenhagen, 1976.

HALD, A. *Statistical theory of sampling inspection*, Part 2: Institute of Mathematical Statistics, University of Copenhagen, 1978.

HASEMAN, J. K. Exact sample sizes for use with the Fisher-Irwin test for 2×2 tables. *Biometrics*. 1978, **34**, 106–109.

ISHIKAWA, K. *Introduction to quality control*. London: Chapman and Hall, 1989.

JURAN, J. M., and GRYNA, F. M., eds. *Juran's quality control handbook*, 4th edition. New York: McGraw-Hill, 1988.

KING, James R. *Probability charts for decision making*. New York: Industrial Press Inc.

MONTGOMERY, D. C. *Introduction to statistical quality control*. New York: Wiley, 1991.

RYAN, T. P. *Statistical methods for quality improvement*. New York: Wiley, 1989.

SCHILLING, E. G. *Acceptance sampling in quality control*. New York: Marcel Dekker, 1982.

THOMPSON, J. R., and KORONACKI, J. *Statistical Process Control for Quality Improvement*. New York: Chapman and Hall, 1993.

WADSWORTH, H. M., STEPHENS, K. S., and GODFREY, A. B. *Modern methods for quality control and improvement*. New York: Wiley, 1986.

WETHERILL, G.B., and BROWN, D. W. *Statistical process control — theory and practice*. London: Chapman and Hall, 1991.

blank

BSI — British Standards Institution

BSI is the independent national body responsible for preparing British Standards. It presents the UK view on standards in Europe and at the international level. It is incorporated by Royal Charter.

Revisions

British Standards are updated by amendment or revision. Users of British Standards should make sure that they possess the latest amendments or editions.

It is the constant aim of BSI to improve the quality of our products and services. We would be grateful if anyone finding an inaccuracy or ambiguity while using this British Standard would inform the Secretary of the technical committee responsible, the identity of which can be found on the inside front cover. Tel: 020 8996 9000. Fax: 020 8996 7400.

BSI offers members an individual updating service called PLUS which ensures that subscribers automatically receive the latest editions of standards.

Buying standards

Orders for all BSI, international and foreign standards publications should be addressed to Customer Services. Tel: 020 8996 9001. Fax: 020 8996 7001.

In response to orders for international standards, it is BSI policy to supply the BSI implementation of those that have been published as British Standards, unless otherwise requested.

Information on standards

BSI provides a wide range of information on national, European and international standards through its Library and its Technical Help to Exporters Service. Various BSI electronic information services are also available which give details on all its products and services. Contact the Information Centre. Tel: 020 8996 7111. Fax: 020 8996 7048.

Subscribing members of BSI are kept up to date with standards developments and receive substantial discounts on the purchase price of standards. For details of these and other benefits contact Membership Administration. Tel: 020 8996 7002. Fax: 020 8996 7001.

Copyright

BSI
389 Chiswick High Road
London
W4 4AL