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)

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

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

$a,b,c\ldots$	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 ith 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.