

Rubber and rubber products — Guidance on the application of statistics to physical testing

ICS 83.060

National foreword

This British Standard is the UK implementation of ISO 19003:2006. It supersedes BS 903-2:1997 which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee PRI/22, Physical testing of rubber.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 28 February 2009
© BSI 2009

ISBN 978 0 580 54170 4

Amendments/corrigenda issued since publication

Date	Comments

INTERNATIONAL STANDARD

ISO 19003

First edition
2006-10-01

Rubber and rubber products — Guidance on the application of statistics to physical testing

*Caoutchouc et produits à base de caoutchouc — Lignes directrices
pour l'application des statistiques aux essais physiques*



Reference number
ISO 19003:2006(E)
Documents
Document Sharing Hub
© ISO 2006

PDF disclaimer

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

Adobe is a trademark of Adobe Systems Incorporated.

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

© ISO 2006

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

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

Published in Switzerland

Contents

Page

Foreword.....	vi
Introduction	vii
1 Scope	1
2 References.....	1
3 Terms and definitions.....	1
4 Symbols	4
5 Limitations of test results	5
5.1 Variability	5
5.2 Accuracy, trueness and precision	6
5.3 Relevance and significance	7
6 Distribution of results and measures of central tendency	7
6.1 Principles	7
6.2 Methodology.....	8
6.2.1 Types of distribution	8
6.2.2 Measures of central tendency	12
6.2.3 Measures of dispersion.....	15
6.2.4 Transformation to normal distribution	16
6.2.5 Test of departure from normality	17
6.3 Applications to rubber testing.....	20
6.3.1 General.....	20
6.3.2 Tensile testing	20
6.3.3 Fatigue	21
6.3.4 Conversion to normal distribution.....	24
6.3.5 Other uses of the median.....	25
7 Confidence limits and significant difference	26
7.1 Principles	26
7.2 Methodology.....	26
7.2.1 Confidence limits and confidence intervals	26
7.2.2 Significant difference	31
7.3 Applications to rubber testing.....	35
7.3.1 General.....	35
7.3.2 Confidence limits and specification limits	35
7.3.3 Comparison of results.....	36
8 Ranking methods	37
8.1 Principles	37
8.2 Methodology.....	37
8.2.1 Friedman's test	37
8.2.2 The outside count test.....	39
8.3 Applications to rubber testing.....	39
9 Criteria for rejecting outliers	40
9.1 Principles	40
9.2 Methodology.....	41
9.2.1 General.....	41
9.2.2 Dixon's test.....	41
9.2.3 Cochran's test for variance	43
9.3 Applications to rubber testing.....	45
9.3.1 General.....	45

9.3.2	Dixon's test applied to individual results	45
9.3.3	Cochran's variance test.....	46
9.3.4	Dixon's test applied to a group of mean values	47
10	Analysis of variance (ANOVA)	47
10.1	Principles	47
10.2	Methodology	48
10.2.1	General	48
10.2.2	One factor with an equal number of replicates	48
10.2.3	One factor with a variable number of replicates	49
10.2.4	Two (and over) factor analysis of variance	50
10.3	Applications to rubber testing	50
11	Regression analysis	54
11.1	Principles	54
11.2	Methodology	55
11.2.1	General	55
11.2.2	Linear least squares	55
11.2.3	Quadratic least squares	56
11.2.4	Cubic least squares	56
11.3	Applications to rubber testing	56
11.3.1	General	56
11.3.2	The effect of temperature on compression set.....	56
11.3.3	Effect of ageing on tensile strength.....	58
11.3.4	Temperature of retraction test.....	59
12	Uncertainty of measurement.....	60
12.1	Principles	60
12.2	Methodology	61
12.2.1	Compilation of a single value for uncertainty	61
12.2.2	Random uncertainty (U_r)	61
12.2.3	Systematic uncertainty (U_s)	62
12.2.4	Deviation of a single value of total uncertainty	63
12.2.5	Reporting of results	63
12.3	Applications to rubber testing	63
13	Sampling	64
13.1	Principles	64
13.2	Methodology	65
13.2.1	General	65
13.2.2	Acceptable quality level and limiting quality	65
13.2.3	Assessment of nonconformity	65
13.2.4	Inspection levels	66
13.2.5	Plans for sampling by attributes	66
13.2.6	Random sampling	67
13.3	Applications to rubber testing	68
14	Number of test pieces.....	68
14.1	Principles	68
14.2	Methodology	69
14.3	Applications to rubber testing	69
14.3.1	General	69
14.3.2	Refinement of confidence limits.....	69
14.3.3	Refinement of a pass/fail status	70
15	Expression of results.....	70
15.1	Principles	70
15.2	Methodology	70
15.2.1	The test report	70
15.2.2	Rounding.....	72
15.3	Applications to rubber testing	73
15.3.1	General	73
15.3.2	Construction of a histogram	73

15.3.3	Examples of rounding	74
16	Precision statements	74
16.1	General	74
16.2	Principles	74
16.3	Methodology	75
16.4	Applications to rubber testing	77
17	Design of experiments	78
17.1	General information and principles	78
17.1.1	General information	78
17.1.2	Principles	79
17.2	Methodology	92
17.2.1	General	92
17.2.2	Descriptive experiments	92
17.2.3	Comparative experiments	93
17.2.4	Response experiments	95
17.3	Applications to rubber testing	95
17.3.1	Descriptive experiments	95
17.3.2	Comparative experiments	97
17.3.3	Response experiments	101
18	Statistical quality control	105
18.1	Principles	105
18.2	Methodology	106
18.2.1	General	106
18.2.2	Control charts by attributes	106
18.2.3	Control charts by variables	106
18.3	Applications to rubber testing	108
18.3.1	General	108
18.3.2	Control charts	108
18.3.3	Cusum chart	111
Annex A	(informative) Mathematical form of the distribution functions referenced in this International Standard	115
Annex B	(informative) Additional forms of mean value	117
Annex C	(informative) Inter-relationships for measures of central tendency in the double exponential and Weibull distributions	118
Annex D	(informative) Equation for the calculation of standard deviation	119
Annex E	(informative) Construction of Weibull probability paper	121
Annex F	(informative) Equations for the calculation of Student's <i>t</i> -values	122
Annex G	(informative) Analysis of variance	123
Annex H	(informative) Equations for the calculation of regression coefficients	127
Annex I	(informative) The intercal method	129
Bibliography		130

Foreword

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

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

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

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

ISO 19003 was prepared by Technical Committee ISO/TC 45, *Rubber and rubber products*, Subcommittee SC 2, *Testing and analysis*.

Introduction

Statistical methods have an important role at all stages of the testing process, from the design of the experiment to the interpretation of results. Hence, those involved in testing require a basic understanding of statistical principles and knowledge of the statistical techniques which need to be applied.

There are many text books and International Standards which describe statistical methods, but it is convenient to have a guide which is a single, easy source of reference to the most commonly used methods and formulae, and which also considers their particular application to the various rubber test methods. This International Standard is therefore complementary both to the general standards on statistics and to the standards on methods of test for rubber.

The approach taken in this International Standard is that, for each subject, the text is structured into principles, methodology and applications to rubber testing. Under principles, the basic concepts of the subject are briefly outlined. Methodology considers the statistical techniques which can be applied; basic procedures and formulae are given but, as appropriate, more detailed matter is placed in annexes and, for less commonly used methods or more advanced treatment, reference is made to other publications. "Applications to rubber testing" indicates how and where the methods may be applied, and gives examples which are particular to rubber properties and tests.

Rubber and rubber products — Guidance on the application of statistics to physical testing

1 Scope

This International Standard provides guidance on the application of statistics to rubber testing. It is intended not to conflict with or replace existing International Standards covering basic statistical techniques, but rather to complement them and provide examples of those techniques applied to particular rubber testing situations.

2 References

This International Standard refers to other publications that provide information or guidance. These standards are listed in the Bibliography.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

NOTE These definitions, which are expressed as far as possible in non-mathematical terms, apply to the main statistical terminology used. More comprehensive and rigorous lists can be found in the various parts of ISO 3534 and in the standards dealing with specific statistical techniques indicated in the Bibliography.

3.1

population

totality of data that could (theoretically) be obtained to characterize the property of the rubber, compounding ingredient or process being measured

3.2

sample

data actually available from the population as a result of an experimental test programme having been undertaken

3.3

variability

tendency for tests performed on nominally identical test pieces to produce different test results

3.4

arithmetic mean

sum of the (population or sample) data divided by the number of values used

NOTE The “average” is the statistic most frequently used to describe a group of data. There are several kinds of average and they are often used in common parlance without specifying the type, which can be a source of confusion. Averages fall into two categories: computational and positional. The arithmetic mean is the most frequently used computational average. Others are considered in Annex B. Positional averages are the median and mode. The calculation of the arithmetic mean is given in Equations (1) and (2) in 6.2.2.2.

3.5
median
middle value (or average of the two middle values) when the data in a sample are arranged in numerically increasing value

3.6
mode
value of the property being measured which occurs with the greatest frequency

3.7
residual
difference (+ or –) between each value and the mean

NOTE The sum of the residuals must be 0.

3.8
variance
arithmetic mean of the squared residuals

3.9
standard deviation
square root of the variance

NOTE The calculation of standard deviation is given in Equations (5) and (6) in 6.2.3.2.1.

3.10
coefficient of variation
ratio of the standard deviation to the mean, generally expressed as a percentage

NOTE The calculation of coefficient of variation is given in Equation (8) in 6.2.3.4.

3.11
range
maximum value minus the minimum value

3.12
standard error
standard deviation of the estimate of the population mean

NOTE The calculation of standard error is given in Equation (7) in 6.2.3.2.3.

3.13
bias
difference between the average statistic and the true value of the parameter it is estimating, arising out of one or more systematic errors

3.14
accuracy
closeness of agreement between a test result and the accepted reference value

3.15
trueness
closeness of agreement between the average value of a large number of test results and the true or accepted reference value

NOTE It is usually expressed in terms of bias.

3.16
precision
closeness of agreement between test results

3.17

repeatability

precision obtained under conditions where independent test results are obtained with the same method on identical test material in the same laboratory by the same operator using the same equipment within a short interval of time

3.18

reproducibility

precision obtained under conditions where independent test results are produced with the same method on identical test material in different laboratories with different operators using different equipment

3.19

level of significance

probability of error associated with a significance test

3.20

distribution function

function describing the probability that a random variable will take a value less than or equal to a number x

3.21

density distribution

slope of the distribution function at every value, i.e. the first derivative of the distribution function

3.22

normal distribution

symmetrical “bell-shaped” density distribution which is fully defined by its mean and standard deviation

NOTE It is also known as the Laplace Gauss or Gaussian distribution.

3.23

double exponential distribution

asymmetrical distribution, fully defined by a single “shape” parameter, which has been used to characterize the distribution of tensile strengths in rubber compounds

3.24

Weibull distribution

symmetrical distribution fully defined by three parameters and found to be useful in characterizing lifetime tests such as fatigue

3.25

degrees of freedom

number of independent differences between the readings available for an estimate of standard deviation

3.26

confidence interval

range within which a value or parameter can be expected to lie with a given probability

3.27

confidence limits

extreme values of the confidence interval

4 Symbols

$a, b, c, \dots$	Constant coefficients in a regression line
C	The coefficient of concordance in Friedman's test, or Cochran's quotient when testing variances for the presence of outliers
C_i	The i th cusum value
C_{pq}	Factors used in the derivation of regression coefficients
C_v	The coefficient of variation
$f(x)$	A property or parameter which is a function of x or a density distribution function
F	The observed value of Snedecor's F -ratio in a given case
F_{cr}	The statistically critical value for F at a given confidence level and for the given degrees of freedom for the lesser and greater mean squares
F_r	The F -value for a regression line
H_0/H_a	The null/alternative hypothesis parameter
K	Friedman's statistic for a rank correlation test
M_z	The mean square for factor z
n	The number of values in a series
$p(x)$	A probability distribution function
P_m	The plot positions for the graphical presentation of a series of values
Q	Dixon's quotient when testing values or means for outliers
r	The repeatability of a test method for a particular test or series of tests
(r)	The repeatability expressed as a percentage of the mean from a test or series of tests
R	The reproducibility of a test method for a particular test or series of tests
(R)	The reproducibility expressed as a percentage of the mean from a test or series of tests
s	The estimate of the population standard deviation from the available sample
s'	the standard deviation of a series of numbers
S	The weighted standard error for the combination of two series of values, or the rank sum for a sample in Friedman's test
S_t	The total sum of the squares of the differences between individual values and their mean
S_z	The sums of squares for factor z
t_α	Student's t -value for a given probability (or confidence level) α
U_r	The random uncertainty in a measurement
U_s	The systematic uncertainty in a measurement
v_z	The number of degrees of freedom for factor z
x	An individual numerical value, such as the tensile strength of a single test piece
x_i	A single value in a series of values, such as a tensile strength in a set of five replicate values
x_{ij}	A single value in a series of values in which two factors are present, such as the tensile strength in sets of replicates obtained at different temperatures

$\bar{x}$	The arithmetic mean of a series of numbers, x_i
Z	The Z -score in hypothesis testing
α, β	The probability of an event occurring
μ	The population mean of a distribution
$\hat{\mu}$	The estimate of the population mean from the available sample
σ	The population standard deviation of a distribution

5 Limitations of test results

5.1 Variability

5.1.1 All measurements are subject to variability. It is necessary to know the sources of variability and make a reliable estimate of its magnitude. From this information, it should then be possible to judge the reliability of the results and hence their uncertainty and significance.

5.1.2 The term population is, expressed simply, the total number of objects in a large group (see 3.1). In testing terms, a population may be, for example, the total number of possible tensile strength results which could be obtained on a particular rubber compound if every piece of the material made was tested.

5.1.3 A sample is a selected number of, for example, parts or tensile results taken from the population.

NOTE 1 To avoid confusion, sample should not be used to mean test piece.

NOTE 2 Sample can have two meanings:

- a) in the physical sense, as in taking five parts from a boxful;
- b) in the statistical sense, as in taking five test results.

5.1.4 If five tensile strength measurements are made from a sheet taken from a batch of rubber, an example of the results which might be obtained is shown in Table 1.

Table 1 — Tensile strength measurements from one batch of rubber

Measurement number	Tensile strength MPa
1	16,8
2	15,4
3	16,3
4	17,7
5	17,6

The sources of variability are:

- a) the intrinsic variability of the sheet rubber, arising from the fact that it is not perfectly homogeneous;
- b) the variability due to the testing procedure, including test piece preparation, machine accuracy and operation error.

If several sheets are tested, there is an additional source of variability due to variations in moulding.

If several batches are mixed, two more sources of variation are added:

- 1) that from the mixing procedure;
- 2) any variation in compounding ingredients.

If sheets which are nominally the same are given to a number of operators, there is variability due to the operators.

Similarly, if a number of different test apparatuses are used, variability due to the machines is introduced. Taking things further, sheets may be tested in different laboratories and between-laboratory variability introduced.

5.1.5 In practice, the magnitude of variability is minimized by carefully controlling the processing operations and the test apparatus and procedures. It is never eliminated altogether and inter-laboratory comparisons have demonstrated that for many rubber tests it can be far greater than was previously thought.

Whatever test is carried out, there is genuine variation due to the material and also variation due to uncontrolled testing errors. It is often very difficult to separate the two. For example, testing errors can arise from

- a) random variations in test piece geometry due to limitations in cutting precision;
- b) variations in the response of the test apparatus;
- c) fluctuations in the operator's performance.

These errors may be large or small and of indeterminate direction so that eventually they tend to cancel out. More serious is systematic error or bias which is unidirectional, for example the error due to a machine being wrongly calibrated or an operator consistently misreading a scale.

5.1.6 Testing error apart, the sample of results will not be representative of the whole population if the physical sample is not representative. Differences between repeat mixes and between repeat mouldings should be expected because of some variation in the quantities and quality of ingredients used, the efficiency of mixing and the time of curing, etc. If gross errors are made, some very atypical results are recorded and it is dangerous to rely heavily on one small sample unless certain that it is representative.

The evaluation of an alternative ingredient by comparison with the standard formula may be considered. The mixes are uniform, the tester follows the procedures correctly and it is concluded, using statistical methods, that the new ingredient is an improvement. It is easily forgotten that this conclusion assumes that the samples of each compound were truly representative of the population. If the variability which would arise from repeat mixings is rather larger than the testing error, as is often the case, then tests on a series of repeat mixes may show no difference between the ingredients or even that the new ingredient was worse.

5.2 Accuracy, trueness and precision

Accuracy is the closeness of agreement between a test result and the accepted reference value (see 3.14), while trueness is the closeness of agreement between the average value of a large number of test results and the true or accepted reference value (see 3.15). Precision, on the other hand, is the closeness of agreement between the test results (see 3.16), independent of any reference value that may exist. To keep variability to a minimum, the test method should be as reproducible as possible, i.e. it should have good precision. However, having high precision may be of little value if the test has a large bias and hence poor accuracy. Both are required and indeed they are related in that poor precision (poor reproducibility) will contribute to lowering the accuracy.

Reproducibility (see 3.18) is the term generally reserved to describe the variation found between different laboratories, and perhaps also at different times. Repeatability (see 3.17) is used to describe the variation between repeats in the same laboratory at essentially the same time. It follows that laboratories may exhibit very good repeatability but, because of bias, the reproducibility between the laboratories is poor.

5.3 Relevance and significance

5.3.1 If accuracy or repeatability were the only interest, testing would be limited to the most accurate or precise methods. However, the test should be relevant in the sense that the results have a useful meaning in terms of material or product performance. All tests are not equal: some have more relevance than others in terms of product performance, material consistency or value as design data. The word significance is sometimes used to mean relevance and applied to the actual test or property measured, but significance is used in this International Standard in the statistical sense as in one material being significantly stronger, for example, than another.

Significance in this sense is concerned with whether observed differences in results are likely to be real or can reasonably be attributed to chance alone. If the probability of obtaining the observed difference through pure chance is small, for example less than 1 in 20, then the difference is said to be significant.

5.3.2 The set of tensile strength results quoted in 5.1 could be compared to other sets obtained on different materials on the same occasion giving, for example, three sets as in Table 2.

Table 2 — Tensile strength measurements from three materials

Measurement number	Tensile strength		
	MPa		
	Material A	Material B	Material C
1	16,8	15,6	16,4
2	15,4	16,4	15,4
3	16,3	14,5	14,3
4	17,7	15,8	14,7
5	17,6	16,0	14,4

The averages of the results for materials A and B are higher than that for C but an assessment should be made as to whether or not they are significantly higher. Without the use of statistical tools it is rather difficult to make this assessment. In fact, using a test for significance as discussed in 7.2.2 it can be proved that A is significantly greater than C with 95 % confidence but that A is not significantly different from B, again with 95 % confidence. This is a useful conclusion but its limitations should be appreciated. The statistical tests prove (with a 1 in 20 chance of being wrong) that results A are significantly greater than C. They do not prove that material A is stronger than material C. It is known that results from one sheet of one mix may not be representative of a formulation and these results from a very small test programme should be treated with caution.

5.3.3 In the above example the differences between the average results were relatively small but tensile strength can be measured accurately with reasonably small variability so that it is not surprising that 10 % difference could be proved significant. For other, less reproducible tests a much greater percentage difference may be needed before the difference can be proved significant. For example, in an electrical resistivity test the mean value for one material was several times higher than that for a second material but the difference could not be proved significant. The deduction can be made that significance is not only dependent on the difference between mean values but also on the amount of variability which is inherent in the test.

6 Distribution of results and measures of central tendency

6.1 Principles

A collection of values, for example individual test results relating to a specific property, are arranged about a mean value. Usually the distribution of results may be represented by a particular mathematical law such as the curve shown in Figure 1.

From the shape of the curve it is possible to obtain useful measures of the tendency towards a central value, the degree of the spread of results and the proportion of results likely to differ by more than a certain amount from the centre.

6.2 Methodology

6.2.1 Types of distribution

6.2.1.1 The normal distribution

The most widely used distribution function is called the normal, or Gaussian, distribution function (see 3.22) which can be completely characterized by two parameters, the mean μ (see 3.4), and the standard deviation σ (see 3.9). These parameters are considered further in 6.2.2 and 6.2.3 respectively. The density distribution is a symmetrical bell-shaped function, the mathematical description of which can be found in Annex A.

Values of the ordinate, i.e. the density of the function, $f(z)$, at given values of z have been tabulated and can be found in any statistics text book. In order to make the tables suitable for general application, the abscissa z , is presented in reduced form, such that z is the number of standard deviations x (the value measured in the experiment) is away from the mean. Since the curve is symmetrical it is usual to find only the positive values of z that are tabulated since $f(z) = f(-z)$.

The proportion of the whole distribution which lies between two values, x_1 and x_2 (i.e. the probability distribution function) can be determined from the integral of the density distribution (see Annex A), but since this integral cannot be expressed analytically, it is more convenient to use tabulated values, which again are available in the form of the reduced variable Z , in any standard text on statistics.

In these tables the value of x_1 is usually set to μ with only the positive reduced variable Z quoted.

For these tables:

when $z = 0$, $p(Z) = 0,0$;

when $z \rightarrow \infty$, $p(Z) \rightarrow 0,5$.

If z is negative, $p(Z) = -p(+Z)$.

To find the proportion of the curve (i.e. the probability of the observation) lying between x_1 and x_2 , ($x_2 > x_1$) the procedure is as follows:

a) Determine Z_1 and Z_2 where

$$Z_1 = (x_1 - \mu)/\sigma,$$

$$Z_2 = (x_2 - \mu)/\sigma.$$

b) Find $p(Z_1)$ and $p(Z_2)$ from the tables.

c) Determine the required probability $p(Z_2 - Z_1)$ where

$$p(Z_2 - Z_1) = p(Z_2) - p(Z_1).$$

NOTE The signs should be taken into account.

6.2.1.2 The double exponential distribution

In the case of the distribution of tensile strengths (and elongations at break) for vulcanizates, there is considerable evidence that the density distribution function is not symmetrical but is skewed towards lower strength values (see References [1] to [6]), although this has been questioned by some investigators (see Reference [7]). The density distribution function which has been found to give a good representation of these skewed data is given by the double exponential distribution (see 3.23), shown in Figure 2 and described mathematically in Annex A.

Although of theoretical interest, the double exponential function has not found widespread application, both because of its complexity and the fact that, with the small number of test pieces normally considered in a tensile test, there is no significant difference in the mean and standard deviation derived from the double exponential function and from the normal Gaussian function.

6.2.1.3 The Weibull distribution

A distribution function that frequently arises out of fatigue data and similar lifetime testing is the Weibull distribution function (see 3.24), the form of which is illustrated in Figure 3 and described mathematically in Annex A.

The distribution is characterized by the following three parameters:

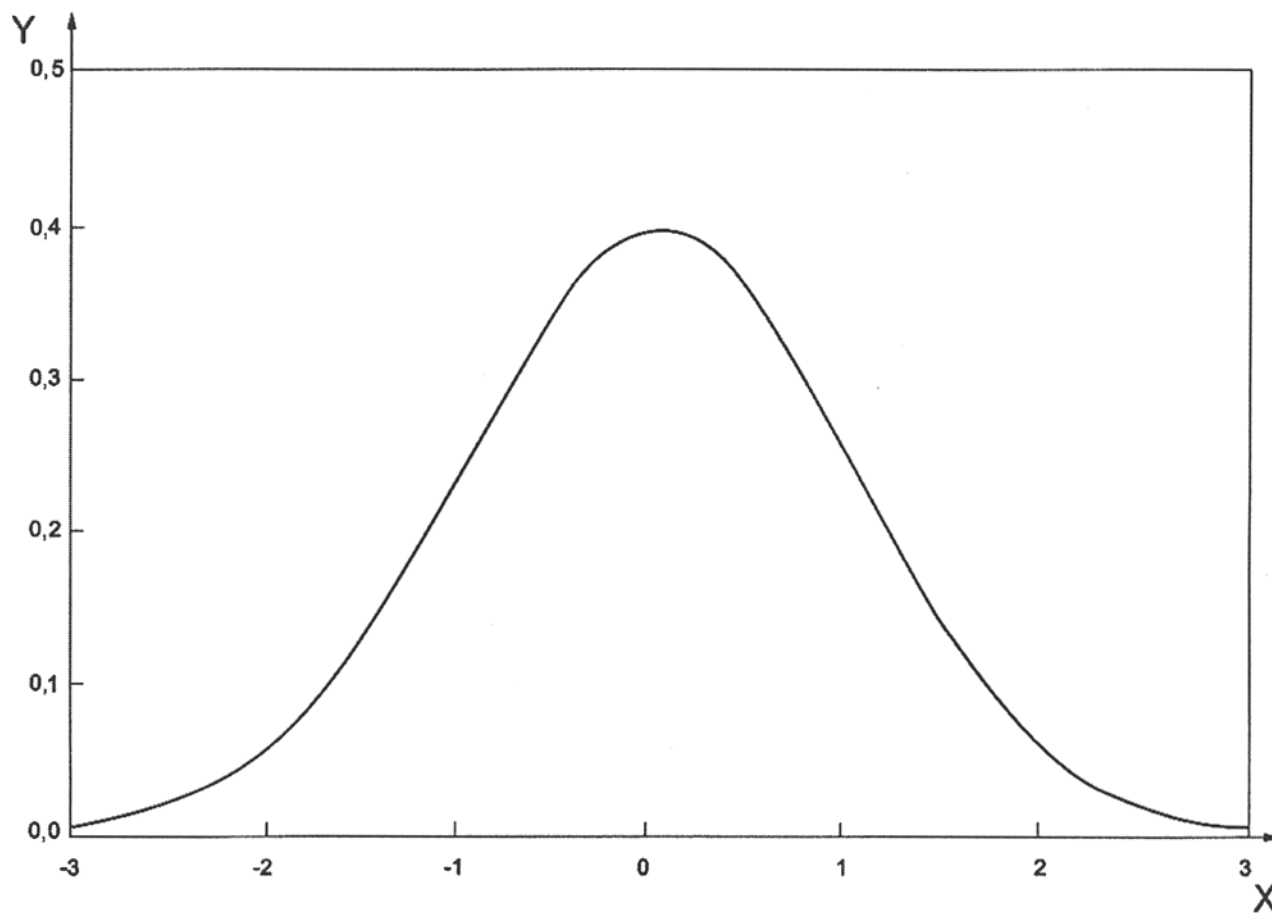
- a) The parameter a represents the minimum life parameter for x at which the probability of failure just reaches zero (that is, giving an infinite lifetime). In most practical applications, a is taken to be zero but, where there is a genuine fatigue limit, a can take a finite, non-zero, value.
- b) The parameter b affects both the spread of results and the peak position of the density function.
- c) The parameter k alters the shape of the density distribution.

When $k = 1$, the function is a simple exponential.

When $k > 1$, the distribution increases from zero (at $x = a$), reaches a maximum and then decreases monotonically, reaching zero again at $x = \text{infinity}$.

When k is approximately 3,44, the distribution is approximately Gaussian, with the mean and median equal to each other.

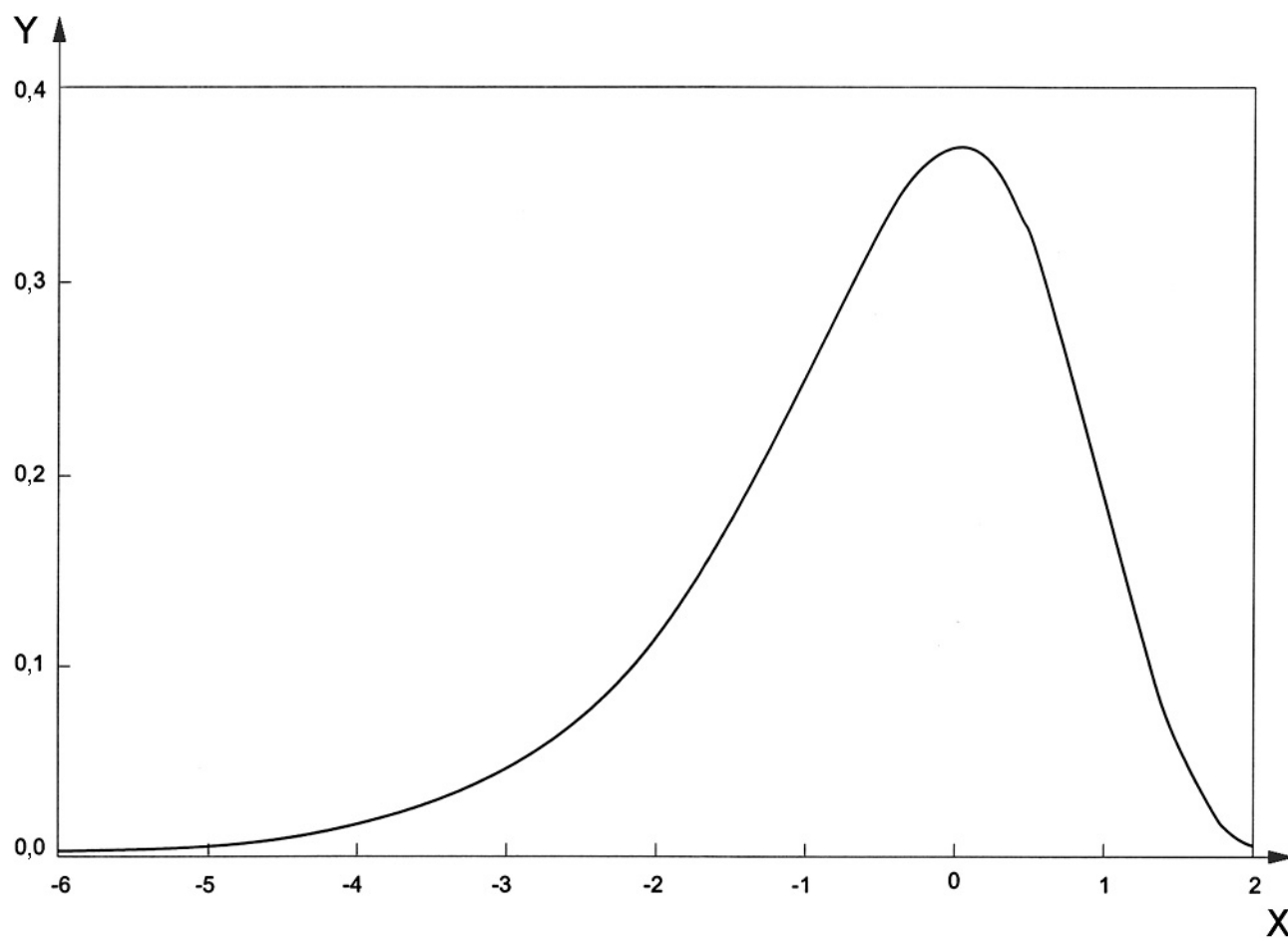
Generally, the Weibull distribution is positively skewed (i.e. skewed towards longer lifetimes).



Key

X parameter y (see Annex A)
Y probability

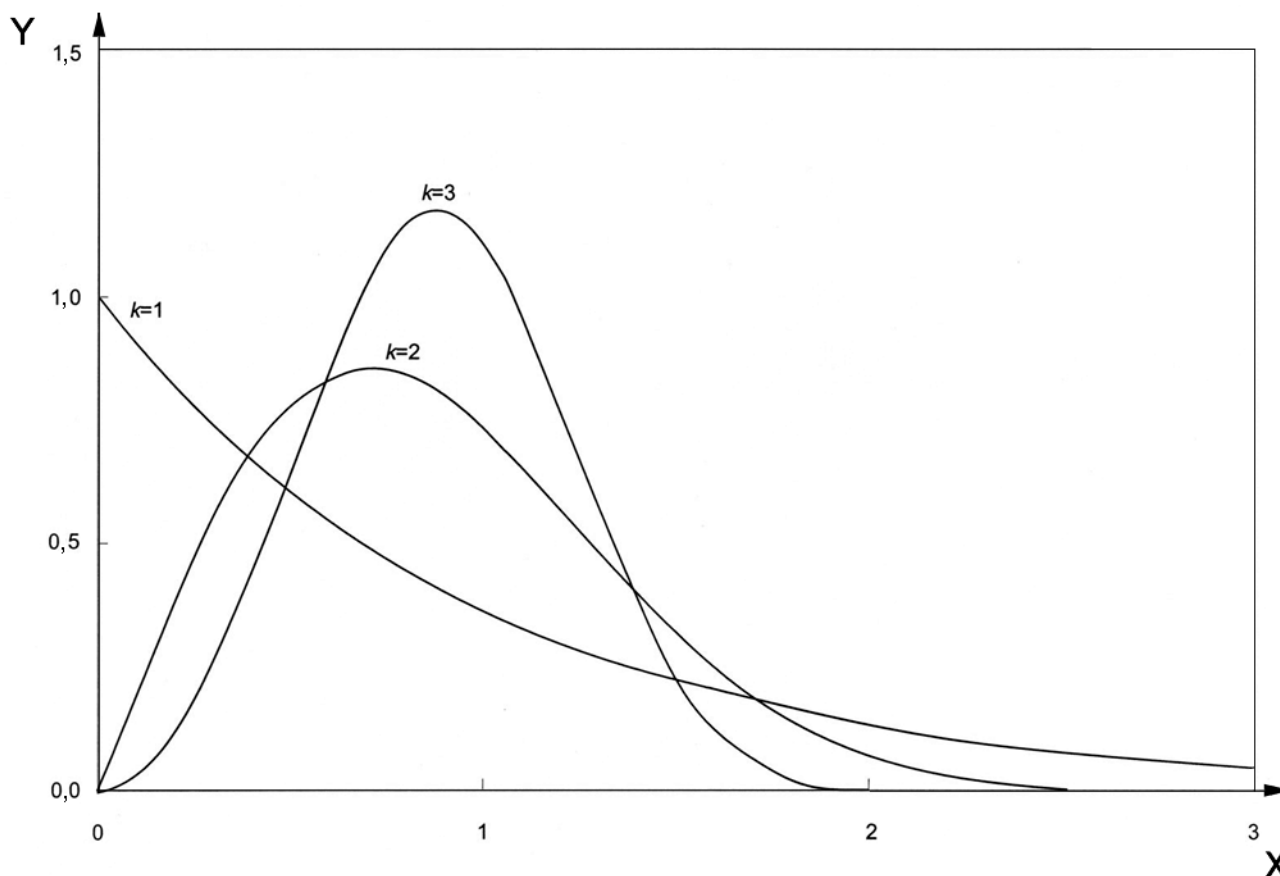
Figure 1 — Gaussian (normal) density function



Key

X parameter ε (see Annex A)
Y probability

Figure 2 — Double exponential density function



Key

X parameter $(x - a)/b$ (see Annex A)
Y probability

Figure 3 — Weibull density function

6.2.2 Measures of central tendency

6.2.2.1 General

Even under the most careful experimental conditions, carrying out repeat measurements on identical material produces a scatter of results. It is useful, therefore, to have some idea of the average or typical value that can be expected of those results. Since typical values tend to lie towards the middle of the data when these are arranged in numerical order, such numbers are also called measures of central tendency.

The small number of experimental results obtained is a sample of the infinite number of results, i.e. the population, that would fully encompass the measurement being made. The more results that are available, the more accurately sample statistics match the population statistics. Clearly, the average of the sample can be determined precisely but, generally, it is the average of the population from which the sample is drawn that is of greater interest and this can only be estimated from the sample, or samples, available. It is appropriate to distinguish sample and population statistics and hence different symbols are used.

6.2.2.2 The mean

The mean (see 3.4) is the most commonly encountered measure of central tendency and it is often also called the average. The precise meaning should be clear when these words are used as there are several ways of averaging a set of numerical values. The arithmetic average or mean is the one that is generally meant and so the word arithmetic is often omitted but, where there is the possibility of confusion, it ought to be included. In

this International Standard, where the word mean is used the arithmetic mean is intended unless otherwise stated.

The mean of a sample, $\bar{x}$, is defined as the sum of the individual numerical values in the sample divided by the number of values in the sample and is given mathematically by the equation

$$\bar{x} = \frac{(x_1 + x_2 + x_3 + \dots + x_n)}{n} \quad (1)$$

which in the shorthand sigma notation becomes

$$\bar{x} = (\Sigma x) / n \quad (2)$$

As noted previously, the mean of the population from which the sample was taken is given the symbol μ . This value is almost never known in practice, but has to be estimated from the sample. The estimated mean of the population $\hat{\mu}$, based on the available sample, is taken to be equal to the sample mean. In other words

$$\hat{\mu} = \bar{x} \quad (3)$$

where $\hat{\mu}$ is an estimate of μ .

Where there are a large number of results having discrete values, it may be more convenient to record the number of occurrences of each value. If each value x occurs f times, then

$$\bar{x} = [\Sigma (fx)] / n \quad (4)$$

NOTE In this case, $n = \Sigma f$.

The same technique can be applied where an infinite variation in value x can occur, but it is more convenient to group the data into bands, counting the number of results in each band.

Other types of mean sometimes encountered are briefly described in Annex B.

6.2.2.3 The median

If the data in the set of results are arranged in numerical order, then the middle value (or the mean of the two middle values where there is an even number of values) is the median (see 3.5).

Geometrically, the median of a density distribution function is the value of the abscissa corresponding to that vertical line which divides the distribution into two equal areas.

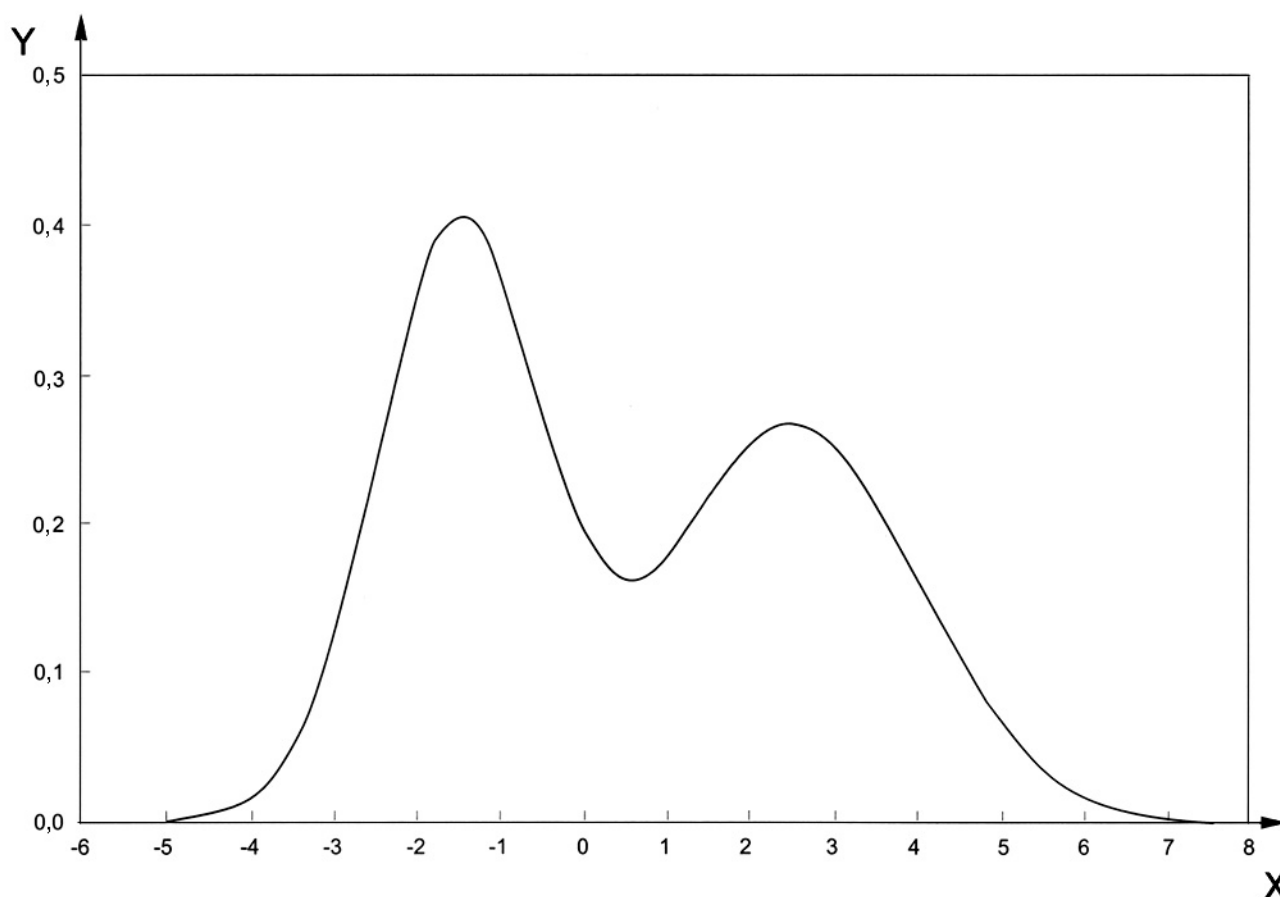
For tensile strength or elongation at break, where the double exponential distribution function is expected to apply, it is strongly recommended that the median value be quoted for the measure of central tendency. The reason for this is three-fold:

- for the small number of replicate results normally involved in a tensile test (typically five or even three), the median is not influenced by any extreme values that may have arisen;
- for a small number of replicates, the median can be deduced by inspection without recourse to any calculation;
- there is an equal probability of an individual observation being greater than or being less than the median.

6.2.2.4 The mode

The mode (see 3.6) of a density distribution function is the value of the abscissa at which the maximum of the function occurs. Although little used in practice, it is referenced here for completeness as it has been used to characterize the double exponential distribution enunciated by Kase from earlier work by Fisher and Tippet ^[1].

Almost all distributions encountered in the rubber industry are unimodal, i.e. having a single maximum value. However, bimodal (having two maxima) or multi-modal distributions can result where two, or more, different mechanisms are occurring simultaneously in a process. Thus, tensile testing at a temperature close to the glass transition of the rubber can lead to a mixture of brittle and rubbery failure. This mixed mode of failure is best analysed by separating the results from the two types of failure and analysing each independently. An example of a bimodal distribution is given in Figure 4.



Key

X parameter y (see Annex A)
Y probability

Figure 4 — Bimodal density function

6.2.2.5 Inter-relationships

It is clear from the definitions of these various measures of central tendency that for symmetrical distributions

Mean = Median = Mode

The inter-relationships for non-symmetrical distributions are necessarily more complex and those for the double exponential and the Weibull distribution are given in Annex C.

6.2.3 Measures of dispersion

6.2.3.1 General

Just as it is useful to know the average value of a set of numbers, so also is their spread or dispersion important. The more loosely packed the numbers are about the mean, the less discrimination there is between two sets of numbers or between the experimental values and, for example, the specification requirement.

The variance and standard deviation are used with the mean, whilst the range is normally used with the median and mode. Also, the quartiles indicate 25 % of the data is higher or lower and the percentiles a specified percent is higher or lower.

6.2.3.2 Standard deviation

6.2.3.2.1 By definition, the standard deviation s' of a sample of results having values of x_i is given as the square root of the average squared deviation of each of the values from their mean ($\bar{x}$). It is given by the equation:

$$s' = \left[\frac{\sum (x_i - \bar{x})^2}{n} \right]^{1/2} \quad (5)$$

An alternative form that is sometimes more convenient to use is given in Annex D. Some of the risks associated with its use are also given.

The symbol s' is used here to represent the standard deviation of the sample of n results. If this sample is representative of the population from which it has been drawn, then the true standard deviation σ can be estimated by equating it with the statistic s defined by applying Bessel's correction to Equation (5). s is given by the equation:

$$s = \left[\frac{\sum (x_i - \bar{x})^2}{n-1} \right]^{1/2} \quad (6)$$

This is the form of the standard deviation that should normally be used when performing statistical tests since it is an estimate for the population as a whole and not simply the particular sample chosen.

6.2.3.2.2 It can be shown that

- a) 68,26 % of values for a normal distribution lie within ± 1 standard deviation of the mean;
- b) 95,44 % lie within ± 2 standard deviations;
- c) 99,73 % lie within ± 3 standard deviations.

For all practical purposes, in a normal distribution (or the distribution of the means of sets of values), the whole population is covered by six standard deviations. Therefore this interval is used in the setting of control charts (see Clause 18).

6.2.3.2.3 Related to the standard deviation is the standard error of the mean, which is determined from the standard deviation by dividing the standard deviation by the square root of the number of observations in the sample. It is, therefore, the standard deviation of the estimate of the population mean (see 3.12). Thus

$$S = s/\sqrt{n} \quad (7)$$

where S is the standard error of the mean.

The standard error is a measure of the expected spread of a series of mean values in the same way that the standard deviation is a measure of the expected spread of the individual values, and it is the standard error which should be used when making statistical comparisons between groups of numbers which are themselves the means of a group of numbers (see 6.3.2.4 and 7.2.2.2).

6.2.3.3 The range

While the standard deviation has valuable mathematical properties, historically it was somewhat cumbersome to calculate without computers and on occasion this might outweigh its value. Where a less precise estimate will suffice, the range, i.e. the maximum value minus the minimum value in the sample, may be used. It is possible to estimate the standard deviation from the range by multiplying the range by a factor which depends on the number of results in the set. For values of n between 2 and 11, the factors A_n are given in Table 3.

Table 3 — Table of factors for converting range to standard deviation

n	A_n
2	0,886
3	0,591
4	0,486
5	0,430
6	0,395
7	0,370
8	0,351
9	0,337
10	0,325
11	0,315

Thus, $s \approx \text{range} \times A_n$.

The range and the quartiles can also be used as measures of spread around the median in an analogous manner to the standard deviation around the mean.

6.2.3.4 Coefficient of variation

Where the relative dispersion of results about their mean is of interest as, for example, in the comparison of the variabilities of the volume swell test with the density test, the ratio of the standard deviation to the mean can be used to normalize the effects of having very different numerical values for the means. It is usual to express this ratio as a percentage and to call it the coefficient of variation (see 3.10). It is given by the equation:

$$C_v = (s/\bar{x} \times 100) \quad (8)$$

where C_v is the coefficient of variation.

6.2.4 Transformation to normal distribution

The individual results obtained from a rubber test may not immediately conform to the normal distribution function. Other possible distribution functions such as the double exponential or Weibull may be theoretically

(or empirically) found to give a better representation of those data (see 6.2.5). Where certain statistical inferences need to be made concerning a set of data, for example the determination of confidence intervals or limits (see Clause 7), knowledge of the distribution function which describes the data is required. Because of the extensive range of tests and techniques that have been developed for the normal distribution, it is worth investigating whether a simple transformation of the data will result in a normally distributed data set to an accuracy sufficient for the situation being analysed.

It is also important to bear in mind that, even where the normal distribution is not found for the individual readings, the distribution of the means of groups of readings (as low as three per group) such as those obtained in the usual tests on rubber nearly always approximate to the Gaussian form (central limit theorem).

The transformation most commonly found to be effective in this regard is to take the logarithms of the values and treat these as the variable to be analysed. The use of log-probability graph paper or computer transformation makes this a very quick and simple test.

Other transformations that have been found to work on occasion include

- a) taking the square root;
- b) taking the reciprocal of the value.

Sometimes the addition of a constant to (or subtraction of a constant from) the value prior to taking logarithms, roots or reciprocals is required. It may be possible to deduce a suitable value for this constant from knowledge of the process being examined, but often it should be established empirically.

6.2.5 Test of departure from normality

6.2.5.1 Normal distribution function

6.2.5.1.1 The simplest way of testing a series of observations for the normality of its distribution is by plotting the results on probability paper. A normally distributed set of observations results in a straight line from which the mean and standard deviation can be derived.

6.2.5.1.2 The procedure is as follows:

- a) Sort the data into ascending numerical order.
- b) As probability paper is printed in the form of a percentage function, calculate the plotting position P_m for point m out of a total of n results using the equation:

$$P_m = 100m/(n + 1) \quad (9)$$

- c) Plot the value of the m th point as ordinate against P_m as abscissa.

6.2.5.1.3 A more or less straight line indicates that the distribution is normal, but a marked deviation from linearity indicates that the distribution is non-Gaussian. Under these circumstances, the nature of the deviation may indicate the kind of distribution function that is more appropriate. In particular, if the larger values are systematically higher than the straight line defined by the lower values, the use of a logarithm or root transformation (see 6.2.4) will often result in a linear plot.

6.2.5.1.4 The above check does not provide a true test for normality in the statistical sense, but does give a rapid indication of the suitability, or not, of the normal distribution function as the model for the data observed. If the plot is not reasonably linear even after transformations have been applied, then more detailed symmetry and kurtosis tests may be required (taking into account the purpose for which the data are being used). These are outside the scope of this International Standard and the interested user is referred to ISO 5479.

Log-normal and Weibull graph papers are also available.

6.2.5.2 Double exponential distribution function

6.2.5.2.1 Where the distribution is expected to follow the double exponential function

- a) order the data into descending numerical value;
- b) plot the value as ordinates against an abscissa of plot positions which are given by Table 4.

6.2.5.2.2 If the double exponential function is valid, then a straight line will result and the following values may be obtained as described:

- a) the mode, which corresponds to the ordinate at which the abscissa is zero;
- b) the standard deviation, which is the difference in ordinates corresponding to a unit difference in the abscissa (i.e. it is the slope of the line);
- c) the median strength, which is the value corresponding to an abscissa of $-0,366\ 5$;
- d) the mean, which is the value corresponding to an abscissa of $-0,577\ 2$.

Table 4 — Plot positions for the double exponential distribution

n^a	Total number of test results, N																								
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25				
1	0,89	0,97	1,92	1,06	1,10	1,13	1,17	1,19	1,21	1,23	1,25	1,27	1,28	1,29	1,31	1,32	1,33	1,35	1,36	1,37	1,37				
2	0,21	0,38	0,48	0,57	0,63	0,69	0,74	0,79	0,81	0,84	0,88	0,91	0,93	0,95	0,97	0,99	1,01	1,02	1,04	1,06	1,07				
3	-0,40	-0,14	0,04	0,18	0,28	0,36	0,44	0,50	0,55	0,59	0,62	0,67	0,70	0,71	0,74	0,77	0,79	0,81	0,83	0,85	0,87				
4	-1,15	-0,68	-0,39	-0,19	-0,05	0,07	0,16	0,24	0,30	0,36	0,41	0,45	0,50	0,53	0,56	0,59	0,62	0,64	0,67	0,69	0,71				
5	-2,54	-1,37	-0,89	-0,59	-0,38	-0,23	-0,11	0,00	0,09	0,15	0,21	0,26	0,31	0,34	0,38	0,43	0,46	0,49	0,52	0,55	0,57				
6		-2,61	-1,55	-1,06	-0,75	-0,54	-0,38	-0,25	-0,15	-0,06	0,02	0,09	0,15	0,18	0,23	0,28	0,31	0,35	0,38	0,42	0,44				
7			-2,76	-1,70	-1,20	-0,90	-0,67	-0,50	-0,38	-0,27	-0,18	-0,10	-0,19	0,02	0,08	0,13	0,18	0,22	0,26	0,29	0,32				
8				-2,88	-1,83	-1,33	-1,02	-0,79	-0,63	-0,49	-0,37	-0,27	-0,29	-0,12	-0,07	-0,01	0,04	0,09	0,13	0,17	0,20				
9					-2,99	-1,94	-1,44	-1,13	-0,91	-0,73	-0,58	-0,47	-0,37	-0,29	-0,21	-0,15	-0,10	-0,04	0,01	0,05	0,09				
10						-3,09	-2,05	-1,53	-1,23	-1,00	-0,82	-0,67	-0,55	-0,46	-0,38	-0,30	-0,23	-0,18	-0,12	-0,07	-0,02				
11							-3,18	-2,14	-1,63	-1,31	-1,08	-0,90	-0,76	-0,64	-0,54	-0,45	-0,37	-0,31	-0,24	-0,19	-0,14				
12								-3,26	-2,22	-1,71	-1,39	-1,16	-0,98	-0,84	-0,71	-0,61	-0,52	-0,44	-0,37	-0,31	-0,25				
13									-3,34	-2,29	-1,79	-1,47	-1,23	-1,05	-0,90	-0,78	-0,67	-0,58	-0,59	-0,43	-0,37				
14										-3,40	-2,36	-1,86	-1,53	-1,30	-1,12	-0,97	-0,85	-0,74	-0,65	-0,56	-0,49				
15											-3,47	-2,43	-1,92	-1,60	-1,36	-1,18	-1,03	-0,90	-0,79	-0,70	-0,62				
16												-3,53	-2,49	-1,99	-1,66	-1,42	-1,24	-1,08	-0,96	-0,84	-0,76				
17													-3,58	-2,55	-2,05	-1,72	-1,47	-1,29	-1,13	-1,01	-0,90				
18														-3,64	-2,61	-2,10	-1,77	-1,52	-1,34	-1,18	-1,06				
19															-3,69	-2,65	-2,15	-1,82	-1,57	-1,39	-1,24				
20																-3,74	-2,70	-2,20	-1,87	-1,62	-1,44				
21																	-3,78	-2,75	-2,23	-1,91	-1,67				
22																		-3,82	-2,79	-2,28	-1,96				
23																			-3,86	-2,83	-2,33				
24																				-3,90	-2,88				
25																					-3,94				

^a n = serial number of test result when arranged in decreasing magnitude.

^a n = serial number of test result when arranged in decreasing magnitude.

6.2.5.3 Weibull distribution function

If the Weibull distribution function is assumed to be valid, the procedure is as follows:

- sort the data into ascending order;
- calculate the plotting positions P_m as for the normal distribution above;
- plot the values for P_m against the observed lifetime on special Weibull probability paper.

NOTE Although this can be purchased directly, it is readily constructed from normal log-log graph paper and Annex E shows how this can be done.

6.3 Applications to rubber testing

6.3.1 General

For many tests, results will approximate to a normal distribution and it is appropriate to express results as the arithmetic mean and the standard deviation as routine practice. Typical examples are given in 6.3.2 to 6.3.5.

6.3.2 Tensile testing

6.3.2.1 The following three sets of 12 replicate tensile strengths shown in Table 5 were observed after testing in accordance with ISO 37 and arranging the results in descending order.

As there are 12 results, the median in each case is the average of the middle two values.

Table 5 — Tensile strength measurements

Measurements in MPa					
Compound A		Compound B		Compound C	
Tensile strength	Median	Tensile strength	Median	Tensile strength	Median
26,7	25,8	28,4	26,4	19,7	18,3
26,2		27,9		19,6	
26,1		27,4		19,2	
26,1		27,1		19,0	
25,9		26,8		18,7	
25,8		26,5		18,4	
25,8		26,3		18,1	
25,8		26,2		17,3	
25,7		26,0		16,4	
25,6		25,9		15,6	
25,1		24,6		15,1	
25,0		24,1		13,5	

6.3.2.2 The information contained in Table 6 has been calculated by the following methods:

- a) the mean has been calculated as defined in 6.2.2.2;
- b) the standard deviation and the standard error of the mean have been calculated as defined in 6.2.3;
- c) the calculated median has been calculated as defined in C.1 (i.e. for a double exponential distribution) using the values obtained in a) and b).

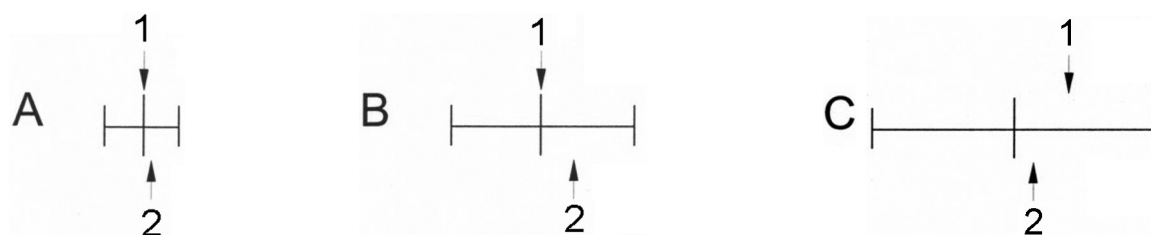
Table 6 — Calculated values for tensile strength measurements

Compound	Mean	Standard deviation	Standard error of the mean	Observed median	Calculated median
A	25,8	0,46	0,13	25,8	26,6
B	26,4	1,24	0,36	26,4	26,6
C	17,6	1,99	0,57	18,3	17,9

6.3.2.3 On comparing the two sets of median values, it is clear that the median obtained from inspection of the data is essentially the same as that calculated from the mean, the standard deviation and Annex C.

6.3.2.4 Examining the mean values of the three compounds (in relation to their standard errors) shows A and B to be within experimental error of each other, but compound C to be significantly different.

6.3.2.5 All the results have been summarized graphically in Figure 5.



Key

- 1 median 1 (observed value)
- 2 median 2 (calculated value)

Figure 5 — Graphical representation of mean and median data for three compounds
(using the tensile strength data of Table 6)

6.3.3 Fatigue

6.3.3.1 A tension fatigue test carried out in accordance with ISO 6943 on 10 replicate test pieces gave the observations listed in Table 7.

Table 7 — Tension fatigue test measurements

Test piece	Cycles to failure
1	219
2	347
3	494
4	593
5	700
6	858
7	1 037
8	1 146
9	1 461
10	1 795

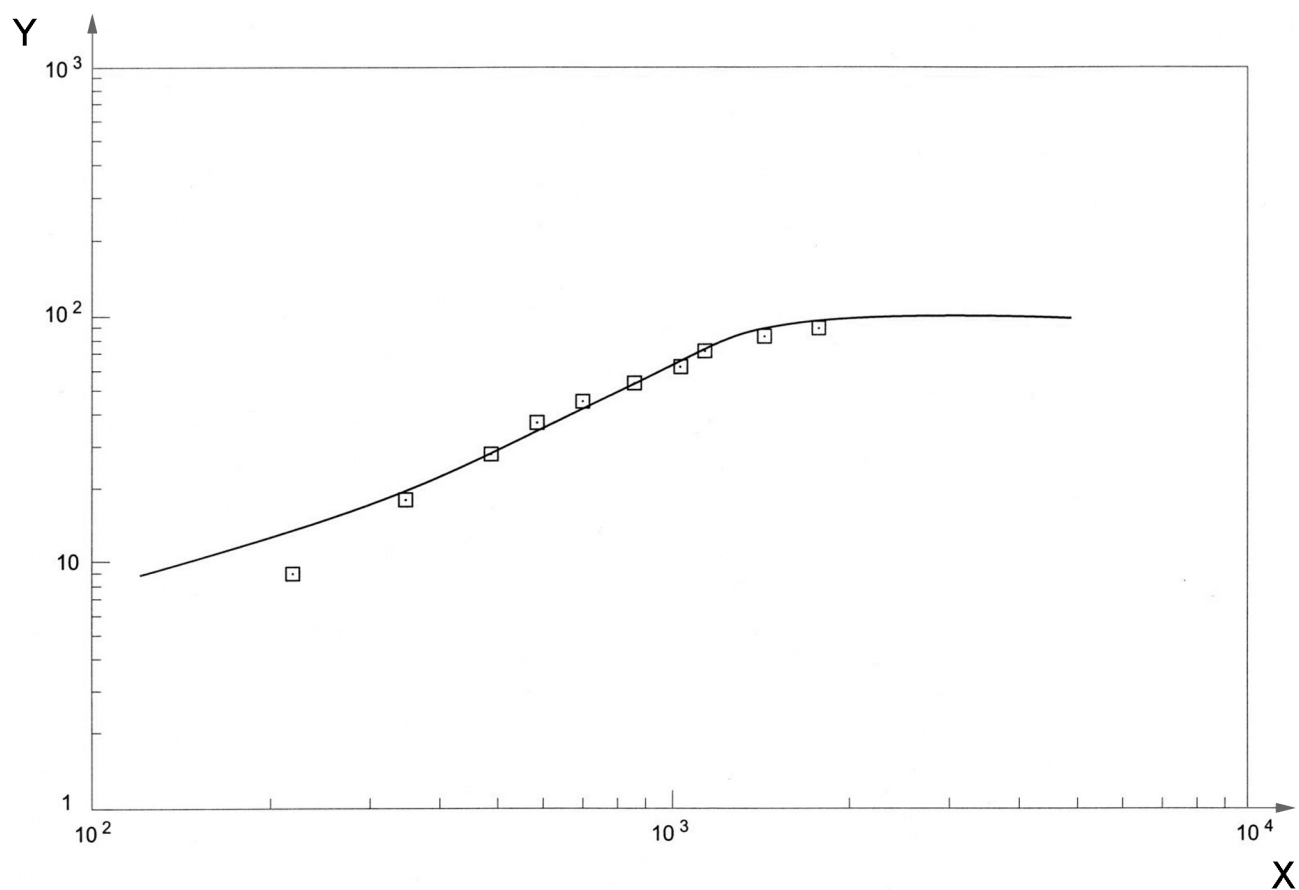
6.3.3.2 When the data in Table 7 are plotted as a normal distribution function, a systematic departure from the expected curve is observed, as shown in Figure 6 where a log-log scale has been used for convenience. The expected normal distribution of lifetimes is shown as the full curve.

6.3.3.3 From the type of test being performed, it would be expected that a Weibull distribution would give a good representation of the data. Hence a Weibull plot, constructed in accordance with Annex E, results in the plot shown in Figure 7. A linear regression analysis (see Clause 11) using the Weibull ordinate for the y -values and the logarithm of the fatigue life as the x -values produces the following values:

$$k = 1,53;$$

$$b = 1\,006.$$

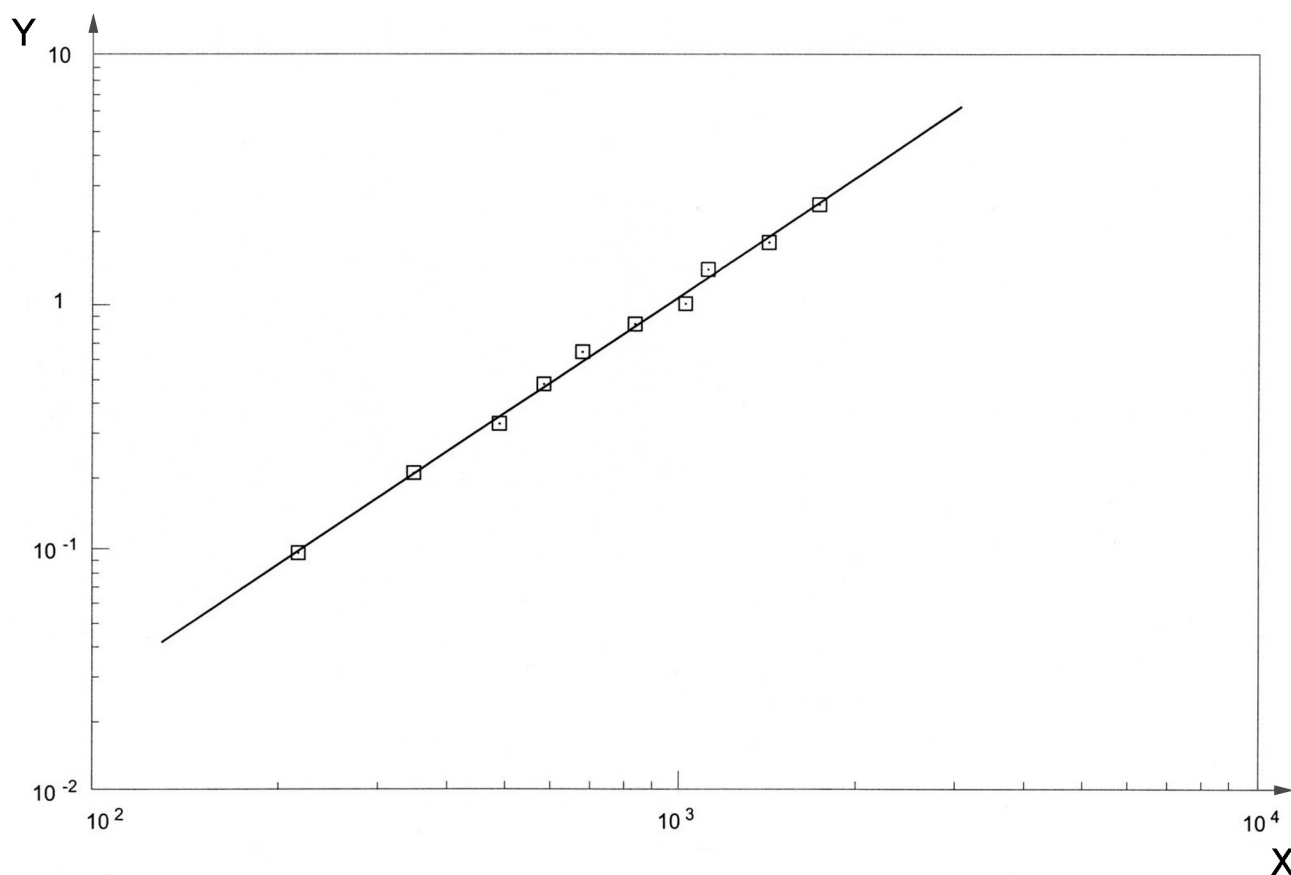
The strong linearity of the plot (the value of the variance ratio is over 5 000 and so the regression is highly significant) indicates that the coefficient a in the Weibull equation is zero.



Key

X fatigue life
Y normal ordinate

Figure 6 — Fatigue data: normal plot (using the data of Table 7)



Key

X fatigue life
Y Weibull ordinate

Figure 7 — Fatigue data: Weibull plot (using the data of Table 7)

6.3.4 Conversion to normal distribution

6.3.4.1 Data for electrical resistivity testing gave the results shown in Table 8.

Table 8 — Electrical resistivity measurements

Test piece	Resistivity $\Omega \cdot \text{cm}$
1	$2,81 \times 10^{11}$
2	$3,54 \times 10^8$
3	$2,68 \times 10^{10}$
4	$2,75 \times 10^9$
5	$1,20 \times 10^{10}$

If the assumption were made that the data are normally distributed, the analysis would give the values:

- a) mean = $6,46 \times 10^{10}$;
- b) standard deviation, $s = 1,21 \times 10^{11}$.

6.3.4.2 Clearly, from the observed mean and standard deviation the data are not normally distributed, a fact which is readily confirmed by plotting the resistivities sorted into ascending value against the plot positions 17, 33, 50, 67, 83 on standard probability paper. A rapidly increasing slope with increasing probability value is observed instead of a straight line.

6.3.4.3 When the same data are plotted as the logarithm of resistivity (on log-probability paper for convenience) a very good straight line is achieved having a mean at approximately $10^{10} \Omega \cdot \text{cm}$. This is very close to the value of the geometric mean, $9,75 \times 10^9$, as would be expected from Annex B.

6.3.5 Other uses of the median

6.3.5.1 Measurements of the hardness of rubber compounds are probably the most frequently made of all the tests. The technique is not a highly precise one and results are usually quoted in whole numbers. Specifications of hardness are almost always given as the required nominal value ± 5 hardness degrees.

6.3.5.2 In this context, there is no mathematical advantage in calculating the mean from the set of five results (almost invariably) taken on a single test piece. Instead, the median, which can be deduced in a few seconds without a calculator, is the preferred measure of central tendency to use. For example, the results in Table 9 were derived on three compounds of nominal hardness 50 IRHD when tested in accordance with ISO 48.

Table 9 — Hardness measurements

Measurements in IRHD			
Result	Compound 1	Compound 2	Compound 3
1	50	52	49
2	51	53	48
3	49	51	47
4	51	50	47
5	50	55	46

6.3.5.3 The values given in Table 10 were calculated from those in Table 9.

Table 10 — Values calculated from hardness measurements

Measurements in IRHD			
Compound	Mean	Standard deviation, s	Median
1	50	1	50
2	52	2	52
3	47	1	47

In all cases, the median is within the 95 % confidence interval (see 7.2.1) of the mean and gives, for all practical purposes, the same result. In some instances, it may be necessary to carry out formal statistical testing in which case the standard deviation as well as the mean are required and there is no advantage in abstracting the median from the list. Also, if large numbers of replicate hardness values are to be processed, the median can become more tedious to determine than the mean.

NOTE The use of the median is purely pragmatic. There is no suggestion that the hardness is anything other than normally distributed about its mean.

7 Confidence limits and significant difference

7.1 Principles

As was stated in Clause 6, the mean and standard deviation derived for a given set of observations can only be estimates of the true mean and standard deviation of the whole population from which these observations are a random selection. To the extent that there is no systematic bias in the observations, the greater the number of results available, the less uncertainty there is over the accuracy of these estimates.

Unless the true mean (or standard deviation) can be deduced via some *a priori* reasoning, it is impossible to state how close a given calculated mean (or standard deviation) is to the true value. But it is possible to indicate with a known degree of uncertainty (the confidence level) that the true value will lie within a particular interval about that calculated value. The greater the degree of certainty required, the larger does this confidence interval become and the further apart are the limits (the confidence limits) of this interval.

Since any calculated measures of central tendency or dispersion are subject to uncertainty, when two such measures are compared, they cannot be expected to agree precisely. The difference between them becomes significant in the statistical sense only when it exceeds a limiting value which could have occurred, with a given probability, purely by chance.

It should be noted that a statistically significant difference between two measures of a property does not imply that the difference has any practical significance. The latter can only be judged in the context of the application being studied and the sensitivity of the application to the measured property. Thus, two compounds having tear strengths of 10 N/mm and 15 N/mm with standard deviations of 1 N/mm based on five results are significantly different at the 99 % level of confidence but, if the specification calls for a minimum tear strength of 25 N/mm, neither of them meets the specification and hence the difference between them is insignificant (and irrelevant) in practical terms.

Significance tests such as Student's *t*-test make several assumptions about the data being analysed, and an awareness of these assumptions is important in order to use the technique successfully. Firstly, the data must represent actual values, such as tensile strength or length; subjective values and grades cannot be treated in this way. Secondly, they must be normally distributed. Finally, they must be independent so that an error in one observation does not influence the error in another observation. If the data are not normally distributed, they can often be made to approximate to normality by a transformation (for example taking logarithms). The tests also assume that the two means being compared come from similar populations, i.e. their variances are not significantly different (see 7.2.2.2 for more information).

7.2 Methodology

7.2.1 Confidence limits and confidence intervals

7.2.1.1 For the mean

7.2.1.1.1 If the population mean, μ , and standard deviation, σ , are known, then it is a simple matter (see 6.2.1.1) to work out what the probability is that the mean of n results lies within, for example, three standard deviations of μ . Using the normal notation of $\bar{x}$ to represent the mean of the n results, then on 99,7 % of occasions

$$\mu - 3\sigma / \sqrt{n} < \bar{x} < \mu + 3\sigma / \sqrt{n} \quad (10)$$

By similar reasoning, if μ is unknown and σ is known, it is logical to assert that

$$\bar{x} - 3\sigma / \sqrt{n} < \hat{\mu} < \bar{x} + 3\sigma / \sqrt{n} \quad (11)$$

Thus, at a confidence level of 99,7 %, it is expected that the population mean would lie within the confidence interval of $\pm 3\sigma/\sqrt{n}$ about the sample mean. If a smaller confidence level, say 95 %, were chosen, the confidence interval would be $\pm 1,96\sigma/\sqrt{n}$.

7.2.1.1.2 It is unfortunate that the exact values of μ and of σ are unknown. Only the estimate s is known from the sample tested. For the same confidence level 99,7 %, the value $3\sigma/\sqrt{n}$ has to be increased to take account of this uncertainty in σ , the amount of the increase being dependent on the value of n . The factors used are given by the Student's t -distribution rather than the normal distribution (which is the limiting value of the t -distribution when n reaches infinity). Further consideration of the t -distribution is outside the scope of this International Standard and reference should be made to one of the many textbooks on statistics that are available. A selection of useful reference works is included in the Bibliography.

7.2.1.1.3 It is, therefore, assumed in the following clauses that the true mean and standard deviation are unknown and that only the estimates of the mean $\bar{x}$ and of the standard deviation s are known. For an exposition of situations in which one or other of the true parameters is known, reference should be made to ISO 2602 and ISO 2854.

The confidence limits for the mean are normally required for the 95 % and the 99 % confidence levels. In either case, the limits are given by the equations:

$$c_L = \bar{x} - (t_{\alpha}s)/\sqrt{n} \quad (12)$$

$$c_U = \bar{x} + (t_{\alpha}s)/\sqrt{n} \quad (13)$$

$$c_I = 2(t_{\alpha}s)/\sqrt{n} \quad (14)$$

where

c_L = lower confidence limit;

c_U = upper confidence limit;

c_I = confidence level.

It is therefore possible to be 95 % (or 99 %, etc.) confident that the true mean μ of the population does lie within this interval about the calculated mean (i.e. the estimated population mean).

In these equations, n is the number of observations and t is the appropriate Student's t -value, obtained from Table 11.

Table 11 — A selected table of Student's *t*-values

<i>n</i>	Confidence level		Confidence level	
	Two-sided case		One-sided level	
	95 %	99 %	95 %	99 %
	$t_{0,975}$	$t_{0,995}$	$t_{0,95}$	$t_{0,99}$
2	12,71	63,66	6,314	31,82
3	4,303	9,925	2,920	6,965
4	3,182	5,841	2,353	4,541
5	2,776	4,604	2,132	3,747
6	2,571	4,032	2,015	3,365
7	2,447	3,707	1,943	3,143
8	2,365	3,499	1,895	2,998
9	2,306	3,355	1,860	2,896
10	2,262	3,250	1,833	2,821
11	2,228	3,169	1,812	2,764
12	2,201	3,106	1,796	2,718
13	2,179	3,055	1,782	2,681
14	2,160	3,012	1,771	2,650
15	2,145	2,977	1,761	2,624
16	2,131	2,947	1,753	2,602
17	2,120	2,921	1,746	2,583
18	2,110	2,898	1,740	2,567
19	2,101	2,878	1,734	2,552
20	2,093	2,861	1,729	2,539
21	2,086	2,845	1,725	2,528
22	2,080	2,831	1,721	2,518
23	2,074	2,819	1,717	2,508
24	2,069	2,807	1,714	2,500
25	2,064	2,797	1,711	2,492
26	2,060	2,787	1,708	2,485
27	2,056	2,779	1,706	2,479
28	2,052	2,771	1,703	2,473
29	2,048	2,763	1,701	2,467
30	2,045	2,756	1,699	2,462
40	2,024	2,707	1,682	2,430
50	2,008	2,680	1,676	2,404
60	2,000	2,664	1,673	2,393
120	1,980	2,617	1,658	2,358
inf.	1,960	2,576	1,645	2,326

7.2.1.1.4 As noted in 6.2.3.2.3, the quantity $s/\sqrt{n}$ is called the standard error of the estimate (of the mean). It can thus be seen that to halve the confidence interval approximately four times as many observations have to be taken.

NOTE t is approximately constant except at very small values of n .

7.2.1.1.5 The value of t_α depends on the confidence level required, the number of observations (or, more precisely, the number of degrees of freedom) and whether a single-sided or a two-sided confidence interval is being sought.

In the simple cases being considered here, the number of degrees of freedom is $(n - 1)$.

7.2.1.1.6 A single-sided confidence interval is used when, for example, a comparison is being made between an observed mean value for a test, such as compression set, and the specification maximum (or minimum) to which it is being tested. This is because the only concern is with those values that might exceed (or not reach) the requirement and there is no interest in the values at the other side of the distribution function, these being those that conform to the specification limit. In this case, t_α is given in the tables under the columns for $t_{0,95}$ for the 95 % and $t_{0,99}$ for the 99 % confidence limits.

7.2.1.1.7 A two-sided confidence interval is used when, for example, it is necessary to know the interval within which the true mean could be expected to lie with the given degree of confidence. In this case, both sides of the distribution function are equally important and will contribute equally to the probability. Thus, for 95 % confidence the $t_{0,975}$ column is required and for 99 % confidence the $t_{0,995}$ column is required.

7.2.1.1.8 There are situations where it might be more convenient to calculate the value of Student's t -factor, for a given probability and number of degrees of freedom, rather than using reference tables. Provided that an error not exceeding 0,5 % of the true t -value is acceptable, then the following equation may be used:

$$t_\alpha = A + BC^{[1/(n-1)]} \quad (15)$$

where the constants A , B and C are as given in Table 12.

Table 12 — t -value constants

t	A	B	C
$t_{0,95}$	0,875 7	0,770 03	7,062 3
$t_{0,975}$	1,053 1	0,909 30	12,819 2
$t_{0,99}$	1,264 0	1,069 9	28,559 0
$t_{0,995}$	1,418 7	1,171 7	53,120 9

7.2.1.2 For the standard deviation

7.2.1.2.1 As in the case of the mean, the standard deviation calculated from a set of data can only be an estimate of the true standard deviation for the population as a whole and as such will have a measure of uncertainty associated with it. Confidence limits can therefore be set which, to a stated degree of confidence, contain the population standard deviation.

7.2.1.2.2 Unlike the confidence limits for the mean, the limits for the standard deviation are not symmetrical about the estimate s . This arises out of the fact that the standard deviation, unlike the mean, cannot be negative.

7.2.1.2.3 As discussed in 7.2.1.1.6 and 7.2.1.1.7 when considering the confidence limits for the mean, there are two cases to be considered. These are the single-sided case and the two-sided case depending on whether just an upper (or lower) limit, or both, is being considered.

$$c_{Us} = \left[\frac{ns^2}{\chi_{\alpha}^2} \right]^{1/2} \quad (16)$$

where c_{Us} is the upper confidence level for s .

$$c_{Ls} = \left[\frac{ns^2}{\chi_{1-\alpha}^2} \right]^{1/2} \quad (17)$$

where c_{Ls} is the lower confidence level for s .

The denominator in these equations comes from the chi-squared distribution function which is defined as the distribution of the sums of the squares of independent standardized normal variates. Some values are given in Table 13 and more comprehensive tables are available in the list of references.

For the single-sided case, $\alpha = 0,95$ or $0,99$ for the 95 % or 99 % confidence limits, respectively.

For the two-sided case, $\alpha = 0,975$ or $0,995$ for the 95 % or 99 % confidence limits, respectively.

As stated in 7.2.1.1.5, the number of degrees of freedom to be entered to find the chi-squared factor is $(n - 1)$.

Table 13 — A selected table of chi-squared values

<i>n</i>	Chi-squared for two-sided case				Chi-squared for one-sided case			
	95 %	95 %	99 %	99 %	95 %	95 %	99 %	99 %
	$\chi_{0,025}^2$	$\chi_{0,975}^2$	$\chi_{0,005}^2$	$\chi_{0,995}^2$	$\chi_{0,05}^2$	$\chi_{0,95}^2$	$\chi_{0,01}^2$	$\chi_{0,99}^2$
1	0,001	5,023	0,000 039	7,879	0,004	3,841	0,000 2	6,635
2	0,051	7,378	0,010	10,597	0,103	5,991	0,020	9,210
3	0,216	9,348	0,072	12,838	0,352	7,815	0,115	11,345
4	0,484	11,143	0,207	14,860	0,711	9,488	0,297	13,277
5	0,831	12,833	0,412	16,750	1,145	11,071	0,554	15,086
6	1,237	14,449	0,676	18,548	1,635	12,592	0,872	16,812
7	1,690	16,013	0,989	20,278	2,167	14,067	1,239	18,475
8	2,180	17,535	1,344	21,955	2,733	15,507	1,646	20,090
9	2,700	19,023	1,735	23,589	3,325	16,919	2,088	21,666
10	3,247	20,483	2,156	25,188	3,940	18,307	2,558	23,209
11	3,816	21,920	2,603	26,757	4,575	19,675	3,053	24,725
12	4,404	23,337	3,074	28,300	5,226	21,026	3,571	26,217
13	5,009	24,736	3,565	29,819	5,892	22,362	4,107	27,688
14	5,629	26,119	4,075	31,319	6,571	23,685	4,660	29,141
15	6,262	27,488	4,601	32,801	7,261	24,996	5,229	30,578
16	6,908	28,845	5,142	34,267	7,962	26,296	5,812	32,000
17	7,564	30,191	5,697	35,719	8,672	27,587	6,408	33,409
18	8,231	31,526	6,265	37,156	9,390	28,869	7,015	34,805
19	8,907	32,852	6,844	38,582	10,117	30,144	7,633	36,191
20	9,591	34,170	7,434	39,997	10,851	31,410	8,260	37,566
21	10,283	35,479	8,034	41,401	11,591	32,671	8,897	38,932
22	10,982	36,781	8,643	42,796	12,338	33,924	9,542	40,289
23	11,689	38,076	9,260	44,181	13,091	35,173	10,196	41,638
24	12,401	39,364	9,886	45,559	13,848	36,415	10,856	42,980
25	13,120	40,647	10,520	46,928	14,611	37,653	11,524	44,314
26	13,844	41,923	11,160	48,290	15,379	38,885	12,198	45,642
27	14,573	43,194	11,808	49,645	16,151	40,113	12,879	46,963
28	15,308	44,461	12,461	50,993	16,928	41,337	13,565	48,278
29	16,047	45,722	13,121	52,336	17,708	42,557	14,257	49,588
30	16,791	46,979	13,787	53,672	18,493	43,773	14,954	50,892

7.2.2 Significant difference

7.2.2.1 General

Closely related to the concept of confidence limits is that of significant difference, where a comparison needs to be made either between two means or two standard deviations. In the following subclauses, it is assumed that there are two sets of observations with the statistics shown in Table 14.

Table 14 — Statistics for observation sets

Observation set	Mean value	Estimated standard deviation	Number of observations
1	$\bar{x}_1$	s_1	n_1
2	$\bar{x}_2$	s_2	n_2

7.2.2.2 For the mean

The means cannot be regarded as likely to be different at the given confidence level α if

$$|\bar{x}_1 - \bar{x}_2| > t_\alpha S \quad (18)$$

where

$|\bar{x}_1 - \bar{x}_2|$ signifies the absolute value of the difference;

t_α is the value of Student's t (Table 11) for the two-sided case, entered for $(n_1 + n_2 - 2)$ degrees of freedom, being 97,5 for the 95 % confidence level or 99,5 for the 99 % confidence level;

S is the weighted standard error for the combined sets of observations, which is calculated by the equation:

$$S = \left[\frac{n_1 + n_2}{n_1 n_2} \times \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \right]^{1/2} \quad (19)$$

In most cases, $n_1 = n_2$ and the equation can be considerably simplified to the form:

$$S = \frac{\sqrt{(s_1^2 + s_2^2)}}{\sqrt{n}} \quad (20)$$

It is assumed in this analysis that there is no significant difference (at a stated level) in the standard deviations s_1 and s_2 . If such a difference is significant, then the two sets of observations cannot be considered to have come from the same population and hence their means would not usually be compared. The test for the significance of the difference between two standard deviations is given in 7.2.2.3.

7.2.2.3 For the standard deviation

7.2.2.3.1 The procedure is as follows:

- a) Determine the ratio of the variances using the equation:

$$F = (s_1/s_2)^2 \quad (21)$$

it being taken that $s_1 > s_2$.

- b) Consult Table 15 where the critical values for Snedecor's F -quotient at the 95 % and the 99 % confidence levels are given.
- c) Use one of the following:
- 1) $\alpha = 0,05$ for a single-sided 95 % confidence level;

- 2) $\alpha = 0,025$ for a two-sided 95 % confidence level;
 - 3) $\alpha = 0,01$ for a single-sided 99 % confidence level;
 - 4) $\alpha = 0,005$ for a two-sided 99 % confidence level.
- d) Establish the critical F -value for the $(n_1 - 1)$ degrees of freedom for s_1 and the $(n_2 - 1)$ degrees of freedom for s_2 by finding the intersection of the column with $(n_1 - 1)$ degrees of freedom and the row with $(n_2 - 1)$ degrees of freedom. If the calculated F -value is greater than this tabulated critical F -value, then the two standard deviations are different at the chosen confidence level.

Table 15 — Snedecor's F -values for selected degrees of freedom
(DF_1 = lesser degrees of freedom, DF_g = greater degrees of freedom)

a) F_{95} for one-sided test															
DF_1	DF_g														
	3	4	5	6	7	8	10	12	15	20	24	30	40	60	120
3	9,28	9,12	9,01	8,94	8,89	8,85	8,79	8,74	8,80	8,66	8,64	8,62	8,59	8,57	8,55
4	6,59	6,39	6,26	6,16	6,09	6,04	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66
5	5,41	5,19	5,05	4,95	4,88	4,82	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40
6	4,76	4,53	4,39	4,28	4,21	4,15	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70
7	4,35	4,12	3,97	3,87	3,79	3,73	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27
8	4,07	3,84	3,69	3,58	3,50	3,44	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97
10	3,71	3,48	3,33	3,22	3,14	3,07	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58
12	3,49	3,26	3,11	3,00	2,91	2,85	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34
15	3,29	3,06	2,90	2,79	2,71	2,64	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11
20	3,10	2,87	2,71	2,60	2,51	2,45	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90
24	3,01	2,78	2,62	2,51	2,42	2,36	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79
30	2,92	2,69	2,53	2,42	2,33	2,27	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68
40	2,84	2,61	2,45	2,34	2,25	2,18	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58
60	2,76	2,53	2,37	2,25	2,17	2,10	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47
120	2,68	2,45	2,29	2,17	2,09	2,02	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35

Table 15 (continued)

b) F_{95} for two-sided test															
DF _l	DF _g														
	3	4	5	6	7	8	10	12	15	20	24	30	40	60	120
3	15,44	15,10	14,88	14,43	14,62	14,54	14,42	14,34	14,25	14,17	14,12	14,08	14,04	13,99	13,95
4	9,98	9,60	9,36	9,20	9,07	8,98	8,84	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,31
5	7,76	7,39	7,15	6,98	6,85	6,76	6,62	6,52	6,43	6,33	6,28	6,23	6,18	6,12	6,07
6	6,60	6,23	5,99	5,82	5,70	5,60	5,46	5,37	5,27	5,17	5,12	5,07	5,01	4,96	4,90
7	5,89	5,52	5,29	5,12	4,99	4,90	4,76	4,67	4,57	4,47	4,42	4,36	4,31	4,25	4,20
8	5,42	5,05	4,82	4,65	4,53	4,43	4,30	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,73
10	4,83	4,47	4,24	4,07	3,95	3,85	3,72	3,62	3,52	3,42	3,37	3,31	3,26	3,20	3,14
12	4,47	4,12	3,89	3,73	3,61	3,51	3,37	3,28	3,18	3,07	3,02	2,96	2,91	2,85	2,79
15	4,15	3,80	3,58	3,41	3,29	3,20	3,06	2,96	2,86	2,76	2,70	2,64	2,59	2,52	2,46
20	3,86	3,51	3,29	3,13	3,01	2,91	2,77	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,16
24	3,72	3,38	3,15	2,99	2,87	2,78	2,64	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,01
30	3,59	3,25	3,03	2,87	2,75	2,65	2,51	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,87
40	3,46	3,13	2,90	2,74	2,62	2,53	2,39	2,29	2,18	2,07	2,01	1,94	1,88	1,80	1,72
60	3,34	3,01	2,79	2,63	2,51	2,41	2,27	2,17	2,06	1,94	1,88	1,82	1,74	1,67	1,58
120	3,23	2,89	2,67	2,52	2,39	2,30	2,16	2,05	1,94	1,82	1,75	1,69	1,61	1,53	1,43

c) F_{99} for one-sided test															
DF _l	DF _g														
	3	4	5	6	7	8	10	12	15	20	24	30	40	60	120
3	29,46	28,7	28,24	27,91	27,67	27,49	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22
4	16,69	15,98	15,52	15,21	14,98	15,80	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56
5	12,06	11,39	10,97	10,67	10,46	10,29	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11
6	9,78	9,15	8,75	8,47	8,26	8,10	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97
7	8,45	7,85	7,46	7,19	6,99	6,84	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74
8	7,59	7,01	6,63	6,37	6,18	6,03	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95
10	6,55	5,99	5,64	5,39	5,20	5,06	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00
12	5,95	5,41	5,06	4,82	4,64	4,50	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45
15	5,42	4,89	4,56	4,32	4,14	4,00	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96
20	4,94	4,43	4,10	3,87	3,70	3,56	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52
24	4,72	4,22	3,90	3,67	3,50	3,36	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31
30	4,51	4,02	3,70	3,47	3,30	3,17	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11
40	4,31	3,83	3,51	3,29	3,12	2,99	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92
60	4,13	3,65	3,34	3,12	2,95	2,82	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73
120	3,95	3,48	3,17	2,96	2,79	2,66	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53