BSI

BS 5324 : 1976

UDC 678.4 : 620.1 : 519.2

BS 5324 : 1976

Guide to

Application of statistics to rubber testing

Guide pour l'application des statistiques aux essais des élastomères

Richtlinie für die Anwendung von Statistiken auf die Gummiprüfung

British Standards Institution

Gr 8

Contents

Foreword

In the preparation of this British Standard, under the authority of the Rubber Industry Standards Committee, use has been made of the guidance given in BS 2987 'Notes on the application of statistics to paper testing'.

The problems of sampling and the reliability of results obtained in rubber testing are clearly cases for the application of statistical methods, and in recent years a number of papers have been published indicating the growing interest in statistical methods in the rubber industry.

The results of rubber tests are usually subject to appreciable variation, due not only to experimental error, but also to variation in the raw materials and to the operation of many unknown or random factors in the manufacturing process. The interpretation of test results therefore requires some care and the chief value of statistical methods in this connection lies in their ability to replace purely subjective judgements of the data by objective criteria such as the tests described in the text. Statistical analysis of the data cannot add to the accuracy of an experiment by turning an uncertain result into a certainty, but it can enable the conclusions to be expressed more precisely in terms of a definite probability. Neither can statistical analysis be used to obtain meaningful results from experiments designed without reference to underlying statistical principles.

Although statistical methods have their foundation in mathematical theories of probability, a knowledge of such theories is not required for the successful application of these methods. Only straightforward arithmetical processes are involved and, in many cases, the calculation does not require the use of a machine. One or two shortcut methods of simplifying calculations are given.

The purpose of the present guide is to give only a brief outline of the subject, but with sufficient detail to enable the reader to carry out the simpler statistical calculations on his test data, such as those for ascertaining the precision of test means and the signficance of differences in test means (a list of formulae for easy reference is given in appendix A). For the more advanced techniques, such as detailed application of analysis of variance, which can be of great utility for the comparison of instruments, for compounding investigations and for the analysis of causes of variation in test readings, reference should be made to appropriate textbooks (see appendix B).

A secondary purpose of this guide is to secure greater uniformity in the presentation of test results and in the use of statistical terms and symbols (see clause **4** and glossary, appendix C).

The problems of sampling depend too much on particular conditions (such as manufacture and storage) to be capable of detailed treatment in this standard. Consideration has been given to these matters in the preparation of BS 1673 'Method of testing raw rubber and unvulcanized compounded rubber': Part 1 'Sampling' and the subject is given an extensive general treatment in BS 5309 'Sampling chemical products': Part 1 'Introduction and general principles' and Part 4 'Sampling of solids'.

In preparing this guide account has been taken of the work of Technical Committee TC 69, Applications of statistical methods, of the International Organization for Standardization (ISO). Attention has been particularly directed to ensure the correspondence of the definitions for statistical terms given in this guide with those approved by TC 69 and published in ISO/R 645. 'Statistical vocabulary and symbols. First series of terms and symbols'.: Part 1 'Statistical vocabulary' and ISO/R 1786 'Statistical vocabulary and symbols. Second series of terms and symbols'.

British Standard Guide to

Application of statistics to rubber testing

Section one. General

1. Scope

This standard has two purposes. One is to arouse the interest of persons engaged in rubber testing in the use of statistical methods, the other is to provide an easy source of reference to the formulae required for the treatment of data, so as to satisfy the general reporting clauses included in most of the rubber testing methods. It is complementary to the British Standards on the testing of rubber.

This standard is not intended to be used for quality control purposes. That subject is dealt with in BS 600 and BS 2564.

2. References

The titles of the British Standards referred to in this standard are listed on the inside back cover.

Section two. General principles

3. A practical problem from rubber testing

Imagine that a rubber manufacturer is accustomed to supplying a certain type of rubber A for a particular job. He knows from experience over a long period of time that this material is of satisfactory modulus, but recently he has developed rubbers B, C and D at an attractive cost per kilogram. As a safeguard against complaints, he decides to test the rubbers before attempting to pass them on to his customers. He carries out tensile tests, ten for each rubber (because he knows how misleading three results can be), under comparable conditions and gets the following results.

Modulus at 300 % elongation (MPa)

A	B	C	D
19.1	19.1	18.4	21.9
18.3	19.2	23.8	20.2
17.6	20.6	21.0	17.5
21.7	19.2	17.0	16.6
17.7	18.7	25.0	11.5
20.7	23.3	18.4	21.8
19.7	22.0	27.5	20.5
19.3	21.8	27.5	18.4
16.3	18.0	20.0	18.8
20.0	19.0	23.0	18.2

The purpose of these tests is to show up differences of modulus of the four rubbers, and probably the first thing the manufacturer will do with the figures is to run his eye up and down the columns. This may be sufficient to answer the questions, but, if the figures are difficult to compare, he will calculate the means or average values of the sets of ten results.

	A	B	C	D
Mean	19.04	20.09	22.16	18.64

In doing this, although he perhaps neither knows nor cares, he has applied statistical methods of analysis to the sets of figures. He has deduced a derived quantity or statistic which gives information in a succinct form about the readings.

What do these mean values tell him? One person may conclude from the four means that rubbers B and C are of higher modulus than A and rubber D is of lower modulus. Another may be more cautious. He knows that any test on rubber is subject to considerable variability (that is why he has done ten tests instead of one) and there is therefore some uncertainty about the mean values. He may conclude that, since the test average of C is 16 % higher than that of A, C is of appreciably higher modulus, and B possibly of higher modulus than A. The result for D is about 3 % lower and, although he may have doubts about the significance of this result, he will perhaps reject rubber D, just to be on the safe side.

The mean value is not enough. The variability of the readings has to be considered. In the unrealistic case where all ten readings on rubber A happened to be 19.04 and those on B 20.09 it could be confidently asserted that B is of higher modulus than A. In practice, some variability is to be expected, and the greater this is the more doubtful does the conclusion become. As the variability increases further, it becomes obvious that there is no reliable evidence of a difference of modulus between the two rubbers.

Ways of assessing variability. A rough impression can be obtained by running over the figures by eye, but conclusions drawn from this procedure are subjective. If the variability has to be expressed in a report the highest and lowest values can be used, or the range, but again the interpretation of these statements is open to doubt. The mean deviation (see glossary, appendix C) is a better expression of variability, but it is of limited usefulness.

The most satisfactory and efficient way of expressing variability is by means of the standard deviation, because it remains more constant than any other measure in repeated sets of figures. The range, for example, may vary relatively more than the standard deviation from one set of readings to another because it depends only on the two extreme values in each set. The standard deviation is more difficult to calculate than the range or mean deviation, but this is more than compensated by its versatility. It is fundamental to statistical analysis.

The coefficient of variation is closely related. It is the standard deviation expressed as a percentage of the mean value. It has the useful property of typifying a material where the standard deviation increases in proportion to the mean value. Under these circumstances the coefficient of variation is constant. This is often the case in rubber testing.

The test portion itself is another variable. Only a small part of the consignment of rubber A was tested and this test portion has been assumed to be representative of the rubber as a whole. If another test portion were to be tested, it is likely that a different mean value of tensile modulus would be obtained and this might lead to different conclusions. When a large series of similar test portions is tested, however, a definite pattern begins to arise. As more and more test portion means are taken, the pattern of the distribution of these means will more closely approach a smooth curve.

On the basis of this phenomenon, statements may be made about a single test portion which take account of their variability.

Why use statistics in rubber testing? In certain simple situations, conclusions may be reached without any reasonable doubt. It should be recognized, however, that personal judgement often plays a large part in arriving at a decision. This may be tempered by experience but, unless statistical methods are used, at least one guess is involved.

Statistics eliminates the guesses. Instead of a forthright statement that the tensile modulus of rubber A is 19.04 MPa (which may be proved wrong by the next sample tested), a conclusion firmly based on statistics might be 'that it is 95 % certain that the true tensile modulus of rubber A lies between 18.05 and 20.03 MPa'.

The question whether it is worthwhile to employ statistics in a particular problem depends entirely upon the importance of the answer. If the most objective conclusion is sought, for scientific or commercial reasons, then statistical methods should be considered.

4. Introduction to statistical methods

4.1 Distribution of results. A collection of values (for example, individual test results) relating to a specific property of the material being tested tends to arrange itself about the arithmetic mean in a manner which may be represented by a distribution curve (such as that illustrated in figure 1). When, as is often the case, the distribution of a large number of values approximates to a particular mathematical law, known as the Normal (Gaussian) distribution, a number of useful calculations can be made, e.g. of the proportion of test readings likely to differ by more than a given amount from the arithmetic mean (see appendix D, table 11 and figure 12).

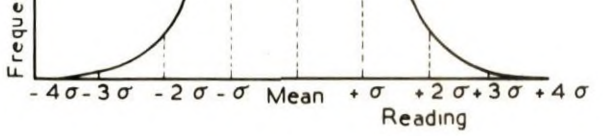


Figure 1. Normal distribution curve

Statistical tests are available to test for the departure from normality for a distribution, and details of some of these, together with the appropriate tables, can be found in *Biometrika tables for statisticians* (see appendix B, ref. 18) pages 61 to 63 and 183 to 184.

In some cases (such as in the Akron abrasion test) the distribution may be found to be markedly asymmetrical and to depart appreciably from the Normal form but, even in such cases, the distribution can often be converted into an approximately Normal form by a simple transformation of the readings, for example, by conversion of each reading to its logarithm (which has been found effective in the Akron abrasion test). Other simple transformations which have sometimes proved effective with experimental data are the square root or the reciprocal of the readings.

Even where the Normal distribution is not found for the individual readings, the distribution of the means of small groups of readings ($\geqslant 3$), such as compose the usual tests on rubber, nearly always approximates to the Normal form.

4.2 Sampling from a population. A set of data obtained in any single investigation can be regarded as a sample from a much larger collection, or population, of results, to which the given data belong. For example, the population might consist of the results of testing every piece of material in a given batch. The usual purpose of carrying out the tests is to estimate the properties of the population distribution from the data provided by the sample. Two properties that it is usually desired to know are the arithmetic mean and the standard deviation, which is a measure of variability.

4.3 Standard deviation (σ). The standard deviation is defined by taking the average, over the whole population, of the squared deviation of each value from the mean, and then taking the square root of this average. This definition of variability serves as a basis for statistical calculations, and various properties of the distribution may be deduced from it.

The frequency of individual readings decreases as they become further from the mean. Thus, for Normal distribution, approximately 1 in 20 readings may be expected to be beyond $\pm 2\sigma$ from the mean, 1 in 80 beyond $\pm 2.5\sigma$ and 1 in 370 beyond $\pm 3\sigma$ (see figure 1).

The distance from the mean that is exceeded by any given proportion of readings can be calculated from the formula $\pm u\sigma$, where u is a tabulated quantity called the standardized Normal deviate.

4.4 Coefficient of variation (v). It is sometimes convenient to express the standard deviation as a percentage of the arithmetic mean, in which case the value so calculated, known as the coefficient of variation, can be used as a measure of the relative variability of sets of readings with different mean values.

4.5 Estimation of the mean and the standard deviation from a sample. An estimate of the population mean is provided by the sample mean, $\bar{x}$, which is simply the arithmetic mean of the results composing the sample.

The best estimate of the population standard deviation is obtained by calculating the sample standard deviation, i.e. by summing the squares of the deviations of individual results from their mean, dividing this sum by one less than the total number of readings, and taking the square root of the quotient.

Thus,

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

where

x is the individual reading;
$\bar{x}$ is the arithmetic mean of the group of readings;
n is the number of readings in the group;
s is the estimate of standard deviation having the same units as x.

The purpose of dividing the sum of squares by $(n-1)$ (this $(n-1)$ is called the 'degrees of freedom') rather than by n is to correct a tendency to underestimate the population standard deviation when n is used as the divisor, especially in the case of small-sized samples. Reference should be made to the glossary in appendix C for a description of this.

Note that the symbol Σ denotes 'the sum of . . .'. In this case the sum is that of all individual values of $(x-\bar{x})^2$.

For methods of computation of s see the worked example in clause **5**.

The quantity s above is referred to as simply the standard deviation, the prefix 'estimate' being omitted, as understood. It is in several respects the best way of expressing the variability of results in a sample.

The population coefficient of variation, as estimated from the small sample, may be calculated as follows:

$$v = 100\,s/|\bar{x}|$$

where $|\bar{x}|$ is the numerical value of $\bar{x}$, regardless of the algebraic sign; that is, v is always positive.

Other measures of 'central tendency' and 'variability' exist. If further statistical processing of the test data is contemplated these are not usually used. They may, however, have application in straight reporting of results and for this reason they are discussed in **4.6.**

4.6 Other measures of central tendency. Other mean values, such as the harmonic and geometric means, can be regarded as the arithmetic mean of transformed data and need not be considered further.

4.6.1 *Median* (middle value). As a statistic this is based primarily on the order of observations and not on their values. It has a larger variance than the arithmetic mean. On the other hand, its value is not affected by an atypical value; moreover the median value may be considered to be more 'typical' than the mean value for skew distributions. In the case of a Normal distribution the mean and median are coincident.

The median may be calculated by arranging the results in ascending or descending order and crossing off the highest and lowest results together until only one (or two values) are left. If two values are left the median may be estimated as the mean of these values e.g.

Set 1. ~~8.2~~ ~~10.1~~ ~~13.7~~ 13.8 ~~14.0~~ ~~14.9~~ ~~14.9~~
median = 13.8

Set 2. 14.7 ~~17.8~~ 20.9 22.3 ~~23.8~~ ~~24.2~~
median = 21.6

4.6.2 *Mode* (most frequently occurring value). This value has the largest frequency density and is therefore intuitively the most typical of the central values. Its value is, however, difficult to estimate, even if the shape of the underlying distribution is mathematically known.

4.6.3 *Mid-range.* This value is defined as the average of the highest and lowest values. It has some drawbacks when used as a measure of central tendency since it is based only on the two extreme observations, which are usually the least reliable; furthermore, its efficiency decreases rapidly for larger series of observations.

From a theoretical point of view this is not a very meaningful statistic, especially when taking into account the fact that a large number of theoretical distributions have an infinite range.

4.7 Other measures of variability. The scatter of data can be described according to three principles:

(a) measures of distance (e.g. the range);

(b) measures compiled from the deviations of every observation from some central value (e.g. mean deviation);

(c) measures compiled from deviations amongst all observations (e.g. mean difference).

4.7.1 *Range.* The range is the numerical difference between the highest and lowest of a set of readings. In a Normal population the average range of a number of measurements is related to the population standard deviation by a factor which depends on the number of readings. An estimate of the standard deviation is given by multiplying the range of a random sample of n observations from a Normal population, by the factor a_n. This is illustrated in table 1, which is an extract from Lindley, D.V. and Miller, J.C.P. *Cambridge elementary statistical tables* (C.U.P.) (tables 6 and 7).

Table 1. Factors for estimating standard deviation from a range of values in a Normal population

n	a_n
2	0.8862
3	0.5908
4	0.4857
5	0.4299
6	0.3946
7	0.3698
8	0.3512
9	0.3367
10	0.3249

For estimating the standard deviation, the range is almost as good as any other method when the number of readings is not larger than about ten and the readings follow a distribution which is not far from the Normal. However, since the exact calculation of the sample standard deviation is relatively quick on ten results or fewer, this would normally be carried out.

With larger numbers of readings, the use of the range is not recommended, except as a check against gross errors in calculating the standard deviation. It is better to use the standard deviation, which utilizes all the information in the data, instead of relying on the extreme readings.

4.7.2 *Interquartile range.* This is the distance between two values between which the central half of the observations fall; e.g. 50 % of the observations fall within the so-called 'interquartile range'.

Interquartile ranges are only used when very large sample sizes are available; they reveal very little, however, about the way the bulk of the observations is condensed inside this range.

4.7.3 *Mean deviation.* This is the mean of the absolute value of the deviations around the central value. As it is excluded from more advanced statistical techniques, its usefulness is restricted; nor does it give a more comprehensible measure of the scatter than standard deviation. Obviously the value obtained will depend on the type of central value chosen (mean, median, or mode).

4.7.4 *Mean difference.* This is the average of the absolute values of the differences of all possible pairs of observations. Again because of the absolute nature of mean difference, this measure of dispersion is excluded from statistical techniques while it virtually offers little more than the standard deviation.

4.8 Unexpectedly high or low results. It is often required to know whether an apparently high or low result in a set of data should be rejected.

As a general principle, a reading should never be rejected unless there are grounds outside the data for so doing. An occasional high or low reading is to be expected, to a varying degree, in the distribution of any rubber property, due to the nature of the rubber itself, and rejection of such readings may lead to a distorted picture of the distribution; the standard deviation, in particular, will be underestimated.

On the other hand, high or low readings are sometimes caused by errors of operation or damaged test pieces, in which case the readings can be legitimately rejected; or again, they may be due to flaws in the test piece, such as thin spots or pin holes. The rejection may then depend on the purpose of the test. For example, in testing the air permeability of rubber sheeting the presence of occasional high readings might be taken to indicate the presence of pin holes in the rubber. If the rubber is being tested as a functional material, then such readings should be retained; but if the air permeability is only required as an index of the rubber's technical properties, possibly in relation to the compounding, then the reading might fairly be rejected.

When a rejectable reading is suspected, and there are no confirmatory indications, statistical criteria outlined by W.J. Dixon (*Processing data for outliers: biometrics*, 1953, 9 (1) 74) can sometimes be used to decide whether or not the reading may be legitimately rejected. An alternative method is the use of Grubb's test described in *Annal. math. stat.* 21 (1) (March 1950).

The problem is approached statistically by considering the sample values to be drawn from a mixed population, consisting of a main population together with a small proportion of values from another, interfering population, which differs from the first in mean, or variance, or in both, so as to yield occasional high (or low) sample values. The object of the test criteria is to remove the more extreme members of the interfering population in order to be able to calculate relatively unbiased estimates of the mean and standard deviation of the main population. It is assumed that both populations follow the Normal distribution law.

The definition of the test criteria, and their critical values, are given in table 2. In this table, x_n is the extreme value to be tested, x_{n-1} is the nearest neighbour to x_n, and so on, the readings being arranged in order of magnitude from x_1 to x_n.

If the value of r calculated from the sample data exceeds the tabulated critical value, the extreme reading may be rejected with a fair degree of confidence (see clause **5** for a worked example).

The critical values correspond to a significance level of 5 %. This means that, if the data follow a pure Normal distribution, only one extreme reading in 20 will, on the average, be rejected unnecessarily.

The method should not be used for test methods where there is reason to think that the distribution departs from the Normal form. Data of this type, however, may often be transformed to nearly Normal form (see **4.1**).

To summarize, no reading should ever be rejected unless either there is evidence of a definite source of error affecting the reading, or the reading is found rejectable according to the above statistical test; the latter may be applied only when the data are known to follow a Normal distribution law, or have been transformed to this form. Indiscriminate use of the rejection test will result in underestimation of the standard deviation.

Table 2. Criteria for rejecting outlying values by Dixon's test

Number of readings, n	Criterion	Critical values
5		0.642
6	$r_n = \frac{x_n - x_{n-1}}{x_n - x_1}$	0.560
7		0.507
8		0.554
9	$r_n = \frac{x_n - x_{n-1}}{x_n - x_2}$	0.512
10		0.477
11		0.576
12	$r_n = \frac{x_n - x_{n-2}}{x_n - x_2}$	0.546
13		0.521

4.9 Confidence interval for the mean, when the population standard deviation (σ) is known. The means of groups of n readings (for example 5, 10 or 20) may be expected to vary about the true mean in a Normal manner, but they will naturally be less variable than individual readings. The distance from the true mean within which a single observed mean may be expected to lie, with a given probability, is given by:

$$\pm u\sigma/\sqrt{n}$$

where

u is the standardized Normal deviate and has the value 1.96 for a probability of 95 % and

$\sigma/\sqrt{n}$ is the standard error of the mean.

Values of the standardized Normal deviate corresponding to other levels of probability may be obtained from tables of the Normal distribution.

The limits given above refer to the distribution of observed means about a known true mean. By using the same limits in an inverse sense, it is possible to state, with a good degree of confidence, the range, about the observed mean, within which the true but unknown mean lies. When used in this sense, the limits are called 'confidence limits', the degree of confidence being expressed by a probability, which is usually taken as 95 %. Thus the 95 % confidence limits of a population mean are given by:

$$\bar{x} \pm \frac{1.96\sigma}{\sqrt{n}}$$

Confidence limits are the usual methods of expressing the precision of an estimate. The use of 95 % limits ensures that the statement that the true mean lies within these limits will be correct, on the average, in 19 cases out of 20. It will be noted that the precision, or closeness of the limits, depends on the number of readings and their standard deviation and also on the probability level (usually 95 %) selected.

The lines in figure 2 show the distance of the 95 % confidence limits from the mean, for typical small numbers (n) of tests. These lines are based on the above formula, but, to make them of more general application, the distance has been expressed as a percentage of the mean, and plotted against coefficient of variation, in the place of standard deviation.

The limits are expressed as the percentage distance from the sample mean, which is based on n readings; these limits are only applicable when the population standard deviation, σ, is known. Limits in the case where the standard deviation, σ, is estimated by s from the sample are given in figure 4.

4.10 Number of tests required for a given precision of the mean, when the population standard deviation (σ) is known. The formula $u\sigma/\sqrt{n}$ given in **4.9** may be used to calculate the number of readings required to estimate the mean with a given degree of precision. For example, in a rubber test having a coefficient of variation of 7 %, the number of readings required to estimate the mean within ± 5 % (95 % confidence limits) is given by the equation:

$$(1.96)\ (7)/\sqrt{n} = 5$$

whence $n = 7.5$.

Therefore eight readings should be sufficient to give the required precision.

4.11 Significance of difference between means of two tests, when the population standard deviations (σ) are known. In comparing the means of two tests it is necessary to have some criterion to decide whether the difference is likely to have arisen by chance, or indicates a real difference between the means of the populations from which the samples were drawn. This criterion is called the least significant difference.

It is so chosen that, as long as no real difference exists between the means of the two populations, there is only a small probability of it being attained or exceeded. This probability, or 'significance level', is usually taken as 1 in 20, but circumstances may required other level to be used.

When the observed difference between means exceeds the least significant difference it is unlikely that the respective true means are the same. The conclusion that a real difference between them exists will be incorrect in fewer than 1 case out of 20 (if the 5 % level of significance is chosen).

A difference that is less than the least significant value, on the other hand, should be taken to indicate not that the true means are necessarily the same, but merely that the difference is 'insignificant', that is, small enough to escape detection by the given data.

For the general case, where the number of readings in the two tests differ, the least significant difference is given by the following formula:

least significant difference between means

$$= 1.96\sigma\sqrt{(1/n_1 + 1/n_2)} \text{ (5 \% level of significance)}$$

where

n_1 and n_2 are the numbers of readings in the two tests, and

σ is the known standard deviation of the population.

For the usual case, where each test contains the same number (n) of readings, the formula becomes:

least significant difference between means

$$= 1.96\sigma\sqrt{(2/n)} \text{ (5 \% ievel of significance)}$$

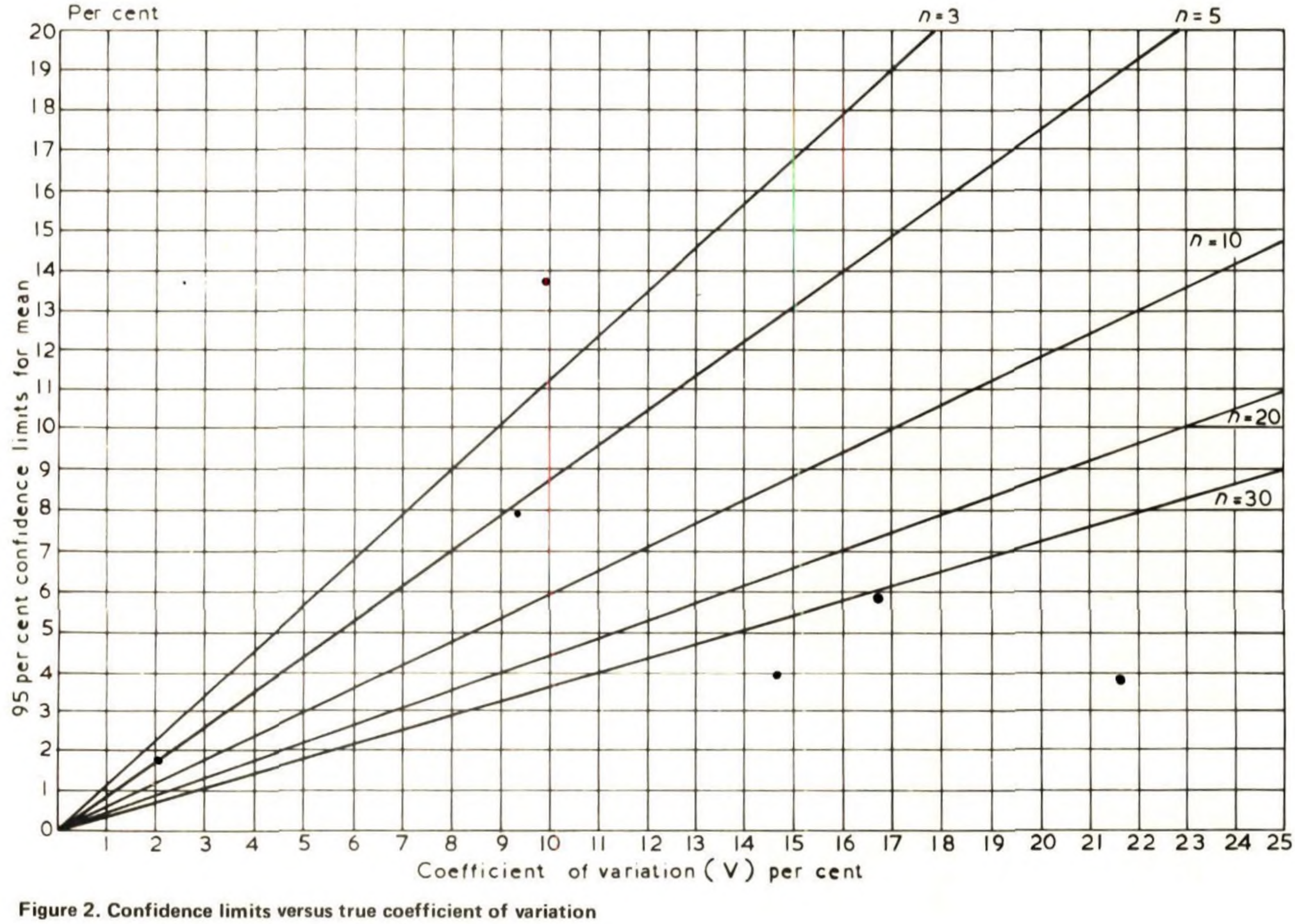


Figure 2. Confidence limits versus true coefficient of variation

The least significant difference, according to the last formula, is plotted in figure 3 for n = 5, 10, 20 and 30 readings in each test. As in figure 2 the coefficient of variation is used in place of the standard deviation; the least significant difference is accordingly expressed as percentage of the mean.

When the standard deviation, s, is estimated from the sample, figure 5 should be used.

4.12 Confidence interval for the mean, when the population standard deviation (σ) is unknown. In **4.10** and **4.11** the precision of the mean and the least significant difference between means have been calculated on the assumption that the variability is accurately known; this is not usually the case, however. It is often not sufficient to rely on 'typical' values of variability based on previous data. The precision of a test mean or the least significant difference between two test means should preferably be calculated from the estimate of standard deviation provided by the test readings themselves. This estimate of standard deviation is subject to some error, since the number of readings available is generally small (10 or 20) and, to allow for this error, the foregoing formulae require some modification, as follows.

Confidence limits for the mean $= \bar{x} \pm ts/\sqrt{n}$

where

t is given, for any chosen probability level, by Student's t-distribution (double-sided) (tables of which are included in nearly all the references; an abbreviated version is also given in table 12, appendix D);

s is the estimate of the standard deviation, calculated from the test readings;

n is the number of readings in the group.

Values of t depend on the number of readings, but, in the tables, the different values of t are usually arranged according to a number γ which is called the number of degrees of freedom. In the present case, the number of degrees of freedom is one less than the number of readings, so that the tables should be entered at $\gamma = (n - 1)$.

For calculating 95 % limits, the appropriate values of t, corresponding to typical values of n, are as follows:

t = 2.78 for n = 5 readings;

t = 2.26 for n = 10 readings;

t = 2.09 for n = 20 readings.

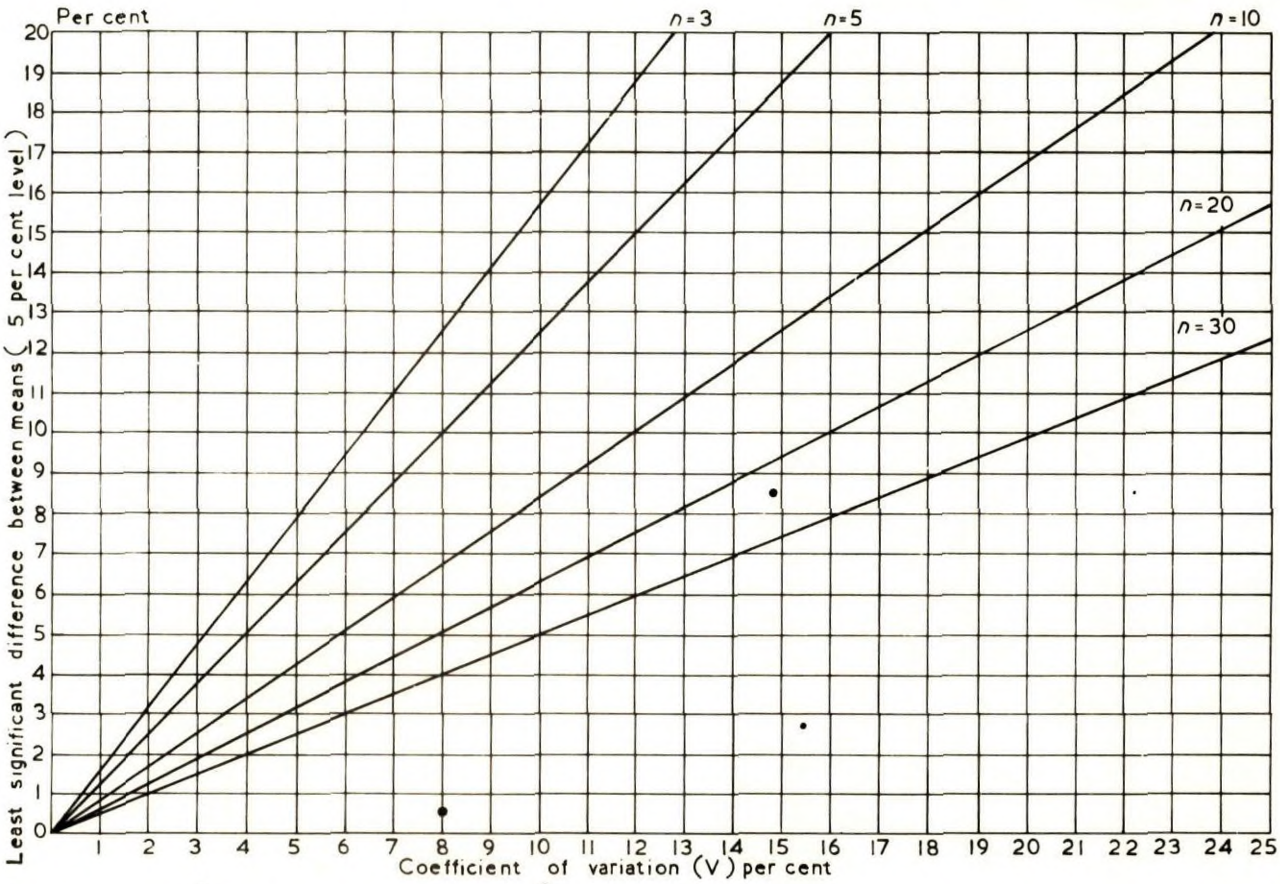


Figure 3. Real difference versus true coefficient of variation

It will be noticed that, owing to the uncertainty in the value of the standard deviation estimated from a small number of readings, the values of t are somewhat larger than the corresponding values of u given by the Normal distribution (u = 1.96, see **4.9**) and the 95 % limits for the mean are correspondingly wider.

Figure 2 refers to the precision of the mean when the standard deviation is known from independent data, as may be the case in quality control. Figure 4 gives the corresponding information when the standard deviation has to be calculated from the test data, which will be the more common occurrence.

4.13 Significance of difference between means of two tests when the population standard deviations are known (Student's *t*-test). For the general case, where the number of readings in the two tests differ:

Least significant difference between means

$$= ts\sqrt{(1/n_1 + 1/n_2)}$$

where

t is given by Student's t-distribution (double-sided) (the tables should be entered at $\gamma = n_1 + n_2 - 2$);

s is a pooled estimate of standard deviation, calculated from the two sets of readings (see **4.15** and worked example in clause **5**);

n_1, n_2 are the numbers of readings in the two test.

For the usual case, where each test contains the same number (n) of readings:

Least significant difference between means

$$= ts\sqrt{(2/n)}$$

(The t tables should be entered at $\gamma = 2n - 2$).

For a 5 % level of significance, typical values of t are:

t = 2.31 for n = 5 readings in each test;

t = 2.10 for n = 10 readings in each test;

t = 2.02 for n = 20 readings in each test.

Figure 5 gives the corresponding information to that in figure 3.

Before this particular significance test can be applied, the validity of obtaining a pooled estimate of the standard deviation should be checked. This is done by applying a test for significance of difference between variances, described in **4.14**. If the variances are not significantly different, then the pooled estimate of the standard deviation is derived as shown in **4.15**, and the calculated value of the pooled standard deviation (s) is used in the above formulae.

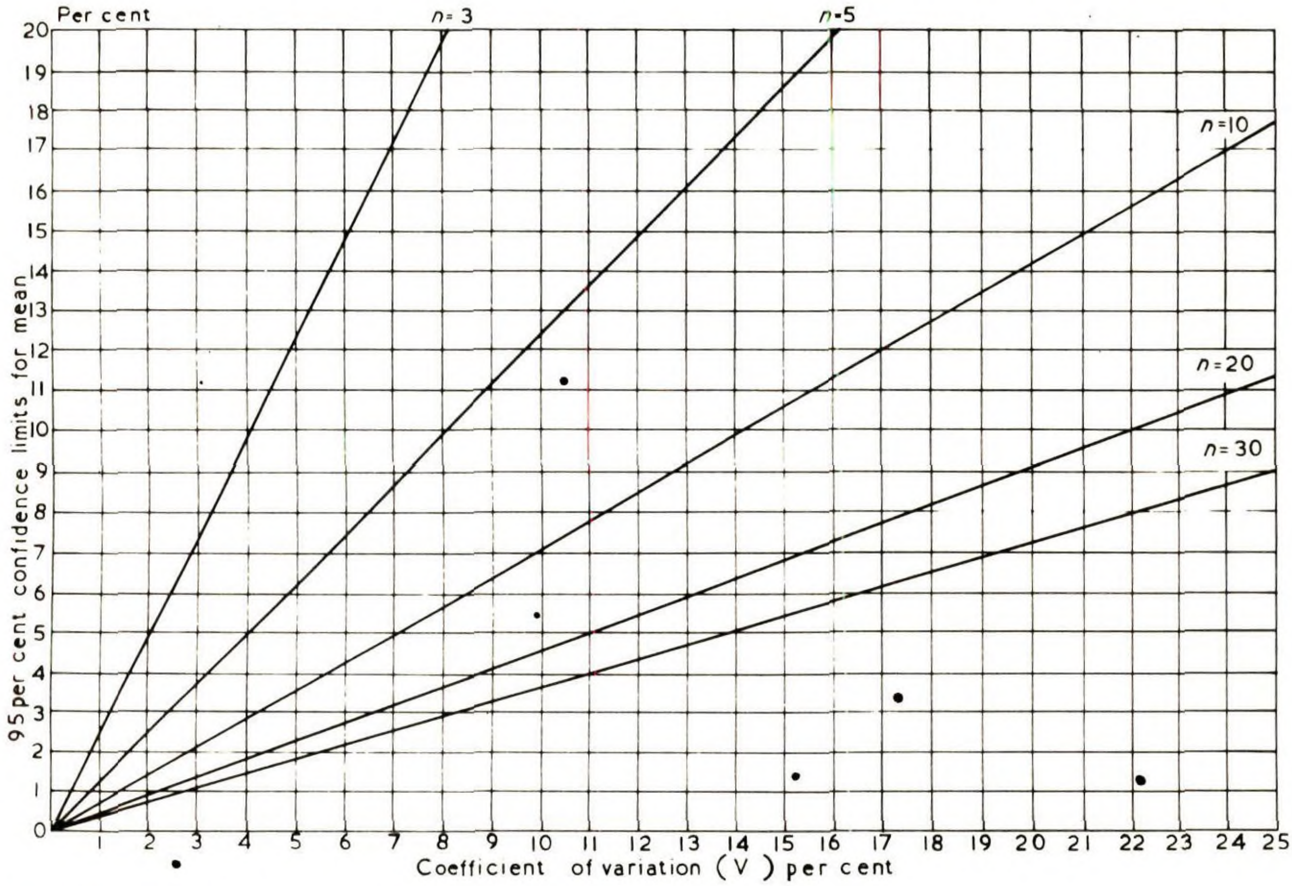


Figure 4. Confidence limits versus estimated coefficient of variation

If the F-test of **4.14** is significant, indicating that the population variances are really different, but it is still essential to compare the means, then a very approximate test would be still to combine the variances, calculate s, and test for a significant difference in the means in the same way as above. *However it is stressed that the test in this case would only be a rough guide.*

4.14 Significance of difference between variances. On some occasions it is not a comparison of mean value which is of major importance, but a comparison of variances. For example, if two operators are each carrying out a number of repeat determinations on a particular sample it will often be useful to assess whether there is any real difference in the repeatability of the two operators; i.e. whether the results of one operator are much more closely grouped around his mean value than the results of the other operator are around his.

(It may also be of interest to test whether there is any bias between the two operators, i.e. whether the means are significantly different; but this is not considered in this subclause.)

Let the consideration be confined to the following specific situation:

Operator 1 carries out n_1 determinations and the variance of the n_1 results is s_1^2.

Operator 2 carries out n_2 determinations and the variance of the n_2 results is s_2^2.

The ratio $\frac{s^2 \text{ (larger)}}{s^2 \text{ (smaller)}}$ is calculated, where s^2 (larger) is the larger of s_1^2 and s_2^2, and s^2 (smaller) is the smaller of s_1^2 and s_2^2.

Appendix D, table 13 is then entered at the appropriate point,

where

γ_1 is the number of degrees of freedom on which s^2 (larger) has been based;

γ_2 is the number of degrees of freedom on which s^2 (smaller) has been based.

If the ratio $\frac{s^2 \text{ (larger)}}{s^2 \text{ (smaller)}}$ is greater than the appropriate value of the tabulated F-distribution (appendix D, table 13) then it can be stated with 95 % confidence that the population variance σ_L^2 (of which s^2 (largest) is an estimate) is greater than the population variance σ_s^2 (of which s^2 (smallest) is an estimate).

If the ratio $\frac{s^2 \text{ (larger)}}{s^2 \text{ (smaller)}}$ is less than the appropriate F-distribution value, then we have not sufficient evidence to refute the hypothesis that the two population variances are equal.

This test (known as the F-test or 'variance ratio' test) could also be applied in the comparison of the variances of various methods of analysis.

For strict correctness this type of test should be applied to the variances s_1^2 and s_2^2 mentioned in **4.15.**

The ideas which have been discussed so far are sufficient for the solution of many common types of problem, and in clause **5** they are applied to an example. Clause **5** also contains hints and precautions on computing which will lighten the arithmetical work.

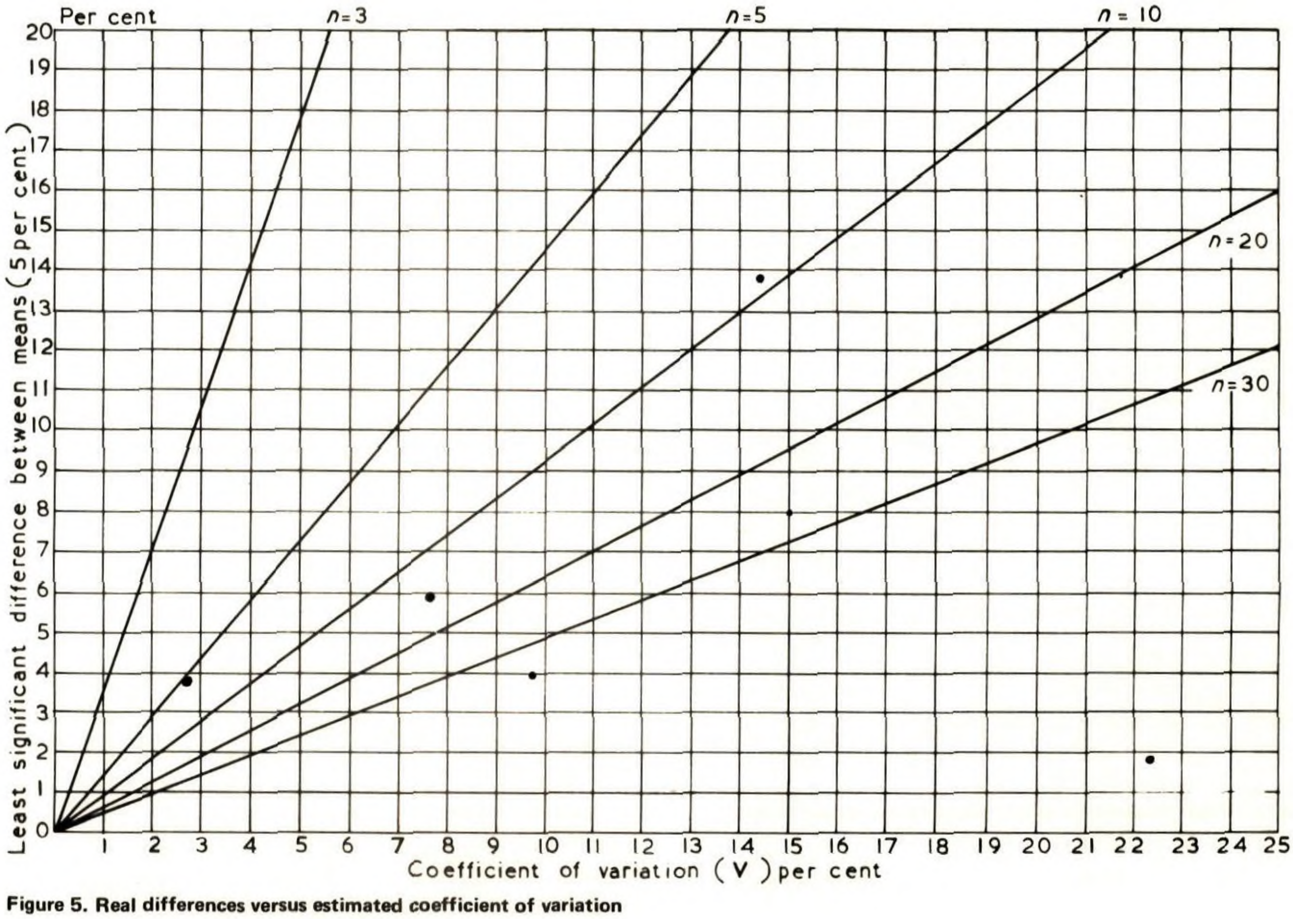


Figure 5. Real differences versus estimated coefficient of variation

4.15 Pooled estimate of standard deviation from two sets of readings. The 'pooled' estimate should be formed only if the difference in variances as described in **4.14** is not significant. When two sets of readings are available, the best 'pooled' estimate of standard deviation is given by the following formula, rather than by the mean of the two separate estimates:

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

When the numbers of readings (n_1 and n_2) in the two tests differ, more weight is given, by this formula, to the test containing the greater number of readings.

When the standard deviations have already been calculated separately for each sample, giving s_1 and s_2 for example, it is seen that the last formula gives the 'pooled' estimate of standard deviation as:

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

Sometimes the variability will be very different in the two sources of material supplied and pooling of the two estimates may not be justified. It would be beyond the scope of this guide to deal here with the procedure to be followed in this but it may be remarked that, provided that the two samples are of nearly the same size, the error arising from pooling will never be serious.

4.16 Single sided significance tests and confidence intervals. A 'double sided significance test' is one in which significant differences in either direction are of importance. A situation in which only a difference in one specific direction is of importance requires what is called a 'single-sided significance test'. The type of test to be used depends completely on the practical application, and is always decided before the experimentation.

In the situations described in **4.11** and **4.13** the general question being asked is: 'Is there a significant difference between . . . . ?' In that case a double-sided significance test was appropriate. However, in practical applications a typical question might be: 'Is the mean of this batch significantly less than the specification?' In such a situation it is of no importance if the batch mean is significantly greater than the specification; only if it is significantly below the specification value need action be taken, and so a single sided test should be used.

(a) If the population standard deviation (σ) is known the following applies.

When a single sided test is appropriate and the test is being carried out at the 5 % significance level, the 1.96 which occurs in the formulae in **4.11** for 'least significant difference between means', is replaced by the value of 1.64. Also a single-sided 95 % confidence interval for the population mean can be calculated. This can be expressed as 'it is 95 % certain that the population mean is greater than $\bar{x} - \frac{1.64\sigma}{\sqrt{n}}$',

or as 'it is 95 % certain that the population mean is less than $\bar{x} + \frac{1.64\sigma}{\sqrt{n}}$',

depending on which is the statement relevant to the practical application.

(b) If the population standard deviation (σ) is unknown the following applies.

When a single-sided test is appropriate, the t-distribution values referred to in **4.13** would be replaced by the corresponding figure from the single-sided t-distribution tables. For a 5 % level of significance and a single sided test, the appropriate t-value would be

t = 1.86 for n = 5 readings in each test

t = 1.73 for n = 10 readings in each test

t = 1.68 for n = 20 readings in each test.

Also a single-sided 95 % confidence interval for the population mean can be calculated. This can be expressed as 'it is 95 % certain that the population mean is greater than $\bar{x} - \frac{ts}{\sqrt{n}}$',

or as 'it is 95 % certain that the population mean is less than $\bar{x} + \frac{ts}{\sqrt{n}}$',

again depending on which statement is relevant to the practical situation. The value of t in these confidence intervals would be that selected from the 5 % column of the single-sided distribution, with the appropriate degrees of freedom.

5. Application of statistical tests to the example in clause 3

The figures in the example in clause **3** can now be treated by the statistical methods developed in clause **4**.

The standard deviation is calculated by the formula in **4.5**.

It is not necessary to calculate and list the individual deviations $(x - \bar{x})$ and their squares in order to estimate the standard deviation. Computation is simplified by the use of the identity:

$$\Sigma(x - \bar{x})^2 = \Sigma x^2 - (\Sigma x)^2/n,$$

where

Σx^2 is the sum of squares, and

$(\Sigma x)^2/n$ is the correction factor,

so that the estimate of standard deviation becomes:

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

The application of this formula will be clear from the worked example. The computing proceeds by the following steps.

(a) Arithmetical work may be reduced by the use of an arbitrary origin, chosen at or near the general level of the readings. In the worked example below, this is done by subtracting 20 from each reading before proceeding with the calculations. The standard deviation is unaffected by this procedure.

(b) Compute the sums Σx and Σx^2, remembering that x may now be the distance of the reading from the arbitrary origin. For the squares of the readings, the use of tables is recommended (for example, Barlow's tables of squares) if a calculating machine is not available. In some cases, as in parts of the example, the numbers are small enough to be squared in the head.

(c) Before proceeding further Σx and Σx^2 should be checked. Rather than merely repeating what has already been done, with the risk of repeating the same error, the following identity can be used:

$$\Sigma(x+1)^2 \equiv \Sigma x^2 + 2\,\Sigma x + n.$$

Calculate the sum of the squares of $(x+1)$ and check that it agrees with the right-hand side of the identity, using the values of Σx^2 and Σx already obtained.

(d) Calculate $\bar{x} = \Sigma x/n$. This, added to the arbitrary origin, is the mean value.

(e) Compute $\Sigma(x-\bar{x})^2$ from the identity given. When working out $(\Sigma x)^2/n$, retain at least as many significant figures as in Σx^2.

(f) Compute the standard deviation s from the formula given.

In some situations it might be necessary to compare all the rubbers in an attempt to assess whether there is in fact a real rubber-to-rubber variation. This is best done by means of 'analysis of variance', a technique which is described in clause **7**. If consideration is restricted to rubbers A and B, then the least significant difference is calculated by the formula in **4.13**. The pooled standard deviation for the pairs being tested is worked out and the value of t used is 2.10 for a 1 in 20 level of significance and 10 readings in each column (see **4.15**). The results are then tabulated as follows.

Rubber	A		B
Mean modulus at 300% elongation (MPa)	19.04		20.09
Observed difference		1.05	
Standard deviation	1.604		1.738
Least significant difference		1.57	

The difference between A and B is less than the significant value and one can only conclude that, on the basis of the tests done, the difference between rubber B and rubber A could have arisen by chance and that there are therefore no statistical grounds for asserting that rubber B is of significantly higher modulus than rubber A.

The difference between the means of A and B is about 1 MPa and the manufacturer still has the feeling, in spite of the conclusion of the ten tests, that a true difference of modulus is being concealed, which he would be able to detect if a large number of tests were carried out.

In pursuit of this idea, he does ten more tensile tests each on second samples of rubbers A and B and gets the following readings:

A 17.0, 26.3, 20.0, 16.4, 18.4, 19.7, 19.8, 18.6, 15.8, 17.0

B 20.1, 19.9, 21.1, 20.2, 17.5, 20.0, 17.6, 20.0, 28.8, 27.4

If these are combined with those of A and B above, the mean values for twenty tests become:

A = 18.97 and B = 20.68

and the standard deviations:

A = 2.341 and B = 2.935.

Had the standard deviations remained constant, the larger number of tests would have meant the least significant difference would have been smaller. However, in this case the apparent increase in variability means that the least significant difference is actually greater, 1.70 as compared to 1.57 previously. Nevertheless, the actual difference between the means is 1.71 and thus we can conclude, on the basis of twenty tests on each rubber, that rubber B is actually of higher modulus than rubber A.

It will be noticed that in the example it was concluded that there was no reliable evidence of any difference between A and B; this conclusion was modified when twenty tests were carried out. A difference of about 1 MPa escaped detection when only ten tests were done. If the manufacturer had been interested only in detecting differences of 2 MPa ten tests would have been sufficient. For a 0.5 MPa difference, more than twenty tests would have been required and in general the number of tests should be related to the minimum size of difference in which one is interested.

Sample D (see clause **3**) contains one very low reading. In **4.8** a simple rule was given for rejecting abnormal results. The rejection criterion r_{11} in this case takes the value 0.495, which is above the limit of 0.477 for 10 tests. The extreme value 11.5 is therefore rejectable. Whether it should be rejected will depend upon considerations discussed in **4.8**.

From the treatment given in this clause to a manufacturer's typical problem, it will be seen that the application of statistical methods enables one to draw objective conclusions as opposed to those which rely on the intuition and experience of the tester. This can be claimed as one of the principal advantages of using statistical methods.

Other more advanced techniques may be used with advantage in rubber problems, but are beyond the scope of this guide, which should be regarded merely as an introduction to the subject. The methods of analysis of variance, for example, are particularly useful in experimental work and, for detailed discussion of such methods, reference should be made to the textbooks (see, however, clause **7**). One way in which the analysis of variance can be of use is in investigating the causes of variation in test results, which may be attributable to the testing instrument itself, to its methods of operation or to irregularity in the rubber sample. In a suitably designed experiment, such methods of analysis will enable these separate factors to be distinguished with a minimum of effort.

Although the statistical methods that have been discussed are very useful for a number of purposes, it should be pointed out that no amount of statistical analysis can compensate for inaccuracies due to unsuitable or unrepresentative choice of rubber samples. Where lack of uniformity exists in a batch of rubber to be sampled, statistical methods can be of use in helping to determine the best methods of sampling.

6. Design and analysis of experiments

6.1 Introduction. During the past few years the industrial chemist has come to accept, although at times somewhat reluctantly, that a mathematical treatment of his data may bring to light useful facts not disclosed by his normal approach to the problem. This bridge having been crossed, chemists and statisticians have been working closely together in most large industrial firms for a considerable time. However, when the subject of 'design of experiments' is mentioned many technologists still feel that this field of work is solely theirs by right. It is not yet fully appreciated that this is an area in which it is most vital that the technologist and statistician should be working intimately together. The statistical approach to the problem is intended to supplement the technologist's knowledge and experience, and not to replace it.

Often, on the plant scale, designed experimentation is out of the question owing to:

(a) lack of plant availability for experimental programmes; or

(b) lack of control of process variables.

However, these restrictions seldom apply to laboratory work and do not always apply to plant work. Wherever experimental planning is possible it is essential that it is undertaken before the work is commenced, for otherwise by the time the problem reaches the statistician all the experimental effort has been spent and he has put before him data which is rarely ideal for analysis. More discussion in the initial stages of the problem could often prevent unnecessary work or point the experimentation in the most appropriate direction. Another common feature of problems of this type is that discussion has taken place, a plan of experimentation has been agreed, but somewhere along the line, probably due to extremely high or low results (or 'intuition'), the technologist is diverted from his plan and the data never reach a statistician. It may be necessary to modify an experimental plan as results become available, but even then hidden benefits will usually emerge from a statistical analysis of the resulting data.

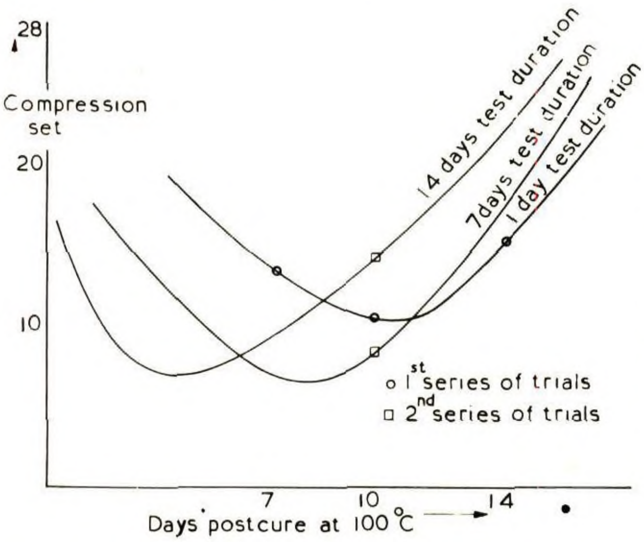


Figure 6. Compression set versus postcure for different durations of test

6.2 'Classical' approach. In most practical situations more than one variable is worthy of consideration, and when this is so it is of the utmost importance that the technologist is guided away from the idea of varying one factor at a time. The following example illustrates well the dangers involved if careful attention is not given to the initial design of experiments.

Suppose we are considering the effect of two factors, duration of test and postcure time at 100 °C, on the compression set of a vulcanizate, and each factor is to be considered at three distinct levels.

If we decide to vary one factor at a time, then we might conduct the experiment as follows.

(a) Fix the test duration at one day and carry out three tests with postcure times of 7, 10, and 14 days respectively.

(b) From figure 6 we see that the lowest compression set value occurs under these conditions at 10 days.

(c) Now fix the postcure time at 10 days, and carry out two further tests with test durations seven and 14 days respectively.

(d) The best result (i.e. lowest compression set) is obtained with a test duration of seven days, so that we might assume that the best possible combination of the two factors is:

(1) test duration of seven days;

(2) postcure time at 100 °C of 10 days, with a resultant compression set value of 8 %.

With this approach to the problem we would obviously not have obtained the best combinations of the levels of the factors for a minimum compression set.

As shown in figure 6, the effect of test duration is different at the different levels of postcure; this effect is known as an 'interaction' between the two factors, and is illustrated in table 3.

Table 3. Factor interaction on compression set

Postcure days	Test duration days		
	1	7	14
7	13	7	9
10	10	8	14
14	15	18	22

Very important interactions would be completely overlooked in varying one factor at a time. In fact, in the plan outlined in the example, the resultant conclusion could have been worse; we have assumed that for each experiment we obtained the true value of compression set, whereas in practice our result would deviate from this due to experimental error.

6.3 Statistical approach to experimental design. The following steps represent the various stages in a typical problem analysis involving experimental design.

(a) One of the first points to emerge from the initial discussion should be the question which one is hoping to answer from this programme of work. The primary concern in most situations is to locate the optimum operating conditions, taking into account both the quality of product and the operating costs.

(b) In order to reach the goal described in (a), it will be of the utmost importance to assess what is significantly affecting the product quality. At this stage a decision has to be made, mainly based on knowledge and experience of the process, as to which factors are going to be deliberately varied in the investigation. It might be seen at this point that a number of variables can be measured, but cannot be specifically controlled, in which case the values of these variables should be recorded for each experiment.

(c) Any data already available concerning this particular process can be of great value. It may be possible to show, from a statistical analysis of this previous data, that some of the factors discussed in (b) are of no importance. Valuable information concerning the experimental error may also be obtained, and consequently this may reduce the size of the experimental programme to be planned.

(d) From the information obtained in (b) and (c) it is necessary to decide the number of levels of each factor, and this in turn will partially govern the number of experiments necessary. If, from the knowledge of the process, it is thought that the effect of a factor will be curved rather than linear, then usually this factor will be considered at three levels at least.

Ideally one would hope to cover many factors, each at many levels, in very few experiments, and still obtain precise information on the relative importance of the factors. In practice this is obviously impossible, and usually the number of experiments which can be carried out governs the number of factors and levels that will be considered. Factorial experimental designs are in frequent use, and in this situation all the possible experimental conditions are covered. For example, if there are three factors (sulphur, accelerator, cure time) with two, three and four levels respectively, then the full set of experiments would be:

Experiment	Sulphur	Accelerator	Cure time
1	0.5	0.5	10
2	0.5	0.5	15
3	0.5	0.5	20
4	0.5	0.5	25
5	0.5	1.0	10
6	0.5	1.0	15
7	0.5	1.0	20
8	0.5	1.0	25
9	0.5	2.0	10
10	0.5	2.0	15
11	0.5	2.0	20
12	0.5	2.0	25
13	2.0	0.5	10
14	2.0	0.5	15
15	2.0	0.5	20
16	2.0	0.5	25
17	2.0	1.0	10
18	2.0	1.0	15
19	2.0	1.0	20
20	2.0	1.0	25
21	2.0	2.0	10
22	2.0	2.0	15
23	2.0	2.0	20
24	2.0	2.0	25

The total number of experiments is given by $2 \times 3 \times 4 = 24$, i.e. the product of the number of levels of the factors. This is by no means the only possible design, but it is both comparatively simple and in frequent use.

(e) In step (d) the number of factors and their levels have been chosen, and hence the total number of experiments determined. *If careful attention is not given to the problem at this point it is possible that useful information will be lost, because often in practice there are nuisance factors confusing the issue.* For example, in the above plan of 24 experiments it could be that only 12 experiments could be done using one batch of raw material, and the other 12 would have to be done using another batch possibly with a different quality. If the 12 experiments with low sulphur were done on batch 1, and the other 12 experiments on batch 2, then the effect of sulphur would be totally confused with the variation in the quality of raw material. This effect is known statistically as 'confounding'. Any factors such as raw material, operator or machine, which may alter during the experimentation, should be brought to light at this stage, and the best way of splitting the design into the required blocks can be decided; but this will generally necessitate the help of a statistician.

(f) Another practical difficulty which is often encountered is the situation where many variables are to be considered, but only a very limited number of experiments can be carried out. If we had, for example, seven factors each at two levels, very few technologists would even remotely consider doing the 128 experiments necessitated by a factorial design. However, in such a situation, by carefully choosing a sub-set of say 32 experiments, much valuable information could still be obtained. Designs such as these are known as fractional factorials and again these will usually require the advice of a statistician. The two ideas of confounding and fractional factorial experimentation are most useful, and most easily applied, when all the factors under consideration have the same number of levels.

(g) When the experiments have been split into any blocks or groups which are required, then the order of the experiments within each block should be randomized, for example by drawing lots.

(h) Once all the experiments have been completed the analysis of the data begins. The methods of analysis of the data are as numerous as the experimental designs, and it is outside the scope of this clause to detail the various techniques. However, to illustrate the analysis of a factorial experimental design, one example is considered in detail. This example is followed through the steps (a) to (h) as described.

Example. This simple example, using only factors at two levels, is chosen to avoid having to introduce sums of squares and other simple but lengthy calculations.

For each experiment the property which is to be measured is the resistance value, and the purpose of this programme of work is to maximize the resistance value.

Two types of additive, A and B, are being considered, and the other factors which it is thought might be affecting the resistance value are the humidity level and the amount of a particular additive. Each of these factors has two distinct levels which are to be covered in the work.

Humidity level	20 % relative humidity
	60 % relative humidity
Amount of additive	0.1 %
	1 %

It is thought that there should be no other variables causing variation in the resistance, and hence blocking of the experiments is not required.

From previous repeat measurements of the resistance value the measurement standard deviation σ has been estimated, and so it is not necessary to carry out any repeat experiments or measurements within this programme.

Summarizing the situation, we have three factors each at two levels, giving us a total of 2^3 (= 8) experiments. The order of the experiments is randomized, e.g. by drawing lots, and the experimental programme is detailed below, together with the results.

Experiment number	Additive type	Humidity level %	Amount of additive %	Resistance $\log_{10} \Omega$
1	A	20	0.1	16
2	B	20	1	7
3	A	60	1	8
4	B	60	0.1	12
5	A	20	1	9
6	B	20	0.1	15
7	B	60	1	6
8	A	60	0.1	11

In this situation an estimate of the effect of any of the factors can be obtained by calculating the average value of the resistance at each level of the factor and then evaluating the difference between these two averages.

Effect of additive type

= average result for additive type A

— average result for additive type B

$= \frac{1}{4}(16 + 8 + 9 + 11) - \frac{1}{4}(7 + 12 + 15 + 6)$

$= 11 - 10 = 1.0$

Effect of humidity

= average result with high humidity

— average result with low humidity

$= \frac{1}{4}(8 + 12 + 6 + 11) - \frac{1}{4}(16 + 7 + 9 + 15)$

$= 9\frac{1}{4} - 11\frac{3}{4} = -2.5$

Effect of amount of additive

= average result with high amount

— average result with low amount

$= \frac{1}{4}(7 + 8 + 9 + 6) - \frac{1}{4}(11 + 12 + 15 + 16)$

$= 7\frac{1}{2} - 13\frac{1}{2} = -6.0$

In this particular type of experimental design the interaction between say additive type and humidity level would be defined as follows:

Interaction between additive type and humidity level

= ½ (average result for additive type A, humidity 60 %

— average result for additive type A, humidity 20 %)

— ½ (average result for additive type B, humidity 60 %

— average result for additive type B, humidity 20 %)

$$= \frac{1}{2}\left(\frac{8+11}{2} - \frac{16+9}{2}\right) - \frac{1}{2}\left(\frac{12+6}{2} - \frac{7+15}{2}\right)$$

$= -1.5 + 1.0 = -0.5$

With similar definitions for the other interactions these can now be evaluated.

Interaction between additive type and amount of additive

$$= \frac{1}{2}\left(\frac{8+9}{2} - \frac{16+11}{2}\right) - \frac{1}{2}\left(\frac{6+7}{2} - \frac{12+15}{2}\right)$$

$= -2.5 + 3.5 = 1.0$

Interaction between humidity level and mount of additive

$$= \frac{1}{2}\left(\frac{8+6}{2} - \frac{7+9}{2}\right) - \frac{1}{2}\left(\frac{12+11}{2} - \frac{16+15}{2}\right)$$

$= -0.5 + 2.0 = 1.5$

In a comparatively simple type of experiment where we have n factors each at two levels, and consequently do 2^n experiments, the standard error of each of the estimated effects or interactions is $\frac{\sigma}{\sqrt{2^{n-2}}}$.

In our example $n = 3$ and let us assume $\sigma = 1$; hence the standard error of any of the estimated effect is $\frac{1}{\sqrt{2}}$.

The 95 % confidence interval for any effect is given by

$$\left(\text{estimated effect} - 1.96 \frac{.1}{\sqrt{2}}\right) \text{ to}$$

$$\left(\text{estimated effect} + 1.96 \frac{.1}{\sqrt{2}}\right)$$

i.e. estimated effect ± 1.38

and if this confidence interval does not include zero then the effect is significant.

Any effect which is assessed as significant has less than a 1 in 20 chance of being zero.

In our example the effects which are significant are:

(1) effect of humidity;

(2) effect of amount of additive;

(3) interaction between humidity and amount of additive.

The three-factor interaction between the factors of additive type, humidity and additive could be considered in a similar manner.

We can summarize our conclusions as follows.

(1) By decreasing the humidity from 60 % to 20 % we expect to increase our resistance by 2.5 ± 1.38 on average.

(2) By decreasing the amount of additive from 1 % to 0.1 % we expect to increase our resistance by 6.00 ± 1.38 on average.

(3) The significant interaction is best illustrated by the following two way table for the resistance.

		Additive	
		0.1%	1.0%
Humidity	20%	15.5	8.0
	60%	11.5	7.0

Humidity has little effect at the higher level of additive but a much larger effect at the lower level of additive.

(4) The suggestion following the analysis would be that the additive type is unimportant, but that further work with lower levels of humidity and additive may be worthwhile.

6.4 Summary. The discussion has so far been of a very general nature, to illustrate that:

(a) statistically designed experimentation is worthwhile;

(b) many practical difficulties can be encountered and overcome;

(c) discussion between statistician and technologist should occur very early in the project wherever possible.

No attempt has been made to describe the many types of experimental design that may be applicable.

7. Analysis of variance

7.1 Introduction. In most rubber testing situations the variability in the quality measurements is contributed to by a number of sources of variation. For example, there is often present:

(a) variation in the quality of one or more of the raw materials;

(b) variation in process operating conditions at one or more stages;

(c) errors arising in sampling and testing the rubber.

Analysis of variance is a technique which can be used to isolate and estimate the effect of those sources of variation which are having a significant effect on the quality measurements.

The more sources of variation of types (a), (b) and (c) that are present, the more complex will be the analysis of the resultant data, and for many commonly occurring large problems a computer programme becomes virtually essential.

To illustrate the basic ideas and steps involved in analysis of variance, a simple example will be considered.

7.2 Example to illustrate analysis of variance. In order to investigate the quality of a consignment of rubber, four samples have been taken and five repeat determinations of the compression set have been carried out on each sample.

The results obtained are as follows.

Sample 1	Sample 2	Sample 3	Sample 4
2	3	6	5
3	4	8	5
1	3	7	5
3	5	4	3
1	0	10	2

Using this data it is required to estimate the true average compression set of the rubber, and also to assess whether or not there is a significant sample-to-sample variation.

The first obvious step is to calculate the mean for each sample to give us some idea of the variation between samples.

Mean for sample 1 = 2
Mean for sample 2 = 3
Mean for sample 3 = 7
Mean for sample 4 = 4

The question that is now under consideration is whether or not we could reasonably expect such differences in the sample means to arise (owing to the presence of the testing error and the limited number of tests per sample) if the true compression set for all the samples was really the same.

In order to do this, an analysis of variance table is constructed which breaks down the total variation in the results into the components for which sources of variation can be identified.

A measure of the total variation present in the data is given by summing the squares of the deviations of each individual result from the overall mean of the results.

If x represents an individual result and $\bar{x}$ represents the overall mean, then $\Sigma (x - \bar{x})^2$ is a measure of the total variation.

A measure of the 'within-samples variation' or 'variation due to testing error' is given by summing the squares of the deviations of each result from its own sample mean.

A measure of the 'between-samples variation' is given by squaring the deviation of each sample mean from the overall mean, multiplying each squared deviation by the number of results on that particular sample, and summing the resultant figures.

In our example the overall mean, $\bar{x}$, = 4.

The total sum of squares (i.e. a measure of the total variation)

$$= (2-4)^2 + (3-4)^2 + (1-4)^2 + \ldots\ldots + (2-4)^2$$

$$= 116$$

Within-samples sum of squares (i.e. a measure of the within-samples variation)

$$= (2-2)^2 + (3-2)^2 + (1-2)^2 + (3-2)^2 + (1-2)^2$$

(from sample 1)

$$+ (3-3)^2 + \qquad\qquad + (0-3)^2$$

(from sample 2)

$$+ (6-7)^2 + \qquad\qquad + (10-7)^2$$

(from sample 3)

$$+ (5-4)^2 + \qquad\qquad + (2-4)^2$$

(from sample 4)

$$= 46$$

Between-samples sum of squares (i.e. a measure of the between-samples variation)

$$= 5(2-4)^2 + 5(3-4)^2 + 5(7-4)^2 + 5(4-4)^2$$

$$= 70$$

An analysis of variance table can be constructed as follows.

Source of variation	sum of squares	degrees of freedom	mean square
Between-samples	70	3	23.3
Within-samples (testing)	46	16	2.9
Total	116	19	6.1

The above procedure has been followed with a view to explaining the concept of analysis of variance, i.e. the breaking down of the total variation into its constituent parts. In practice, however, if the calculations are to be done by hand or using a calculating machine the following method should be followed. It is easier to apply and will necessarily give identical results to the above calculations.

$$\text{Total sum of squares} = \Sigma x_{ij}^2 - \frac{(\Sigma x_{ij})^2}{n}$$

where

Σx_{ij}^2 is the sum of the squares of the individual results;

Σx_{ij} is the sum of the individual results;

n is the total number of results.

$$\text{Between-samples sum of squares} = \frac{\Sigma T_i^2}{p} - \frac{(\Sigma x_{ij})^2}{n}$$

where

T_i is the sum of the results on the ith sample;

p is the number of tests on each sample.

The within-sample sum of squares is then obtained by subtracting the between-sample sum of squares from the total sum of squares.

In the sample considered earlier in this clause,

$$\Sigma x_{ij}^2 = (2^2 + 3^2 + 1^2 + 3^2 + 1^2 + 3^2 + 4^2 + \dots + 3^2 + 2^2) = 436$$

$$\Sigma x_{ij} = (2 + 3 + 1 + 3 + 1 + 3 + 4 + \dots + 3 + 2) = 80$$

$$n = 20$$

$$T_1 = 10 \quad T_2 = 15 \quad T_3 = 35 \quad T_4 = 20$$

$$p = 5$$

Then

$$\text{Total sum of squares} = 436 - \frac{80^2}{20} = 116$$

Between-sample sum of squares

$$= \frac{(10^2 + 15^2 + 35^2 + 20^2)}{5} - \frac{80^2}{20} = 70$$

$$\text{Within-samples sum of squares} = 116 - 70 = 46$$

In any similar analysis of variance in which we have say n experimental results available, the total number of degrees of freedom will be $(n - 1)$.

For any factor which is considered to be a possible cause of variation in the test result, the number of degrees of freedom associated with the variation attributable to that factor will be $(r - 1)$, where r is the number of levels of that particular factor. Hence in this particular example, where four samples are being considered, the degrees of freedom associated with the between-samples variation is three.

The within-samples degrees of freedom can be obtained by (total degrees of freedom – between-samples degrees of freedom) = 16.

In each instance the 'mean square' (which is similar to a variance) is sum of squares divided by degrees of freedom. If the sampling variance is denoted by σ_s^2 and the testing variance by σ_t^2, then it can be shown that

(a) 2.9 (= 46/16) is an estimate of σ_t^2;

(b) 23.3 (= 70/3) is an estimate of $\sigma_t^2 + 5\sigma_s^2$.

In (b) the 5 occurs because five tests have been carried out on each sample. If the number of tests per sample was not constant, the appropriate number could be calculated, but from a more complex formula.

The ratio $\dfrac{\text{between-samples mean square}}{\text{within-samples mean square}}$ is now calculated and if this ratio is less than the appropriate tabulated F-distribution value (see table 13 appendix D; in this instance the F-distribution value would have $\gamma = 3, \gamma_2 = 16$) then it can be reasonably assumed that the two mean squares under consideration are in fact estimates of the same thing. From the last paragraph this can only be so if $\sigma_s^2 = 0$. In that situation the 'best' estimate of the testing variance would be the total mean square.

If the ratio when evaluated is greater than the appropriate F-distribution value, this suggests that the between-samples mean square and the within-samples mean square cannot reasonably be estimating the same thing. This can only be so if $\sigma_s^2 \neq 0$.

In our particular example the ratio is $\dfrac{23.3}{2.9} = 8.1$ and the F-distribution 5 % point for 3 and 16 degrees of freedom is 3.24. Hence we feel at least 95 % certain that $\sigma_s^2 \neq 0$; that is, there is present real sample-to-sample variation. In this situation the best estimate of σ_t^2 is 2.9, and an estimate of σ_s^2 is obtained from

$$\frac{(\text{between-samples mean square} - \text{within-samples mean square})}{5}$$

$$= 4.1$$

Often it would be necessary not only to estimate from the available data the true average compression set but also to calculate a confidence interval within which we feel sure the true average will lie.

If p tests have been carried out on each of r samples, and σ_t^2 and σ_s^2 estimated as above, then an approximate 95 % confidence interval for the true average compression set is given by

$$\bar{x} \pm 2\sqrt{\frac{\sigma_t^2}{rp} + \frac{\sigma_s^2}{r}}$$

In our particular example this confidence interval is 4.0 ± 2.16.

In some instances it is required to assess whether certain of the differences between the samples or items are in fact statistically significant. If data are only available on two samples then the method detailed in clause **5** should be followed. If the data are from more than two samples, and assuming that the analysis of variance has indicated that there is real sample-to-sample variation present, the situation is more complex and a statistician should be consulted.

7.3 Summary. Only a very simple example has been used to illustrate the idea of analysis of variance. As many different forms of experimental design are available so analysis of variance tables are constructed in many different but appropriate ways. In practically every instance certain basic steps can be identified.

(a) There is the need to identify the possible causes of variation. These may be:

(1) only sampling and testing;

(2) the main effects and interactions of many factors;

(3) the dependence on some uncontrolled variable.

(b) The breakdown of the total variation into the parts that can be attributed to each source of variation.

(c) The construction of the analysis of variance table, including the correct allocation of the degrees of freedom.

(d) The application of the appropriate F-distribution tests to assess the significance of the various sources of variation.

The book *Design and analysis of industrial experiments* edited by O.L. Davies, published by Oliver and Boyd, contains many excellent worked examples of various forms of analysis of variance.

8. Application of regression analysis

8.1 Introduction. Regression analysis is concerned with the estimation of the relationship between some response variable (e.g. tensile strength) and some other variable or group of variables (e.g. quality of raw materials or operating conditions) which may be affecting the response. It is an attempt to explain the variation in the response in terms of the variation in the variables. Regression techniques can be used to assist in the analysis of carefully planned and controlled experimental laboratory work, or they can be applied to data arising from routine manufacture where the planning of experiments and control of the variables are often virtually impossible.

If data is available only on a response and one other variable then often valuable information can be obtained from a simple graphical plot (see figure 7). Confining our attention to the data points marked x, it appears obvious that 'age of material' has a definite effect upon 'duration of flexing life'. However, if the four extra points marked ⊡ become available we are much less certain whether the two variables are really related. A more precise method of assessment than the graphical plot and visual judgment is now required. In practical situations many variables are usually present (see table 4) and faced with such a mass of figures visual inspection cannot help us, while unfortunately graph plotting is confined to two dimensions. In such situations as these regression techniques are invaluable.

Analysis of data by regression techniques can roughly be broken down into three stages:

(a) the assessment of which variables are really having an effect upon the response;

(b) the estimation of the best relationship between these variables and the response;

(c) the calculation of the variation in the response explained by the equation, and the calculation of the residual or unexplained variation.

8.2 Correlation coefficient. The simplest way to assess whether or not there is a significant linear relationship between two variables is to calculate the 'correlation coefficient'. The data required to calculate this parameter are a series of 'pairs' of observations on the response (e.g. tensile strength) and the other variable (e.g. percentage of natural rubber in polymer) which is often called the independent variable.

If y represents the response and x represents the independent variable, we may have n pairs of observations

$$(y_1, x_1)\ (y_2, x_2)\ \dots\ (y_n, x_n).$$

The correlation coefficient (usually denoted by r) between the response and the independent variable is given by:

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2\,\Sigma(y-\bar{y})^2}}$$

$$= \frac{\Sigma xy - \dfrac{(\Sigma x)(\Sigma y)}{n}}{\sqrt{\left(\Sigma x^2 - \dfrac{(\Sigma x)^2}{n}\ \Sigma y^2 - \dfrac{(\Sigma y)^2}{n}\right)}}$$

From the way in which it is calculated, the correlation coefficient will never be less than −1 and never greater than +1.

If all the points lie on some line with positive slope, then $r = 1$, and if all the points lie on some line with negative slope, then $r = -1$.

Usually the correlation coefficient will not take either of these extreme values but will lie somewhere in between. The sign of the evaluated correlation coefficient will indicate whether the 'best' straight line through the points has a positive or negative slope. Using table 5, it is also possible to assess whether or not a calculated correlation coefficient indicates a significant relationship between the two variables, and this test depends on the number of data points available. For example, if we had only six pairs of observations available we could say that we were 95% certain that there was a real relationship between the variables if the calculated correlation coefficient was less than −0.811 or greater than +0.811. However, if 72 data points were available we would be 99% certain that there was a real relationship between the variables if the calculated correlation coefficient was less than −0.302 or greater than +0.302. If the calculated correlation coefficient does indicate a significant relationship *it is very important to realise that this is not necessarily a cause and effect relationship.*

If there are a large number of data points available, then a value of r as low as 0.1 might be classed as significant. The implication of this would be that on average as x increases then y also increases, but that whereas we feel confident that x and y are varying together there is still a great deal of unexplained variation in the response y.

If, on the other hand, when the significance test is carried out the value of r does not indicate a real linear relationship, then it is still possible that there could be some curved or more complex relationship between y and x. If this is thought to be the case multiple regression analysis or some more complicated curve fitting-technique should be used.

To summarize this subclause, if there are only two variables of interest, the correlation coefficient can be quickly calculated and used to give a good indication of the significance of the relationship between these two variables. It does *not*, however, enable us to estimate the best straight line through the available data points. The correlation coefficient between duration of flexing life and age of material is evaluated in the example in **8.7**.

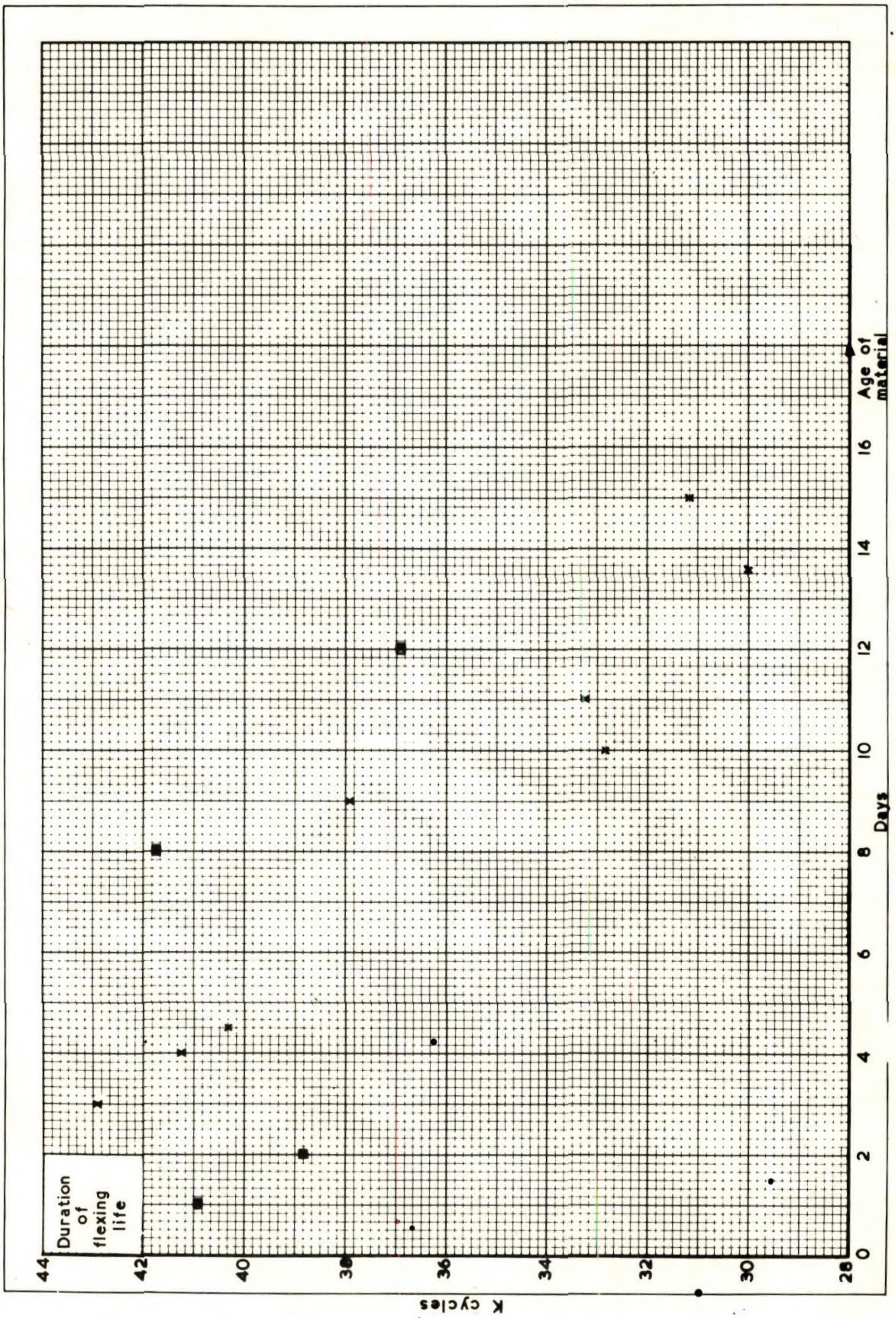


Figure 7. Dependency of flexing life on age of material

8.3 Simple linear regression. Interest is still confined to the situation where there are only two variables, but now it is required to estimate the 'best' straight line through the points. We are interested in the equation

$$y = a + bx$$

where

y is the response (e.g. tensile strength);

x is the so-called 'independent' variable (% natural rubber in polymer);

a and b are parameters to be estimated from the data:

b is the slope of the line (the estimated increase in y for a unit increase in x);

a is the intercept on the y axis (the value of y when $x = o$ if this is meaningful).

We would not expect this equation to predict the exact value of y from the corresponding value of x, but we hope that it would help to quantify the suspected relationship between x and y. We require the values of a and b that in some way give us the fitted line as close as possible to the data points. The vertical distance of a data point from the fitted straight line is known as its 'residual' (see figure 8). The values of the coefficients a and b are so chosen as to minimize the sum of the squares of these residuals.

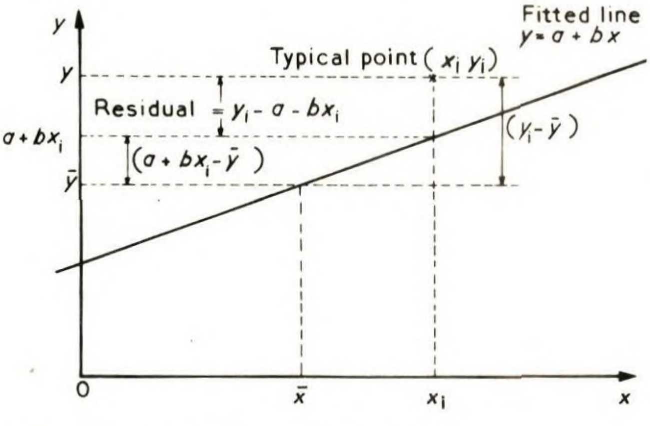


Figure 8. Illustration of simple linear regression

From figure 8 we see that the residual for the point (y_i, x_i) is given by $y_i - a - bx_i$. Remembering there are n data points, we are obtaining a and b to minimize:

$$\sum_{i=1}^{n} (y_i - a - bx_i)^2$$

By differentiating with respect to a and b, and equating the resultant expressions to O, we are left with two simultaneous equations involving a and b. Solving these equations we arrive at our solution:

$$b = \frac{\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} \qquad a = \frac{\Sigma y}{n} - \frac{b\Sigma x}{n}$$

Even at this stage we do not know the true values of the intercept and slope, but we have obtained the best estimates from the available data. If one more pair of points becomes available, then the best estimates of the intercept and slope are likely to be slightly different, and we therefore need some measure of the accuracy of our estimates. It can be shown that the 95 % confidence interval for the true slope is:

$$b \pm \frac{t_{n-1}s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

If this confidence interval excludes O then we are 95 % certain that the true slope is non-zero and therefore that x is definitely affecting y.

In the above confidence interval,

t_{n-2} is the 5 % point of the double-sided t-distribution on $(n - 2)$ degrees of freedom;

s is the residual standard deviation (i.e. the standard deviation of the residuals) and equals:

$$\sqrt{\frac{\sum_{i=1}^{n}(y_i - a - bx_i)^2}{n-2}}$$

A guide to the proportion of variation in y which can be attributed to its dependence on x is given by what is called the percentage fit of the equation.

The total variation in y is measured by $\sum_{i=1}^{n}(y_i - \bar{y})^2$.

The unexplained (residual) variation in y is measured by $\sum_{i=1}^{n}(y_i - a - bx_i)^2$ which can be shown to be equal to:

$$\Sigma y^2 - \frac{(\Sigma y)^2}{n} - b\left(\Sigma xy - \frac{\Sigma x \Sigma y}{n}\right)$$

which in turn can be shown equal to:

$$\Sigma(y - \bar{y})^2\,(1 - r^2)$$

and the percentage fit

$$= 100\left(1 - \frac{\sum_{i=1}^{n}(y_i - a - bx_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right) = 100\,r^2.$$

If all the points lie on the fitted line, percentage fit = 100, but in general most of the points will not be on the fitted line and the percentage fit will be less than 100.

Often we will be interested in the accuracy of the equation for predicting the value of y corresponding to a given value of x. For a given value of x, say X, we can say:

we are 95 % certain that the real or true value of y lies in the range

$$(a + bX) \pm t_{n-2}s\sqrt{\frac{1}{n} + \frac{(X - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

However, because the observed value of y for any experiment varies about its true value, the confidence interval for the observed value is wider:

we are 95 % certain that the observed value of y will lie in the range

$$(a + bX) \pm t_{n-2}s \sqrt{1 + \frac{1}{n} + \frac{(X - x)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

The example in **8.7** illustrates the technique of simple linear regression.

8.4 Multiple regression. Simple linear regression is of restricted use in the practical environment of industry, because in most situations there are more than two variables of interest. However, regression techniques can be powerful tools in the study of data when they are extended to a number of independent variables, or to higher degrees of equation (quadratic or possibly cubic). In such applications the use of a computer becomes virtually essential, but a number of quick and efficient computer programmes are already available for carrying out such regression analyses.

In 'multiple linear regression' we are considering an equation of the form:

$$y = a + b_1x_1 + b_2x_2$$

where, for example,

y is the tensile strength;

x_1 is the carbon black;

x_2 is the % natural rubber in the polymer.

This equation cannot be represented by a graph in two dimensions, but we can still measure its usefulness by evaluating the percentage fit (i.e. the percentage of variation in y accounted for by the equation).

In 'multiple quadratic regression' the equation of interest is of the form:

$$y = a + b_1x_1 + b_{11}x_1^2 + b_{12}x_1x_2 + b_2x_2 + b_{22}x_2^2$$

This form of equation allows for the fact that the relationship between the response and the independent variables may be curved rather than linear. The x_1x_2 term allows for the possibility of an interaction between variables x_1 and x_2, and would be very important if the effect of x_1 is dependent on the level of x_2.

8.5 Rejection of insignificant terms. If we fit the two equations

$$y = a + b_1x_1 + b_2x_2 \quad \text{and}$$

$$y = a + b_1x_1$$

to the same set of data, the former equation will aways give the higher percentage fit. In other words, including an extra variable will always increase the percentage fit. However it could be that the increase in percentage fit, due to including the second variable, is not statistically significant (i.e. possibly the extra variable is having no real effect and the increase in percentage fit is just due to chance).

In practice we fit an equation which includes all the variables that might be having an effect upon the response, and then we apply statistical tests to reject all those variables which are not making a significant contribution to the percentage fit. The theory of this statistical significance testing in multiple regression is more complex, and is not included here.

In some instances a variable may be retained in the final equation as having a statistically significant effect upon the response, and yet the coefficient of that variable might be so small as to make it of no practical importance. If the equation is being used to predict the response at specific values of the independent variables, such practically insignificant terms could be left out of the equation for ease of calculation purposes. On other occasions a variable might be rejected from the final equation when it is known from the technology of the situation that this variable really does affect the response. In some such instances it may be necessary to ensure that such a variable is retained in the final equation.

If, when we have carried out our regression analysis, the variation in the response unaccounted for by the equation is still comparatively large, then there are a number of possibilities:

(a) the wrong type of equation has been fitted;

(b) some important variable(s) has not been covered in the analysis;

or

(c) there is a large error in the determination of the response.

8.6 Summary. Regression analysis in all its forms is a most useful guide to control or optimization of some response (or even group of responses), but it is seldom that equations can be determined to be used for accurate predictions. Many words of caution have been written on the subject of regression analysis, because, being such a powerful technique, there is a danger of indiscriminate or irresponsible usage. It is, however, worthwhile to be reminded of some of the dangers.

(a) If an attempt is made to analyse inappropriate or unsuitable data nonsensical conclusions may well be reached.

(b) The problem should be formulated carefully, as it is not always clear whether certain variables should be classed as independent variables or responses.

(c) It can be very dangerous to use a regression equation to predict response values outside the experimental region already covered. Extrapolation can be very misleading.

(d) Using regression analysis is not a substitute for thinking about the problem. Knowledge of the situation being investigated is valuable, both before and after the regression analysis.

Table 4. Response values for combinations of the three independent variables

Response (duplicates)		Variable 1	Variable 2	Variable 3
12.0	13.9	1	20	0
1.4	0.9	1	20	3
2.6	1.3	1	20	7
3.4	2.4	1	20	14
35.6	34.2	1	70	0
8.0	7.8	1	70	3
4.3	2.7	1	70	7
9.9	8.1	1	70	14
18.5	19.3	3	20	0
0.9	1.4	3	20	3
2.7	6.9	3	20	7
10.8	9.7	3	20	14
42.1	44.7	3	70	0
16.7	19.6	3	70	3
11.5	11.0	3	70	7
12.6	13.5	3	70	14
25.3	24.7	7	20	0
10.2	9.4	7	20	3
16.3	16.9	7	20	7
22.1	23.2	7	20	14
58.8	55.7	7	70	0
19.2	17.9	7	70	3
22.1	24.4	7	70	7
12.1	12.1	7	70	14
29.4	33.6	14	20	0
16.1	17.4	14	20	3
14.8	15.0	14	20	7
11.2	11.4	14	20	14
67.5	67.7	14	70	0
32.8	32.9	14	70	3
17.6	16.0	14	70	7
14.7	16.5	14	70	14

Table 5. Correlation coefficient v numbers of observations

Number of pairs of observations	Value of correlation	Coefficient for significance	
	10 % level	5 %	1 %
3	0.9877	0.9969	0.9998
4	0.9000	0.9500	0.9900
5	0.805	0.878	0.9587
6	0.729	0.811	0.9172
7	0.669	0.754	0.875
8	0.621	0.707	0.834
9	0.582	0.666	0.798
10	0.549	0.632	0.765
11	0.521	0.602	0.735
12	0.497	0.576	0.708
13	0.476	0.553	0.684
14	0.457	0.532	0.661
15	0.441	0.514	0.641
16	0.426	0.497	0.623
17	0.412	0.482	0.606
18	0.400	0.468	0.590
19	0.389	0.456	0.575
20	0.378	0.444	0.561
21	0.369	0.433	0.549
22	0.360	0.423	0.537
27	0.323	0.381	0.487
32	0.296	0.349	0.449
37	0.275	0.325	0.418
42	0.257	0.304	0.393
47	0.243	0.288	0.372
52	0.231	0.273	0.354
62	0.211	0.250	0.325
72	0.195	0.232	0.302
82	0.162	0.217	0.283
92	0.173	0.205	0.267
102	0.164	0.195	0.254

8.7 Example on simple linear regression. The problem is to assess whether or not the 'age of material' is having a significant effect upon the 'duration of flexing life', and to determine the best linear relationship showing the dependence of 'duration of flexing life' on the 'age of material'.

Twelve pairs of observations are available, as shown (see also figure 7).

Duration of flexing life (y)	Age of material (x)
40.8	1
38.7	2
42.9	3
41.4	4
40.5	4.5
41.7	8
37.8	9
32.7	10
33.3	11
36.9	12
30.0	13.5
31.2	15

Calculations

$$\Sigma x = 93 \qquad \Sigma x^2 = 967.5$$

$$\Sigma y = 447.9 \qquad \Sigma y^2 = 16937.91$$

$$\Sigma xy = 3277.65 \qquad n = 12$$

$$\Sigma x^2 - \frac{(\Sigma x)^2}{n} = 246.75$$

$$\Sigma y^2 - \frac{(\Sigma y)^2}{n} = 220.0425$$

$$\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n} = -193.575$$

$$\text{Correlation coefficient } r = \frac{-193.575}{\sqrt{246.75 \times 220.0425}}$$

$$= -0.831$$

Comparing this value with the significance levels of the correlation coefficient in table 5, we see that, since our value is less than −0.708, we are 99 % certain that there is a relationship between these two variables.

For the 'best' straight line, we evaluate a and b, as follows:

$$b = \frac{-193.575}{246.75} = -0.7845$$

$$a = \frac{447.9}{12} - b \cdot \frac{93}{12} = 43.405$$

Hence the best linear relationship showing the dependence of y on x is

$$y = 43.405 - 0.7845x$$

Total sum of squares $= \Sigma(y - \bar{y})^2 = 220.04$.

Residual sum of squares $= \Sigma(y - a - bx)^2$

$= 68.18$.

Percentage fit $= 100 \left[1 - \frac{68.18}{220.04}\right] = 69.0$.

Residual standard deviation $= \sqrt{\frac{\text{residual sum of squares}}{n-2}}$

$= 2.61$.

95 % confidence interval for the true slope of the line is

$$b \pm \frac{t_{n-2}(5\ \%)s}{\sqrt{\Sigma(x - \bar{x})^2}}$$

which is $-0.785 \pm \frac{2.228 \times 2.609}{15.708}$

i.e. from − 1.155 to − 0.415.

9. Reporting of results

The methods of presentation of conclusions and results arising from the application of statistical techniques vary as much as the statistical techniques themselves, and even for an application of a specific technique the presentation of results would vary from one person to another. It is virtually impossible to eliminate this subjective element and consequently the following remarks are offered for guidance only.

Confining attention to the case where measurements have already been made on some particular variable, it is desired to present a concise summary of the total information available. One of the first steps, which is usually informative, if not essential, provided sufficient results are available, is to construct a histogram which condenses the data from a large number of individual results into a more meaningful graphical form. To construct a 'histogram' the results are combined into say approximately 10 distinct groups, with at the most say 5 % of the observations lying outside the interval covered by these groups. Usually the groups are of equal intervals on the measured variable and attention is here restricted to this case. The major step in constructing the histogram is to count the number of observations in each interval. Assuming that in this case all the results are recorded to one decimal place, a 'verbal' way of presenting the information is:

No. of observations	⩽ 10		is 0
No. of observations	in the range	10 to 14.9	is 10
No. of observations	in the range	15 to 19.9	is 10
No. of observations	in the range	20 to 24.9	is 20
No. of observations	in the range	25 to 29.9	is 35
No. of observations	in the range	30 to 34.9	is 50
No. of observations	in the range	35 to 39.9	is 70
No. of observations	in the range	40 to 44.9	is 40
No. of observations	in the range	45 to 49.9	is 30
No. of observations	in the range	50 to 54.9	is 15
No. of observations	in the range	55 to 59.9	is 5
No. of observations	⩾ 60		0

The graphical illustration of the information so accumulated is shown below in figure 9. The similarity between this and the frequency distribution discussed in clause **5** is obvious.

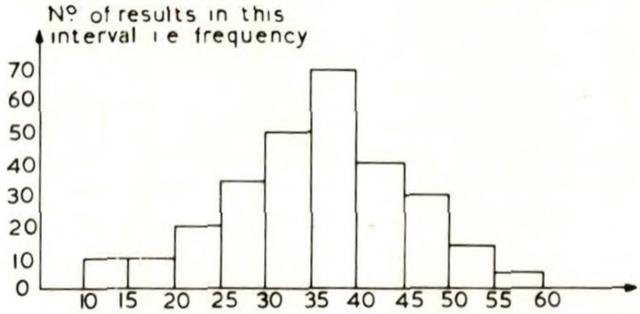


Figure 9. Histogram of results

Often, an additional section needs to be inserted at each end of the diagram; the height of the left-hand section would represent the number of observations < 10 and at the right-hand end the height would represent the number of observations ⩾ 60. In this graphical form we have not clearly defined our interval; obviously a result of 15 should only be included in one group, but which one? The way in which we define our group intervals exactly depends on the accuracy to which our results are measured; in the example if the results were always integers the first interval could be 10 to 14.9, whilst for results recorded to one decimal place we would have say 10 to 14.99.

This graphical presentation provides a good indication of how closely the individual points are clustered round a 'central' value, and whether or not the results tend to be occurring in higher frequencies at one end of the scale. In the example, one might suggest that about 37 seems to be a central value, and the graph is reasonably symmetrical about this axis.

The graphical presentation includes no statistical calculations, and is suitable when a large number of results are available. With fewer results, the number of groups would need to be decreased to make the histogram worthwhile. Other pictorial techniques are available in this situation and one of these is explained in **12.3**.

If numerical, rather than pictorial, consideration of results is required, and in any case with small numbers of observations, a measure of central tendency (e.g. the mean) and a measure of spread should be calculated as described previously. For a small number of results (say < 10) the median value should preferably be calculated and quoted as the measure of central tendency and the range method used for estimation of the standard deviation. For numbers of results less than five there is little point in calculating confidence limits, since these will usually be very wide due to the uncertainty associated with the estimation of the mean and standard deviation.

In practice, if more than five results are available attention is very often directed specifically towards the 'true' mean value of some process (and not the obtained mean value $\bar{x}$) and from $\bar{x}$ and s calculated as above we can assess how certain we are that this 'true' mean value will lie within a certain region. This statement might take the following form for the example considered earlier:

'We are 95 % certain that the 'true' mean value lies in the range 35 to 39'.

This confidence interval depends on the values of $\bar{x}$, s and n.

To some technologists a statement of the calculated values of $\bar{x}$ and s would be most meaningful, while to others the confidence interval is going to provide all the required information. Because of this individual preference it is essential that both items should be contained in the tabulated results, where appropriate.

For the example considered earlier, whether or not a histogram has been constructed, the following values calculated from the raw data should provide the necessary information:

number of observations	= 285;
arithmetic mean	= 37.0;
standard deviation	= 16.1.

We are 95 % certain that the true mean (of the process) lies in the range 35.1 to 38.9

It is possible to go further and construct a confidence interval for the true standard deviation of the process, but unless there is specific interest in the standard deviation rather than the mean, this information could tend to confuse rather than clarify the issue.

The presentation of results cannot really be discussed without some mention of the subject of 'rounding'. This arises in two distinct ways: the rounding of individual observations before carrying out any calculation and secondly the rounding of the results in their final presentation. Considering the latter first, a logical rather than theoretical rule is to quote the arithmetic mean, the standard deviation and the confidence interval each to one more significant figure than the individual results are observed. To round off to the same number of significant figures as the observations could be dangerous if any form of comparison of mean against a target value was to be made. The main purpose of rounding the observations before doing any calculations is to facilitate the arithmetic involved, and the need for this is not as strong as it was before calculating machines eased the task. More complex rules are available (see BS 2846 : 1957, page 40), but for most practical situations an effective guide would be to use the original data when this did not involve too much effort; the last significant figure could be rounded if this seems both reasonable and necessary.

10. Ranking methods

10.1 Ranking. Ranking methods are applied when the observations cannot be expressed as absolute values, but can be put in an order of increasing merit. In rubber testing they are particularly useful in judging exposed rubber samples by visual inspection with respect to, for instance, ozone cracks or discoloration.

10.2 Friedman's test. Each of m observers independently arranges n samples in order of increasing merit so that the jth observer assigns the rank R_{ij} to the ith sample, the best sample having a rank $R = n$. When ties are present they are given the relevant average rank.

As a measure for the differences between the samples the value

$$K = \sum_{i=1}^{n} (S_i - \bar{S})^2 \text{ is determined.}$$

The rank sum (S_i) for the ith sample is obtained by summing the ranks for this sample over the observers,

$$S_i = \sum_{j=1}^{m} R_{ij} \quad \text{and} \quad \bar{S} = \frac{\sum_{i=1}^{n} S_i}{n}$$

Whenever the value of K equals or exceeds a critical value K_{cr} (see table 6) it can be concluded that there are significant differences between the samples. Although this does not mean that significant differences exist within any pair of samples, it is generally meaningful to report the average ranks, $\bar{R}_i = S_i/m$.

Whether or not significance is obtained depends on the differences between the samples as well as on the degree of agreement among the observers. It may therefore be useful to report the 'degree of concordance' $C = \dfrac{12K}{m^2\,(n^3 - n)}$.

Its value may vary between 0 (no agreement) and +1 (complete agreement).

NOTE. In order to obtain a high degree of concordance it is necessary to describe the criterion of judgment clearly and to restrict the criterion to one aspect.

Table 6. Friedman's test critical values K_{cr} for $\alpha = 0.05$ level of significance

m^n	3	4	5	6	7	8	9	10	11	12	13	14	15
2	-	20	38	64	96	138	192	258	336	429	538	664	808
3	18	37	64	104	158	225	311	416	542	691	865	1063	1292
4	26	52	89	144	217	311	429	574	747	950	1189	1460	1770
5	32	65	113	183	277	396	547	731	950	1210	1512	1859	2254
6	42	76	137	222	336	482	664	887	1155	1469	1831	2253	2738
7	50	92	167	272	412	591	815	1086	1410	1791	2233	2740	3316
8	50	105	190	310	471	676	931	1241	1612	2047	2552	3131	3790
9	56	118	214	349	529	760	1047	1396	1813	2302	2871	3523	4264
10	62	131	238	388	588	845	1164	1551	2014	2558	3189	3914	4737
11	66	144	261	427	647	929	1280	1706	2216	2814	3508	4305	5211
12	72	157	285	465	706	1013	1396	1862	2417	3070	3827	4697	5685
13	78	170	309	504	764	1098	1512	2017	2618	3326	4146	5088	6159
14	84	183	333	543	823	1182	1629	2172	2820	3581	4465	5479	6632
15	90	196	356	582	882	1267	1745	2327	3021	3837	4784	5871	7106

10.3 Outside count test. This is a rough and ready method for the comparison of two specific samples out of n (e.g. one experimental and one reference). It is carried out as follows:

(a) count the number of values in the sample containing the highest value that are higher than the highest value in the other sample;

(b) count the number of values in the other sample that are lower than the lowest value of the first sample.

If the sum of the two counts totals seven or more, it may be concluded that the two samples are different at the 5 % level.

10.4 Example of the use of Friedman's test. Ten vulcanizates containing different antiozonants have been simultaneously exposed in an ozone cabinet. Each of five observers rank the 10 vulcanizates with respect to the degree of cracking, the criterion being crack length (see table 7).

$$K = \sum_{i=1}^{n} (S_i - \bar{S})^2 = (-10)^2 + (-17.5)^2 + \ldots$$

$$+ (9)^2 + (20.5)^2 = 1929$$

As K_{cr} for $n = 10$ and $m = 5$ is 731 it is concluded that there are significant differences between volcanizates.

The coefficient of concordance $C = \dfrac{12K}{m^2\,(n^3 - n)}$

$= \dfrac{12 \times 1929}{25(1000 - 10)} = 0.94$ so that the observers agree well in their judgment.

When applying the outside count test to vulcanizates 9 and 10 it is easily seen that the ranks do not overlap at all, so that the 'outside count' is 10 (twice the number of observers). As this value is larger than seven it is concluded that vulcanizates 9 and 10 differ significantly.

Section three. Statistical techniques applicable to specific tests

11. Introduction

In this section statistical techniques not necessarily described before that are applicable to specific rubber tests are discussed in detail. The absence of any particular test from this section should not, however, be taken to mean that no such specific techniques are available or desirable.

12. Tensile testing (BS 903 : Part A2)

12.1 Introduction. Tensile testing involves the measurement of three basic properties, stress at a given strain (modulus), tensile strength and elongation at break. With a small number of test pieces, e.g. three to five, the median should always be used. This clause indicates how to obtain more representative values for these parameters when a larger number of repeat test pieces are available.

12.2 Stress at given strain (modulus). Treatment of modulus results is usually quite straightforward since the distribution of results is in most cases Gaussian. For small numbers of tests the median should be used. Larger numbers of test results may be treated as in clause **5**.

12.3 Tensile strength. The distribution of tensile strength is not in general Gaussian but is markedly skew. This distribution frequently follows the so-called double exponential distribution (see figure 10). In the analysis of such a distribution the most important measure of which central tendency is the mode, it is often desirable to obtain an estimate from a set of results. It is clearly not usually practical to measure so many test pieces that a full distribution curve can be drawn. It is therefore necessary to consider methods whereby the mode can be estimated from relatively few test results.

Table 7. Ranking of 10 rubber vulcanizates by five observers

Observer	Vulcanizates									
	1	2	3	4	5	6	7	8	9	10
A	4	1	5½	5½	2	3	8	10	7	9
B	3½	2	5	6	1	3½	7	9	8	10
C	3	2	4	6	1	5	7½	9	7½	10
D	3	3	6	5	1	3	9	8	7	10
E	4	2	4	6	1	4	10	8	7	9
Sum S_i	17½	10	24½	28½	6	18½	41½	44	36½	48
Mean sum $\bar{S}$	27½	27½	27½	27½	27½	27½	27½	27½	27½	27½
$S_i - \bar{S}$	−10	−17½	− 3	+ 1	−21	− 9	+14	+16½	+ 9	+20½
Mean rank	3.5	2	4.9	5.7	1.2	3.7	8.3	8.8	7.3	9.6

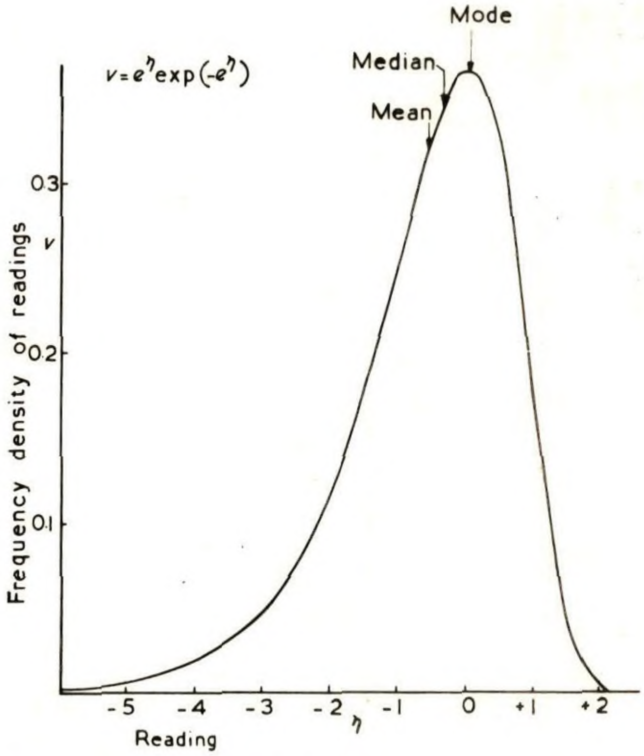


Figure 10. Double exponential distribution

12.3.1 *Graphical method.* If the results are arranged in numerically descending order and plotted on specially designed probability paper it is possible to determine whether a number of observations have been drawn from a population with a double exponential distribution and, if so, its mode and standard deviation. This technique is best illustrated by an example. The data below shows the results of tensile measurements on three sets of 12 dumb-bells cut from single sheets of vulcanized rubber. The results have been ranked in descending order and are not in the order of testing.

Tensile strength (MPa)

A	B	C
28.4	26.7	19.7
27.9	26.2	19.6
27.4	26.1	19.2
27.1	26.1	19.0
26.8	25.9	18.7
26.5	25.8	18.4
26.3	25.8	18.1
26.2	25.8	17.3
26.0	25.7	16.4
25.9	25.6	15.6
24.6	25.1	15.1
24.1	25.0	13.5

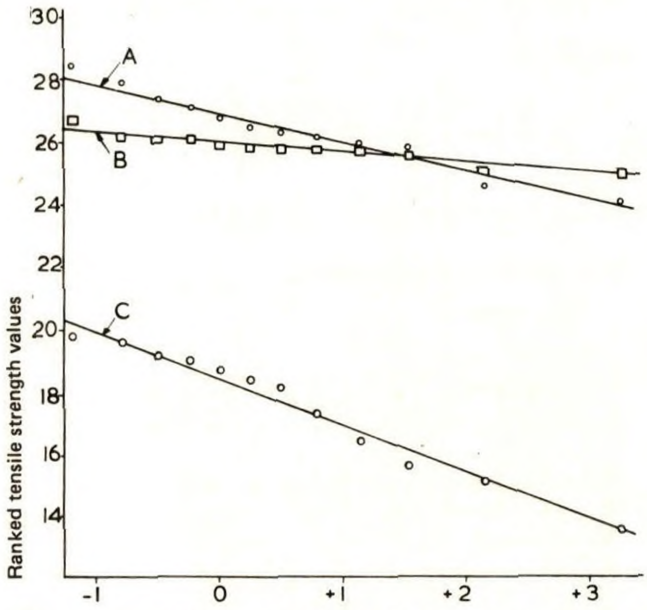
These results are plotted (see figure 11) using abscissae (in this case for 12 test results) obtained from table 8. After drawing the best straight line through these points the value of the tensile strength corresponding to plotting point zero and the slope of the line are read off. The former gives the mode value and the latter the standard deviation. If one or two points do not lie on the straight line these may be considered atypical and ignored. In cases where a reasonably representative straight line cannot be drawn through the points the data cannot be assumed to be drawn from the double exponential distribution and this technique should not be used.

Reading from figure 11 the following results are obtained:

A	mode 27.0	standard deviation	0.9
B	mode 26.0	standard deviation	0.33
C	mode 18.4	standard deviation	1.5

Vulcanizates A and B are essentially similar except for differences in the mixing giving differing dispersions. In practice it has been found that the value of the standard deviation is often a useful indicator of the uniformity of dispersion, poor dispersion giving rise to high values of standard deviation.

This graphical technique, whilst having the advantage of enabling atypical results to be disregarded, is time consuming and it is often preferable to use a purely numerical method.



Standardized deviates (plot positions taken from data in clause **3**)

Figure 11. Analysis of tensile strength results

12.3.2 *Numerical method.* In this case the test results are again ranked in decreasing order and each value multiplied by a weighting factor obtained from table 9 or 10. The sum of these terms gives in effect the mode and standard deviation that would have been obtained by drawing a least squares line through the data plotted as shown in **12.3.1**. Thus:

$$\text{Mode} = S_1w_1 + S_2w_2 + S_3w_3 + ...$$

$$\text{Standard deviation} = S_1d_1 + S_2d_2 + S_3d_3 + ...$$

When the process is carried out on the data in **12.3.1** the following results are obtained:

A	mode 26.96	standard deviation	0.94
B	mode 25.99	standard deviation	0.33
C	mode 18.41	standard deviation	1.52

Where more than 12 data are used it is normally better to use the graphical method, but where the number of results lies between 5 and 12 the numerical method is very convenient, especially if a simple desk calculator is available.

12.4 Elongation at break. This follows the same type of distribution as tensile strength and may be treated similarly.

Table 8. Plot positions (double exponential frequency distribution function)

n \ 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	+0.19	+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

N: total number of test results;
n: serial number of test result, increasing with decreasing magnitude.

Table 9. Weighting factors *w* for calculating mode

n \ *N*	5	6	7	8	9	10	11	12
1	0.327	0.274	0.236	0.207	0.185	0.167	0.152	0.139
2	0.269	0.233	0.205	0.182	0.164	0.150	0.137	0.126
3	0.217	0.197	0.179	0.163	0.149	0.137	0.126	0.117
4	0.153	0.159	0.159	0.144	0.134	0.125	0.117	0.109
5	0.034	0.111	0.125	0.124	0.120	0.114	0.107	0.101
6		0.025	0.086	0.100	0.103	0.101	0.098	0.094
7			0.015	0.069	0.084	0.087	0.088	0.086
8				0.010	0.056	0.071	0.075	0.076
9					0.005	0.047	0.061	0.066
10						0.002	0.039	0.053
11							0.000	0.034
12								−0.001

Table 10. Weighting factors *d* for calculating standard deviation

n \ *N*	5	6	7	8	9	10	11	12
1	0.213	0.187	0.162	0.142	0.128	0.116	0.106	0.097
2	0.116	0.115	0.107	0.100	0.092	0.086	0.080	0.075
3	0.028	0.053	0.063	0.066	0.065	0.063	0.062	0.059
4	−0.079	−0.013	0.019	0.034	0.040	0.044	0.045	0.045
5	−0.278	−0.096	−0.031	−0.001	0.015	0.024	0.028	0.032
6		−0.246	−0.098	−0.042	−0.013	0.003	0.019	0.018
7			−0.220	−0.098	−0.047	−0.022	−0.006	0.004
8				−0.200	−0.095	−0.051	−0.027	−0.012
9					−0.184	−0.092	−0.052	−0.031
10						−0.170	−0.089	−0.053
11							−0.158	−0.086
12								−0.148

N is the total number of test results;
n is the serial number of test result, increasing with decreasing magnitude.

13. Tear strength (crescent) (see BS 903 : Part A3)

13.1 Use of linear regression to give better discrimination of the test. Tear strength, in this test, is a measure of the force required to break a nicked test piece. Obviously the larger the nick the lower the force required to break the test piece. Within a limited range of nick depths, their relation to tear strength can be regarded as linear.

The method permits the nick depth to be in the range 0.42 mm to 0.58 mm. This is because of obvious practical difficulties in controlling the depth of cut made.

Depending on the formulation of the rubber this range of nick depths can give tear strengths differing by as much as 40 %, or as little as 5 %.

In situations requiring maximum discrimination between tear strength values the amount of variation in the test results can be reduced by removing the variation attributable to nick depth differences. This can be done in the following manner.

Prepare and nick about 12 test pieces. Vary the depth of nick from test piece to test piece so that the range 0.3 mm to 0.6 mm is fairly evenly covered. Measure nick depth on both sides of the test piece. Use the average value in subsequent calculations. Measure the tear strength of each test piece in random order by the recommended method.

Calculate the linear regression equation and the correlation coefficient.

Let $x_1 x_2$ mm ... x_n mm be the nick depths.

Let $y_1 y_2$ N ... y_n N be the sample tear strengths.

Determine constants a and b in the regression line:

$$y = a + bx.$$

(a) Obtain the totals

$x_1 + x_2 + ... x_n$ i.e. Σx

$x_1^2 = x_2^2 + ... x_n^2$ i.e. Σx^2

Similarly calculate Σy and Σy^2

$x_1 y_1 + x_2 y_2 + ... x_n y_n$ i.e. Σxy

(b) Calculate the sums of squares

$$\Sigma x^2 - \frac{(\Sigma x)^2}{n} = C_{11}$$

$$\Sigma y^2 - \frac{(\Sigma y)^2}{n} = C_{yy}$$

$$\Sigma xy - \frac{\Sigma x \cdot \Sigma y}{n} = C_{y1}$$

(c) Calculate the following.

(1) Correlation coefficient, $r = \dfrac{C_{y1}}{\sqrt{C_{11} \cdot C_{yy}}}$

(2) Slope of regression line, b. $b = \dfrac{C_{y1}}{C_{11}}$

(3) y – intercept, a $a = \dfrac{\Sigma y - b\Sigma x}{n}$

(4) [Residual error]^2, $(S_r)^2$

$$= \frac{(1 - r^2) \cdot C_{yy}}{n - 2}$$

(d) Calculate the value of y from the equation

$y = a + bx$

using $x = 0.50$ mm.

This value of y is the best estimate of tear strength at a nick depth of 0.50 mm which can be used for further calculations and comparisons.

(e) Calculate

$$(a + 0.5b) \pm t_{n-2} \cdot S_r \sqrt{\frac{1}{n} + \frac{(0.5 - \bar{x})^2}{X_1^2}}$$

where t is the Student's t.

This will give the 95 % confidence limits on the true y value corresponding to a nick depth of 0.5 mm.

14. Abrasion resistance (see BS 903 : Part A9)

14.1 Use of analysis of variance. This test measures the volume loss in 1000 revolutions of the abrasive. These notes illustrate the procedure for four vulcanizates A_1, A_2, A_3, A_4, tested in triplicate.

The procedure is suitable for a hierarchical analysis of variance to give guidance on judging the differences between the four vulcanizates.

The following calculation procedure illustrates how this would be done.

NOTE. Four vulcanizates are considered but the method need not be restricted to four.

(a) Form a table of the volume loss results for each measurement:

Test	1	2	3	Row total
Vulcanizate				
A_1	x_1	x_2	x_3	R_1
A_2	x_4	x_5	x_6	R_2
A_3	x_7	x_8	x_9	R_3
A_4	x_{10}	x_{11}	x_{12}	R_4
Column total	C_1	C_2	C_3	Grand total (GT)

(b) Obtain the row totals, $(R_1 = x_1 + x_2 + x_3)$

Obtain the column totals $(C_1 = x_1 + x_4 + x_7 + x_{10})$

Obtain the grand total $(GT = C_1 + C_2 + C_3)$, check $= R_1 + R_2 + R_3 + R_4$

(c) Obtain from this table the following totals:

Σx, Σx^2, ΣR, ΣR^2, ΣC,

(note $\Sigma x = \Sigma R = \Sigma C = GT$)

(d) Calculate the 'correction factor', $CF = \dfrac{(GT)^2}{12}$ (for 12 observations).

(e) Calculate the total sum of squares, T

$T = \Sigma x^2 - CF$

(f) Calculate the vulcanizates sum of squares, B

$B = \dfrac{\Sigma R^2}{3} - CF$ (the 3 is for the number of tests per vulcanizate).

(g) Form and complete the analysis of variance table.

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio
Between vulcanizates	3	B	$B/3$	$\frac{B}{3} \times \frac{8}{T-B}$
Between repeats within vulcanizates	12–1–3 = 8	$T - B$	$\frac{T-B}{8}$	
Total	12–1 = 11	T		

(h) Check in variance ratio tables (see appendix D, table 13) under 3, 8 degrees of freedom for the significance of the $\dfrac{B}{3} \times \dfrac{8}{T-B}$ value.

(i) If it is significant this indicates there is a significant variation in the mean values of the vulcanizates.

15. Crack growth and fatigue testing

15.1 Introduction. Failure in vulcanized rubber undergoing repeated deformations can occur as a result of crack growth. Cracks develop from imperfections in the material and can ultimately cause complete severance of an article. Flex cracking resistance can be measured by the De Mattia apparatus as described in BS 903 : Parts A10 and A11. Alternatively the fatigue life, or number of cycles to failure, can be measured using suitable machines.

15.2 Methods for analysis of results

15.2.1 *Use of linear regression (see BS 903 : Part A10).* BS 903 : Part A10 recommends that results be quoted as the number of kilocycles flexing to reach each successive stage or grade of cracking. The grades defined are in order of increasing severity of cracking and are given the letters A to F.

For certain vulcanizates, especially those based on natural rubber, the rate of development of cracks through the grades A to F can be adequately represented by a straight line which relates the number of kilocycles flexing received by the test piece to the grade of cracking produced. If the grades A to F are represented by the numbers 1 to 6 then this relation has the form

$y = a_0 + a_1 x$

where

y is the number of kilocycles to reach a given grade;

x is the numerical grade of crack pattern;

a_1 is the slope of the relationship;

a_0 is the y-intercept.

The slope, a_1, will represent a 'flexing life per grade' and can be used as a measure of the flexcracking resistance of the vulcanizate. Similarly, in principle, the intercept a_0 would represent an 'initiation' period, which implies that a certain amount of flexing is required to produce the first visible signs of damage whereupon the growth of this damage proceeds at a different rate. It is not necessarily true true that a_0 represent an initiation period in the usually accepted meaning of the phrase. It could be that the numbering of the grades A to F should be say from 3 to 8 and that a straight line through the origin is representative. However, there is evidence from work with different polymers of a long initiation type interval followed by a rapid degradation through the cracking stages.

BS 903 : Part A10 recommends that at least three test pieces be used to characterize a vulcanizate and that mean values for each grade of cracking be quoted. When one has to compare such means for two different vulcanizates some account has to be taken of the variability of the results from which each mean value was obtained. As there are twelve such mean values the process of comparison is not obvious.

Providing that a straight line adequately relates the flexing kilocycles to the flex grade then the slope of this line will characterize the test piece while the spread of results about this line will give a measure of the variability with which to judge the difference between two such slopes.

The mathematical procedure to do this is given next.

Calculation of various statistical quantities

Let the observations be:

No. kilocycles flexing	Grade reached
y_1	$A \dots x_1$
y_2	$B \dots x_2$
y_3	$C \dots x_3$
y_4	$D \dots x_4$
y_5	$E \dots x_5$
y_6	$F \dots x_6$

Calculate Σy; Σx; Σy^2; Σx^2; Σxy.

Calculate $$\Sigma x^2 - \frac{(\Sigma x)^2}{n} = C_{11}$$

$$\Sigma x^2 - \frac{(\Sigma y)^2}{n} = C_{yy}$$

$$\Sigma xy - \frac{\Sigma x \Sigma y}{n} = C_{yi}$$

where n is the number of grades considered.

Calculate

(a) slope $a_1 = \frac{C_{y1}}{C_{11}}$;

(b) intercept $a_0 = \bar{y} - a_1\bar{x}$ or $\frac{\Sigma y - a_1 \Sigma x}{n}$;

(c) correlation coefficient $r = \frac{C_{y1}}{\sqrt{C_{11} \cdot C_{yy}}}$ and percentage fit $= 100r^2$;

(d) residual variance of regression line

$$s_r^2 = \frac{C_{yy} - \frac{(C_{y1})^2}{C_{11}}}{n-2};$$

(e) variance of slope $s_s^2 = \frac{s_r^{222}}{C_{11}}$

Use of the calculated values.

(1) The flex life of the test piece is estimated by a_1 and is quoted in kilocycles per grade. Its confidence limits (95%) are $a_1 \pm t\frac{s_r}{C_{11}}$.

(2) A guide to the adequacy of the calculated slope to represent the data can be gained from the correlation coefficient r, or better, the percentage fit. If this is less than about 80 %, then it is recommended that the calculated 'life' or slope be used with caution.

(3) For replicate test pieces giving values of

slope $= a_1, b_1, c_1$ etc.

and variance of slope $= s_{sa}^2, s_{sb}^2$, etc.

in order to judge whether these test pieces give identical slopes it is essential that the biggest difference between two slopes be smaller than

$t\sqrt{s_{sa}^2 + s_{sb}^2}$ where $t = 2.31$ for $2(n-2)$ $= 8$ degrees of freedom.

(4) A representative slope for a given formulation can be found by taking the average slope of all test pieces from that vulcanizate.

i.e. $$\frac{a_1 + b_1 \dots + n_1}{\text{number of test pieces}} = \text{average life, } L$$

(although this is not the most accurate statistical estimate)

(5) If L_1 and L_2 represent average lives for two different vulcanizates, the significance of the difference between them can be judged by performing a comparison of means calculation.

Results to be quoted. It is recommended that the following values are given in reporting results for each test piece:

slope a_1

intercept a_0

percentage fit

residual variance about the regression line, s_r^2 } or quote s_r and s_s residual standard deviation, and

variance of the slope, s_s^2. } standard error of slope.

15.2.2 *Use of quadratic regression (see BS 903 : Part A11).* The test requires that a standard cut is made in the test piece which is then flexed repeatedly and the length of the growing crack is measured at known intervals of time. The initial cut measures about 2 mm and the test is complete when its length is about 12 mm.

It is required to know how many flexing cycles are needed to cause the initial cut to grow by 2 mm; a further 4 mm; a further 4 mm. From a smooth curve of the plot of length of crack against number of flexing cycles these values can be read.

It can be advantageous to interpolate the required values from a derived quadratic regression equation, especially if adequate calculating equipment is available. To do this, the coefficients of the following equations are calculated:

$$y = A_0 + A_1 x + A_2 x^2$$

where

x is the crack length, mm;

y is the number of kilocycles flexing;

L is the initial cut length, mm.

It is required to find the values of y when $x = L$ and $L + 2$ (or $L + 2$ and $L + 6$ or $L + 6$ and $L + 10$).

To avoid having to extrapolate select those observations which just exceed the range of x under consideration, e.g. L to $L + 2$.

Let these be $x_1, y_1; x_2, y_2; .... x_n, y_n$.

(a) Calculate

$$x_1 + x_2 + ... x_n = \Sigma x$$

$$x_1^2 + x_2^2 + ... x_n^2 = \Sigma x^2$$

Similarly Σx^3, Σx^4, Σy, Σy^2, Σxy, $\Sigma x^2 y$

Then calculate the sums of squares.

(b) $$\Sigma x^2 - \frac{(\Sigma x)^2}{n} = C_{11}$$

$$\Sigma x^4 - \frac{(\Sigma x^2)^2}{n} = C_{22}$$

$$\Sigma x^3 - \frac{\Sigma x \Sigma x^2}{n} = C_{12}$$

$$\Sigma y^2 - \frac{(\Sigma y)^2}{n} = C_{yy}$$

$$\Sigma xy - \frac{\Sigma x \cdot \Sigma y}{n} = C_{y1}$$

$$\Sigma x^2 y - \frac{\Sigma x^2 \cdot \Sigma y}{n} = C_{y2}$$

(c) Use these sums of squares in the following equations and solve for A_1 and A_2.

$$C_{11} \cdot A_1 + C_{12} \cdot A_2 = C_{y1}$$

$$C_{12} \cdot A_1 + C_{22} \cdot A_2 = C_{y2}$$

(d) Find A_0 from

$$nA_0 = \Sigma y - A_1 \Sigma x - A_2 \Sigma x^2$$

(e) Calculate $D = A_1 \cdot C_{y1} + A_2 \cdot C_{y2}$

and $T = C_{yy}$

(f) Complete the analysis of variance table:

Source of variance	Degrees of freedom	Sums of squares	Mean square	Variance ratio
Due to regression	2	D	$D/2$	$\frac{D}{2} \cdot \frac{n-3}{T-D}$
Residual	$n-3$	$T-D$	$(T-D)/(n-3)$	

(g) Check the significance of the regression by comparing the value obtained for the variance ratio with the appropriate value in the statistical tables of variance ratio (see appendix D, table 13).

(h) Use the derived response equation to calculate:

(1) y for $x = L$ (or $L + 2$ or $L + 6$);

(2) y for $x = L + 2$ (or $L + 6$ or $L + 10$);

(3) the difference between these y values.

This will represent the number of flexing cycles to cause the cut to grow from L to $L + 2$ (or $L + 2$ to $L + 6$ or $L + 6$ to $L + 10$).

15.2.3 *Fatigue testing.* Fatigue testing involves measurement of the number of cycles to failure (the fatigue life) of test pieces subjected to repeated deformations. A British Standard for tension fatigue is in preparation and this will give general guidance on the presentation and interpretation of results. A brief description of statistical aspects of fatigue behaviour is given below together with a simple method for the assessment of results.

The inherent variability in fatigue life is very much greater than in other strength properties, such as tensile strength; this reflects the greater sensitivity of fatigue life to factors which influence failure, such as flaw size. The extent of the variation depends on vulcanizate compositions, particularly the type of rubber used; for example, the overall variation for vulcanizates of natural rubber (NR) or isoprene rubber (IR) is typically two-fold or less, whereas for styrene-butadiene rubber (SBR) or butadiene rubber (BR) it can be an order of magnitude or more. The 'nature of the distribution' is also influenced by vulcanizate details and no single distribution is applicable to all rubbers. Thus for NR or IR vulcanizates, the distribution of fatigue lives often approximates to a Normal (Gaussian) distribution; for SBR, on the other hand, the distribution of fatigue lives tends to be markedly skew, but that of their logarithms may be essentially Normal. In view of these differences in behaviour and the complexities they present, particularly in relation to the treatment of blends of different rubbers (which are very widely used in practice), it is likely that a generally applicable method of analysis, which may be along the simple lines given below, will be recommended in the proposed standard for tension fatigue.

For each set of tests, the following should be reported:

(a) number of test pieces used (minimum 6);

(b) individual fatigue lives in ascending order of magnitude;

(c) median fatigue life;

(d) measure of dispersion.

It is important that some measure of dispersion is quoted. Apart from standard techniques described earlier in this guide, use of the ratio of highest to lowest life has provided a simple measure that has been found useful in the particular area of fatigue testing. In principle this ratio involves some disadvantages, but for the numbers of test pieces that are normally involved it has been found to correlate closely with the coefficient of variation and is much easier to handle. Due to the complexity of fatigue testing and the differences in the behaviour of the various rubbers, care should be taken in applying statistical tests described earlier in this guide. Appropriate statistical tests are available (e.g. distribution free tests) but are beyond the scope of this guide.

Low results should be disregarded if (and only *if*) there is positive non-statistical evidence that they are unrepresentative (e.g. the presence of an abnormally large flaw in the fracture surface which is clearly attributable to a fault in test piece preparation).

For NR and IR, six test pieces should give a representative measure of the median but for SBR and rubbers that behave similarly 12 test pieces are likely to be required, particularly when only one strain cycle is used. Apart from the general reasons given earlier in this guide, the median provides a more satisfactory measure of central tendency than the arithmetic mean for rubbers such as SBR the (fatigue) lives of which follow a skew distribution. Attention is drawn to the lowest fatigue life since this is often primarily of concern from a service viewpoint.

16. Resistance to low temperatures (see BS 903 : Part A13)

16.1 Use of cubic regression to give better estimates of regidity modulus. The test determines the apparent rigidity modulus of rubbers at different temperatures. Practice usually covers the temperature range from − 75 °C to + 15 °C. Most technologically useful materials change from the 'frozen' solid state to the 'rubbery' state somewhere in this region. Their modulus decreases from about 10 000 Pa to about 1 Pa. The shape of the modulus versus temperature curve is sigmoid.

Often, the fitting of a cubic regression equation to the results gives a better estimate of the temperatures for given moduli. The calculations are as follows.

It is required to determine the values of the constants in the following equation and to check their usefulness or significance:

$$y = A_0 + A_1x + A_2x^2 + A_3x^3$$

where

x is the apparent rigidity modulus;

y is the temperature.

Data points available are x_1y_1; x_2y_2; .... x_ny_n

(a) Obtain the totals

$$\Sigma x, \Sigma x^2, \Sigma x^3, \Sigma x^4, \Sigma x^5, \Sigma x^6, \Sigma y, \Sigma y^2, \Sigma xy, \Sigma x^2y, \Sigma x^3y$$

where

$$\Sigma x = x_1 + x_2 + ... x_n;$$

$$\Sigma x^2y = x_1^2y_1 + x_2^2y_2 + ... x_n^2y_n \text{ etc.}$$

(b) Calculate the sums of squares

$$C_{11} = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

$$C_{22} = \Sigma x^4 - \frac{(\Sigma x^2)^2}{n}$$

$$C_{33} = \Sigma x^6 - \frac{(\Sigma x^3)^2}{n}$$

$$C_{12} = \Sigma x^4 - \frac{\Sigma x \Sigma x^2}{n}$$

$$C_{13} = \Sigma x^4 - \frac{\Sigma x \Sigma x^3}{n}$$

$$C_{23} = \Sigma x^5 - \frac{\Sigma x^2 \Sigma x^3}{n}$$

$$C_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$

$$C_{y1} = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$$

$$C_{y2} = \Sigma x^2y - \frac{\Sigma x^2 \Sigma y}{n}$$

$$C_{y3} = \Sigma x^3y - \frac{\Sigma x^3 \Sigma y}{n}$$

(c) Use these sums of squares in the following equations, and solve for A_1, A_2 and A_3.

$$C_{11}\ A_1 + C_{12}\ A_2 + C_{13}\ A_3 = C_{y1}$$

$$C_{12}\ A_1 + C_{22}A_2 + C_{23}\ A_3 = C_{y2}$$

$$C_{13}\ A_1 + C_{23}\ A_2 + C_{33}\ A_3 = C_{y3}$$

(d) Find A_0 from

$$nA_0 = \Sigma y - A_1\Sigma x - A_2\Sigma x^2 - A_3\Sigma x^3$$

(e) Calculate $D = A_1\ C_{y1} + A_2\ C_{y2} + A_3\ C_{y3}$

and $T = C_{yy}$

(f) Complete the analysis of variance table:

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio
Due to regression	3	D	$D/3$	$\frac{D}{3} \cdot \frac{n-4}{T-D}$
Residual	$n-4$	$T-D$	$(T-D)/(n-4)$	
Total	$n-1$	T		

(g) Check the significance of the regression by comparing the value obtained for the variance ratio with the appropriate value in the statistical tables of variance ratio.

(h) Use the derived response equation to calculate the (predicted) temperature at any required modulus.

17. Ozone resistance (see BS 903 : Part A23)

Results from this test are sometimes qualitative and sometimes quantitative. Qualitative results can be treated as indicated in clause **10**. Quantitative results can be treated by appropriate statistical techniques, e.g. regression or Student's *t* test.

Appendix A

Formulae for easy reference

Mean $\bar{x} = \Sigma x/n$

Standard deviation

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

Coefficient of variation

$$v = \frac{s}{|\bar{x}|} \times 100\ \% \text{ or } \frac{\sigma}{|\bar{x}|} \times 100\ \%$$

Where the standard deviation is accurately known:

Confidence limits for mean

$$\bar{x} \pm u\sigma/\sqrt{n}$$

Least significant differences between two means

$$u\sigma\sqrt{(1/n_1 + 1/n_2)}$$

Where the standard deviation is estimated from the sample(s):

Confidence limits for mean

$$\bar{x} \pm ts/\sqrt{n} \text{ (where } v = n-1)$$

Pooled estimate of standard deviation

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

$$= \sqrt{\frac{\Sigma x_1^2 - (\Sigma x_1)^2/n_1 + \Sigma x_2^2 - (\Sigma x_2)^2/n_2}{n_1+n_2-2}}$$

Least significant difference between two means

$$ts\sqrt{(1/n_1 + 1/n_2)} \text{ (where } v = n_1+n_2-2)$$

Appendix B

References (useful books and tables)

B.1 Introductory works

1. BROOKES, B.C. and DICK, W.F.L. Introduction to statistical methods. Heinemann.
2. CROW, E.L. and others. Statistics manual. Dover.
3. DAVID, F.N. Statistical primer. Griffen.
4. DAVIES,* O.L. Design and analysis of industrial experiments. Oliver and Boyd.
5. DAVIES, O.L. *and* GOLDSMITH, P.L. Statistical methods in research and production. 4th edition. Oliver and Boyd.
6. DRAPER and SMITH. Applied regression analysis. Wiley.
7. HAYSLETT, H.T. Statistics made simple. W.H. Allen.
8. KENDALL and BUCKLAND. Dictionary of statistical terms. Oliver and Boyd.
9. LANGLEY, R. Practical statistics. Pan Books.
10. MORONEY, M.J. Facts from figures. Penguin Books.
11. TIPPETT, L.H.C. Technological applications to statistical methods. Williams and Norgate. Chapters 8 and 9.
12. BS 600, see inside back cover.
13. BS 2846, see inside back cover.
14. BS 6000, see inside back cover.

B.2 Tables

15. Barlow's tables. Edited by L.J. Comrie.
16. FISHER, R.A. and YATES, F. Statistical tables. Oliver and Boyd.
17. LINDLEY, D.V. and MILLER, J.C.P. Cambridge elementary statistical tables. Cambridge University Press.
18. PEARSON, E.S. and HARTLEY, H.O. Biometrika tables for statisticians. Cambridge University Press.

Appendix C

Glossary of statistical terms and symbols

Term	Definition
population (or universe)	In rubber testing, the aggregate of all test readings on a single property which could be made on the rubber under investigation.
sample	The test readings actually made, usually selected to constitute a random selection from the population.
sample size, n	The number of test readings in the sample.
variate, x	An individual test reading, subject to variation.
Σ	The sum of ...
arithmetic mean (mean), $\bar{x}$	The sum of the readings, divided by their number.
mean deviation	The arithmetic mean of the deviations of individual readings from the mean, neglecting their sign.
standardized normal deviation, u	A variate, Normally distributed about a mean at zero and with unit standard deviation. (For tables of the Normal (Gaussian) distribution, see textbooks and appendix D.)
Student's *'t'*, t	The ratio of a sample mean, Normally distributed about a mean at zero, to the estimate of its standard deviation. (For tables of the distribution see textbooks and appendix D.)
degrees of freedom, γ (sometimes referred to as ϕ)	The number of independent differences between the readings available for an estimate of standard deviation. For example, in estimating the standard deviation from a sample of n readings, there are only $(n-1)$ independent values of $(x-\bar{x})$, the last value being determined by the requirement that $\Sigma(x-\bar{x})=0$, i.e. one degree of freedom has been 'lost' in the calculation of the mean $\bar{x}$.
precision	The closeness of an estimate, usually expressed as 95 % confidence limits.
confidence limits (for a mean)	The range of values about the sample mean within which the true mean can, with a certain degree of confidence, be stated to lie.
least significant difference (between two sample means)	A value which the difference between two sample means must exceed in order to establish, with a certain small risk of error, that the true means in fact differ.
level of significance	The probability of error associated with significance tests. (See least significant difference.)
factor	One of the variables in a planned experiment whose effect it is desired to estimate, and which can be controlled at a specified value.
interaction	When the effect of one variable is different at different levels of a second variable, then there is said to be an interaction between the two variables.

Term	Definition
factorial experimental design	A plan of experiments in which an experiment is done at each of the possible combinations of levels of the factors.
effect	When a factor is only considered at two levels, the effect of the factor is: average result at the higher level of the factor, minus result at the lower level of the factor.
correlation coefficient, r =	$\frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2 \cdot \Sigma(y-\bar{y})^2}}$ where $-1 \leqslant r \leqslant +1$ A measure of the extent to which the two variables in question vary together.
percentage fit	A measure of the proportion of variation in the response which is accounted for by variation in the other variables. Equal to $100\,r^2$.
residual	When the best relationship has been fitted, the difference between the observed value of the response and the predicted value of the response corresponding to the same values of the independent variables, is called the residual.
residual standard deviation	This is simply the standard deviations of the residuals, and is a measure of the scatter of the points about the fitted equation.
variance of population	The arithmetic mean of the squares of the deviations of the population measurements from the arithmetic mean of the population.
variance of sample	The sum of the squares of the deviations of the sample observations from the arithmetic mean of the sample, divided by $n-1$.
standard deviation of population, σ	The square root of the variance of the population.
standard deviation of sample, s	The square root of the variance of the sample.
response	Any property which we are measuring and attempting in some way to optimize, and which is thought to be dependent on some other variables is termed a response (e.g. tensile strength, hardness).
independent variable	A variable which is thought to be causing variation in a response variable, or which can be used to account for variation in a response variable.
standard error	A measure of the accuracy of the estimate of a parameter.

Appendix D

Statistical reference tables

Table 11. The Normal (Gaussian) distribution function

Single sided				Double sided	
U	F(U)	U	F(U)	U	G(U)
−2.6	0.005	0.1	0.540	0.0	0.000
−2.5	0.006	0.2	0.579	0.1	0.080
−2.4	0.008	0.3	0.618	0.2	0.158
−2.3	0.011	0.4	0.655	0.3	0.236
−2.2	0.014	0.5	0.692	0.4	0.310
−2.1	0.018	0.6	0.726	0.5	0.384
−2.0	0.023	0.7	0.758	0.6	0.452
−1.9	0.029	0.8	0.788	0.7	0.516
−1.8	0.036	0.9	0.816	0.8	0.576
−1.7	0.045	1.0	0.841	0.9	0.632
−1.6	0.055	1.1	0.864	1.0	0.682
−1.5	0.067	1.2	0.885	1.1	0.728
−1.4	0.081	1.3	0.903	1.2	0.770
−1.3	0.097	1.4	0.919	1.3	0.806
−1.2	0.115	1.5	0.933	1.4	0.838
−1.1	0.136	1.6	0.945	1.5	0.866
−1.0	0.159	1.7	0.955	1.6	0.890
−0.9	0.184	1.8	0.964	1.7	0.910
−0.8	0.212	1.9	0.971	1.8	0.928
−0.7	0.242	2.0	0.977	1.9	0.942
−0.6	0.274	2.1	0.982	2.0	0.954
−0.5	0.308	2.2	0.986	2.1	0.964
−0.4	0.345	2.3	0.989	2.2	0.972
−0.3	0.382	2.4	0.992	2.3	0.978
−0.2	0.421	2.5	0.994	2.4	0.984
−0.1	0.460	2.6	0.995	2.5	0.988
−0.0	0.500	3.1	0.999	2.6	0.990

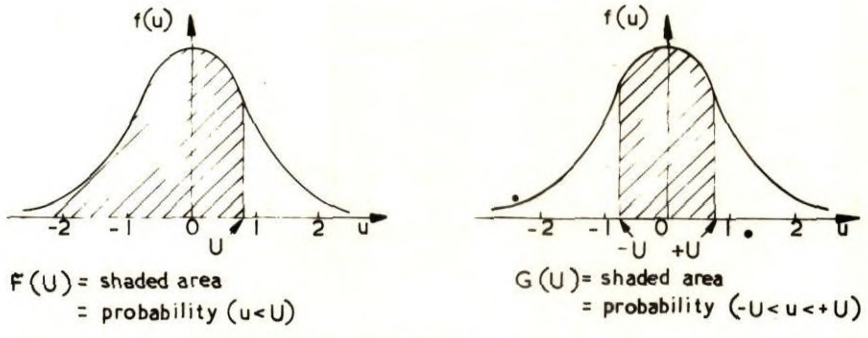


Figure 12. Normal (Gaussian) distribution function

Table 12. Table of percentage points of Student's t-distribution

Degrees of freedom (γ)	For use in double-sided significance test (or confidence interval)			For use in single-sided significance test (or confidence interval)		
	Percentage point			Percentage point		
	1 %	5 %	10 %	1 %	5 %	10 %
1	63.66	12.71	6.31	31.82	6.31	3.08
2	9.92	4.30	2.92	6.97	2.92	1.89
3	5.84	3.18	2.35	4.54	2.35	1.64
4	4.60	2.78	2.13	3.75	2.13	1.53
5	4.03	2.57	2.02	3.37	2.02	1.48
6	3.71	2.45	1.94	3.14	1.94	1.44
7	3.50	2.36	1.89	3.00	1.89	1.42
8	3.36	2.31	1.86	2.90	1.86	1.40
9	3.25	2.26	1.83	2.82	1.83	1.38
10	3.17	2.23	1.81	2.76	1.81	1.37
11	3.11	2.20	1.80	2.72	1.80	1.36
12	3.05	2.18	1.78	2.68	1.78	1.36
13	3.01	2.16	1.77	2.65	1.77	1.35
14	2.98	2.15	1.76	2.62	1.76	1.35
15	2.95	2.13	1.75	2.60	1.75	1.34
16	2.92	2.12	1.75	2.58	1.75	1.34
17	2.90	2.11	1.74	2.57	1.74	1.33
18	2.88	2.10	1.73	2.55	1.73	1.33
19	2.86	2.09	1.73	2.54	1.73	1.33
20	2.85	2.09	1.72	2.53	1.72	1.33
25	2.79	2.06	1.71	2.49	1.71	1.32
30	2.75	2.04	1.70	2.46	1.70	1.31
40	2.70	2.02	1.68	2.42	1.68	1.30
60	2.66	2.00	1.67	2.39	1.67	1.30
120	2.62	1.98	1.66	2.36	1.66	1.29
∞	2.58	1.96	1.64	2.33	1.64	1.28

Table 13. 5 % points of F distribution

γ_2 \ γ_1	1	3	5	7	8	10	12	24	∞
1	161.4	215.7	230.2	236.8	238.9	241.9	243.9	249.0	254.3
3	10.13	9.28	9.01	8.89	8.85	8.79	8.74	8.64	8.53
5	6.61	5.41	5.05	4.88	4.82	4.74	4.68	4.53	4.36
7	5.59	4.35	3.97	3.79	3.73	3.64	3.57	3.41	3.23
8	5.32	4.07	3.69	3.50	3.44	3.35	3.28	3.12	2.93
10	4.96	3.71	3.33	3.14	3.07	2.98	2.91	2.74	2.54
12	4.75	3.49	3.11	2.91	2.85	2.75	2.69	2.51	2.30
24	4.26	3.01	2.62	2.42	2.36	2.25	2.18	1.98	1.73
∞	3.84	2.60	2.21	2.01	1.94	1.83	1.75	1.52	1.00

Standards publications referred to

BS 600 Application of statistical methods to industrial standardization and quality control
BS 903 Methods of testing vulcanized rubber
Part A2 Determination of tensile stress-strain properties
Part A3 Determination of tar strength (crescent test piece)
Part A9 Determination of abrasion resistance
Part A10 Determination of resistance to flex cracking
Part A11 Determination of resistance to crack growth
Part A13 Determination of the stiffness of vulcanized rubbers at low temperatures (Gehman test)
Part A23 Determination of resistance to ozone cracking under static conditions
BS 2564 Control chart technique when manufacturing to a specification, with special reference to articles machined to dimensional tolerances
BS 1673 Methods of testing raw rubber and unvulcanized compounded rubber
Part 1 Sampling
BS 2846 Guide to the statistical interpretation of data
BS 2987 Notes on the application of statistics to paper testing
BS 5309 Sampling chemical products
Part 1 Introduction and general principles
Part 4 Sampling of solids
BS 6000 Guide to the use of 6001. Sampling procedures and tables for inspection by attributes
ISO/R 645 Statistical vocabulary and symbols. First series of terms and symbols
Part 1 Statistical vocabulary
ISO/R 1786 Statistical vocabulary and symbols. Second series of terms and symbols

This British Standard, having been prepared under the direction of the Rubber Industry Standards Committee, was published under the authority of the Executive Board on 31 August 1976.

ISBN: 0 580 09450 2

Contract requirements
Attention is drawn to the fact that this British Standard does not purport to include all the necessary provisions of a contract.

Revision of British Standards
British Standards are revised, when necessary, by the issue either of amendment slips or of revised editions. **It is important that users of British Standards should ascertain that they are in possession of the latest amendments or editions.**

The following BSI references relate to the work on this standard:
Committee references RUC/10, RUC/10/4, RUC/10/4/3
Draft for comment 72/54662DC

Cooperating organizations

The Rubber Industry Standards Committee, under whose supervision this British Standard was prepared, consists of representatives from the following Government departments and scientific and industrial organizations:

British Association of Synthetic Rubber Manufacturers
* British Rubber Manufacturers' Association Ltd.
Department of Industry
* Malaysian Rubber Producers Research Association
* Ministry of Defence
* Rubber and Plastics Research Association of Great Britain
Rubber Growers' Association
* Society of Motor Manufacturers and Traders Ltd.

The Government department and scientific and industrial organizations marked with an asterisk in the above list, together with the following, were directly represented on the committee entrusted with the preparation of this British Standard:

British Railways Board
Chemical Industries Association
Department of the Environment
Electrical Research Association
Institution of Mechanical Engineers
Institution of Municpial Engineers
Institution of Water Engineers and Scientists
National College of Rubber Technology
Post Office
Royal Institute of Chemistry
Individual experts

Amendments issued since publication

Amd. No.	Date of issue	Text affected

British Standards Institution · 2 Park Street London W1A 2BS · Telephone 01-629 9000 · Telex 266933

7608 –2k–B