SCAN-G 2:07
Revised 2007

SCANDINAVIAN PULP, PAPER AND BOARD
TESTING COMMITTEE

Pulp, paper and board

Statistical treatment of test results

0 Introduction

The value of statistical methods lies in the fact that they make it possible to interpret test results according to strictly objective criteria. A statistical analysis of test data does not increase the experimental precision by transforming uncertain results to certainties, but it does make it possible to express numerically, in the form of definite probabilities, the significance to be attached to the conclusions drawn from the results.

The purpose of this Guideline is to give a brief description of those statistical methods which are recommended for use in the treatment of test data derived according to SCAN-test Methods and to promote uniformity in the use of statistical terms and symbols, and in the mode of expressing test results.

This SCAN-test Guideline replaces SCAN-G 2:63. This revised version differs from the earlier version in that certain equations which were considered helpful before computerised help was available have now been omitted. In addition, the sections relating to the calculation of uncertainties have been extended, and information is provided as an aid to the development of precision statements in SCAN-test Methods.

This Guideline can with advantage be read in conjunction with SCAN-G 6:00 *Uncertainty of results from physical testing.* The ISO Technical report ISO/TR 24498 *Paper, board and pulps – Estimation of uncertainty for test methods* may also be helpful.

1 Scope

This SCAN-test Guideline gives a brief description of the simple statistical methods commonly used in the treatment of test data and provides the necessary equations. It does not give detailed derivations of the equations, nor does it provide proofs of the theorems presented. The equations are presented in a manner which enables persons without statistical training to apply the methods, but it is recommended that a person with statistical training and experience be consulted before conclusions are drawn from the results.

A number of numerical examples are presented in Annex A as an illustration of how the methods are to be applied.

2 Terms and definitions

2.1 *Measurement* – processing of a single test piece or test portion

> *Note* – The number of measurements required for a test is usually stated in the method.

2.2 *Result of a measurement* – value of the property obtained by a single measurement

2.3 *Test* – complete procedure, including preparation of the test material, performing the number of measurements required, and making the necessary calculations

2.4 *Result of a test* – value (e.g. mean, standard deviation etc) reported for the test and calculated from the results of all the measurements

2.5 *Population* – finite or infinite amount of material or number of units

> *Note* – It is normally not practical to measure every unit in a population, and a *sample* must therefore be taken.

2.6 *Sample* – limited amount of material taken from a population

> *Note 1* – The material is taken for the purpose of providing a finite number of test pieces on which a certain property is to be measured, in order to obtain an estimate of the value of that property in the material from which the sample is taken.

> *Note 2* – In this Guideline, the term *sample* is used in its statistical sense. In physical or chemical test methods this word may have another meaning.

2.7 *Random sample* – limited amount of material selected at random from a population, i.e. in such a manner that any unit in the population has an equal probability of being selected, so that the sample may be considered to be fully representative of the population from which it is taken

2.8 *Test piece* – piece or quantity of material taken from the sample for use in a single measurement in a chemical or physical test

2.9 *Statistic* – single value calculated from and used to represent a set of measurement results, usually as an estimate of some parameter of a population from which a sample has been taken

2.10 *Mean, arithmetic mean,* μ or $\overline{x}$ – a statistic describing the population calculated as the sum of the individual measurement results divided by the number of measurements

> *Note* – the symbol μ is used to denote the mean of the population, and the symbol $\overline{x}$ is used to denote the mean of a set of measurements, and thus the estimated mean of the population.

2.11 *Median,* x_M – statistic describing a set of measurements chosen so that 50 % of the values are above and 50 % of the values are below x_M

2.12 *Range* – difference between the largest and the smallest values of a set of measurement results

2.13 *Dispersion* – measure of the extent to which the results of the individual measurements are scattered about the mean

2.14 *Variance of a population,* σ^2 – sum of the squares of the deviations of the individual values of the property from the calculated mean divided by the number of measurements, n

2.15 *Variance of a sample,* s^2 – sum of the squares of the deviations of the individual values of the property from the calculated mean divided by a factor equal to one less than the number of measurements, $(n-1)$

> *Note* – Division by this factor ensures that the calculated variance of the sample is an unbiased estimate of the variance of the population from which the sample is taken.

2.16 *Standard deviation,* σ *or* s – square root of the variance

> *Note* – This is the most commonly used measure of the dispersion. The symbol σ is used to denote the standard deviation of the population, and the symbol *s* is used to denote the standard deviation of a set of measurements, and thus the estimated standard deviation of the population.

2.17 *Coefficient of variation, CoV* – standard deviation divided by the mean

> *Note* – The coefficient of variation is expressed as a percentage. It is sometime referred to as the relative standard deviation.

2.18 *Straggler* – member of a set of values which is unusually high or unusually low and is inconsistent with the other members of that set at a 5 % probability level

> *Note* – If the result of a statistical test shows that a result is a straggler but not an outlier, it should normally not be rejected.

2.19 *Outlier* – member of a set of values which is extremely high or extremely low and is inconsistent with the other members of that set at a 1 % probability level

> *Note* – If the result of a statistical test shows that a result is an outlier, this result should be excluded from the subsequent statistical calculations.

2.20 *Significance* – extent to which the data indicate that the observed effect has a given probability of not being due solely to chance

Note – The level of significance is usually expressed either as the probability *p* that the given result is due to chance or, more commonly, as the probability (1 – *p*) that the result is not due to chance. Significance tests are frequently applied for p < 0,05, i.e. for (1 – *p*) > 0,95, commonly called the 95 % significance level. In critical situations, a significance of 99 % (*p* <0,01) may be required.

2.21 *Degree of freedom* – number of independent comparisons which can be made between the members of a sample

Note – The number of degrees of freedom is usually *n-1* where *n* is the number of independent measurements of tests being considered.

2.22 *Confidence limits* – limits which define with a given probability the range within which a given statistic is estimated to lie

Note – The confidence limits most commonly referred to are those associated with the estimated mean of a population calculated from measurements made on a limited sample taken from the population.

2.23 *Student's t-distribution* – probability of a continuous random variable used to assess the significance of a measured statistic.

Note 1 – Values of the distribution provided in tabular form for different degrees of freedom are used in the calculation of *confidence limits.*

Note 2 – This distribution was published in 1908 by "Student", the pseudonym of W S Gosset.

2.24 *Uncertainty* – measure of the uncertainty associated with a given statistic

Note 1 – In the case of the mean, the simple uncertainty is equal to the estimated standard deviation of the population.

Note 2 – Information about the calculation of combined uncertainties associated with different sources of uncertainty is given in SCAN-G 6.

2.25 *Expanded uncertainty, U* – simple uncertainty multiplied by a coverage factor, *k*, so that $U = k \cdot s$

Note – In the case of the mean, the expanded uncertainty is analogous to the confidence limit.

2.26 *Repeatability conditions* – conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within a short interval of time

Note – In the testing of pulps, paper and board, it is not possible to make measurements strictly under repeatability conditions if, as is often the case, the test is destructive.

2.27 *Repeatability standard deviation,* s_r – standard deviation of measurement results obtained under repeatability conditions

2.28 *Repeatability limit, r* – value less than or equal to which the absolute difference between two test results obtained under repeatability conditions is expected to be with a given probability

2.29 *Reproducibility conditions* – conditions where the test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.

Note – In the testing of pulps, paper and board, it is not possible to make measurements strictly under reproducibility conditions if, as is often the case, the test is destructive

2.30 *Reproducibility standard deviation,* s_R – standard deviation of test results obtained under reproducibility conditions

2.31 *Reproducibility limit, R* – value less than or equal to which the absolute difference between two test results obtained under reproducibility conditions is expected to be with a given probability

3 Distributions

In any population of different values of a variable, the different values occur with different frequencies. The nature of the population can thus be described by the *distribution* of the frequencies with which the different values of the variable occur. This *frequency distribution* can be illustrated graphically by a *frequency curve*.

When tests are carried out on a number of test pieces taken from a sample of the population, the values obtained are also distributed with different frequencies, which are estimates of the *probabilities* that these values occur with these frequencies in the population. The *probability distribution* thus describes the probable frequencies with which different events are expected to occur in a sample taken at random from the population.

This probability distribution can be described by a *probability curve* which will resemble the frequency curve of the total population.

The primary purpose of a statistical analysis is (a) to calculate suitable *statistics* to describe the distribution and (b) to assess the reliability of these *statistics* as a description of the population.

3.1 Histogram

If a set of measurements has been made on test pieces taken from a sample from the population, it is possible to obtain an approximate idea of the appearance of the frequency curve of the population by constructing a histogram, *Figure 1*, where measurement results are grouped in classes indicated by the figures at the base of each rectangle. The height of each rectangle represents the number of measurement results assigned to the interval.

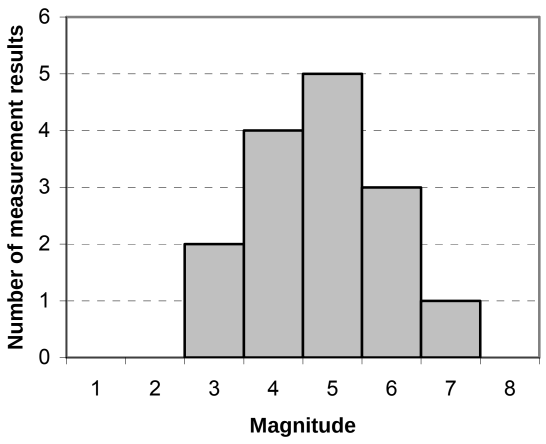


Figure 1. Histogram

Figure 1 is a histogram for a set of 15 measurements, the magnitudes of the results of which are between 2 and 7. Two of the results lie between 2 and 3, four between 3 and 4, five between 4 and 5, three between 5 and 6, and one between 6 and 7.

If this test is extended by increasing the number of measurements and if the class interval is decreased, the stepped form of the histogram will approach a smooth frequency curve. Such a frequency curve can often be given a definite mathematical form.

4 Measures of location

To describe a distribution in simple terms, some measure of the location of the distribution is required, i.e. some measure of the centre of the distribution.

4.1 Mean

The (arithmetic) mean or average value of a set of results of measurements of a given variable is calculated as the sum of the individual measurements divided by the number of measurements:

$$\bar{x} = \frac{(x_1 + x_2 + ... + x_n)}{n} = \frac{\sum x}{n} \qquad [1]$$

4.2 Weighted mean

If the data are sorted into groups with different numbers of values in the different groups, this must be taken into consideration when the mean is calculated.

The mean of a frequency distribution is calculated as the weighted mean where the different values recorded are multiplied by their respective frequencies and the sum is then divided by the total number of measurements:

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} \qquad [2]$$

where f_i are the frequencies of the values x_i .

4.3 Median

In some cases, particularly when a distribution is extremely skew, it is inappropriate to calculate the mean since this may be unduly affected by the magnitude of the extreme values. In this case it may be more appropriate to calculate the median x_m which is defined as:

$$x_1, x_2, x_3 ... x_n < x_m < x_n, x_{n+1}, x_{n+2} ... x_{2n} \qquad [3]$$

where x_1, x_2, x_3 …… are the results of the individual measurements arranged in ascending order.

If the number of measurements is odd, i.e. if the highest number is x_{2n+1} , the median is calculated as

$$x_m = x_{n+1} \qquad [4]$$

If the number of measurements is even, i.e. if the highest number is x_{2n} , the median is calculated as

$$x_m = \frac{(x_n + x_{n+1})}{2} \qquad [5]$$

> *Note* – It is possible to extend this procedure to divide the data into *quartile*s, where 25 % of the values are larger than the *upper quartile* and 75 % of the values are larger than the *lower quartile* etc.

5 Measures of dispersion

The mean alone is not sufficient to describe the character of a distribution. Some measure of the degree of dispersion around the mean is also required. The measure of dispersion mostly used is the standard deviation σ, which is the positive square root of the *variance*.

From a set of n measurement results $x_1\ x_2\ ...x_n$, the variance σ^2 of the population is estimated by calculating s^2 according to the expression:

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1} \qquad [6]$$

The term n-1 is used instead of n, since division by n has been shown to give a biased estimation of σ^2, particularly for low values of n.

A measure of the standard deviation in the population σ is consequently obtained by calculating:

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} \qquad [7]$$

In addition to the standard deviation, it is sometimes of interest to report the *relative standard deviation*, i.e. the magnitude of the standard deviation in relation to the mean value. This is called the *coefficient of variation*, Co*V*, and it is usually expressed as a percentage:

$$CoV = 100\frac{s}{\bar{x}} \qquad [8]$$

6 Tests for and rejection of outliers

Occasionally one member of a set of measurement values may appear to differ abnormally from the others. The question then arises as to whether or not this difference is because the abnormal value belongs to a different population. Several statistical criteria have been suggested for answering this question.

It is clear that a value which does not properly belong to the test series should not be included in the calculation of the mean or in any subsequent analysis, but it must be emphasised that extreme caution should be exercised before a value is rejected. A value should preferably be rejected only if it has been established that an actual mistake has been made. An unusually high or low value may often be due to the natural variation in the material and, if this value is rejected, the remaining values will give an incorrect picture of the distribution and particularly the standard deviation will be too small.

In order to test whether a value is unreasonably far from the mean, Grubb's test (cf. ISO 5725) can be applied to check whether an exceptionally high or an exceptionally low value x_i is to be regarded as an outlier. Grubb's statistic, G, is calculated as:

$$G = \frac{|x_i - \bar{x}|}{s} \qquad [9]$$

Critical values for Grubb's test at the 1% probability level are given in *Table 1*. If the value obtained for G is higher than the value given in the table, x_i is an outlier and can be rejected. Note that this procedure should be used only once for each set of data.

Table 1. Critical values for Grubb's test

No of values	Outliers p<0,01	Stragglers p<0,05
5	1,764	1,175
6	1,973	1,887
7	2,139	2,020
8	2,274	2,126
9	2,387	2,215
10	2,482	2,290
11	2,564	2,355
12	2,636	2,412
13	2,699	2,462
14	2,755	2,507
15	2,806	2,549
16	2,852	2,585
17	2,894	2,620
18	2,932	2,651
19	2,968	2,681
20	3,001	2,709
21	3,031	2,733
22	3,060	2,758
23	3,087	2,781
24	3,112	2,802
25	3,135	2,822

7 The normal distribution

Experience has shown that the results of sets of measurements are often distributed in a manner which shows a close agreement with the mathematically well-known *normal distribution* which has a bell-shaped form.

The equation for the frequency curve is a function of the mean μ and the standard deviation σ, *viz*.:

$$y = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)} \qquad [10]$$

where
y is the relative frequency;
x is the measured quantity.

The frequency curve is shown in *Figure 2*, where the measured quantities *x* are given as multiples of σ. The curve reaches its maximum at $x = \mu$, and is symmetrical around this point.

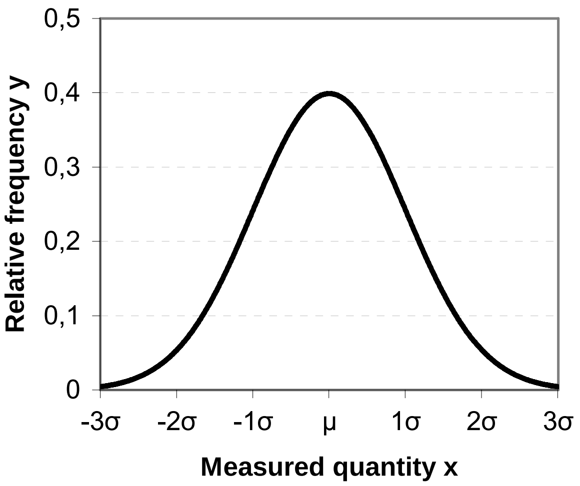


Figure 2. Normal frequency curve.

In the particular case where μ = 0 and σ = 1, the equation is reduced to:

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad [11]$$

This equation can be integrated up to a certain value, or between certain limits. In this way it is possible to calculate the probability that the result of a single measurement, *x,* will fall above or below a certain value or that *x* will occur between two given limits.

It is useful to note that the probability that a single value of a normally distributed variable deviates from its mean by more than s is 31,7 %, by more than 2σ is 4,6 %, and by more than 3σ is 0,3 %, and that a deviation of more than 4σ occurs on average only once in 10 000 measurements. Alternatively, it can be said that approximately 68 % of all values lie within ± σ from the mean, and that 95 % of all value lie within ± 2σ.

8 Accuracy and precision

It is important in the presentation of test results to distinguish between *accuracy* and *precision*.

The *accuracy* of a test result is a statement of the extent to which it conforms to the true value.

The *precision* of a test result is a statement of the extent to which the set of measurement results on which the test result is based are dispersed about the mean. The precision is expressed as the confidence limits, or as the uncertainty of the mean.

Both the accuracy and the precision can be reported in either absolute or relative measures. In any report it is important to state which measure has been used, especially if the results are expressed as percentages.

9 Systematic and random errors

All errors can be divided into two types, systematic and random errors.
A *systematic error* affects the accuracy of the method and of the values measured, but the existence and magnitude of a systematic error are normally unknown, since they can be estimated only by comparison of the calculated mean for a set of measurements with the “true” mean, which has been calculated or determined in an independent manner. If the magnitude of a systematic error is indeed known, steps will normally be taken to eliminate this error.

A *random error* is a measure of the precision or reproducibility of a method, the magnitude being indicated by the dispersion.

A systematic error is regular and may be due, for example, to a defect in the measuring device used, whereas a random error is irregular and is due to variations by chance in the measured results.

10 Random variations

In a statistical context, the term *random error* is often used to describe variations which are not errors but which are true variations in the material.

If a measurement can be repeated on a *single* test piece, the results of the repeat measurements will vary within a certain range (and the variations will usually be normally distributed) and this variation is the random error associated with the method or the instrument.

If measurements are made on a set of *different* test pieces, the variation in the results is due not only to any random error in the method or instrument but also to a real variation within the sample. These two possible causes cannot normally be separated.

Unless information to the contrary is available, it is usually correct in a pulp, paper or board context to assume that the differences among measurement results are indeed due to variations in the material. The standard deviation is thus often an important material property which should be reported together with the mean.

11 The confidence interval

After having obtained an estimate of a certain quantity on the basis of a test series, it may be necessary to state the degree of uncertainty associated with this estimate.

This is usually done by giving the limits within which the true value of the assessed magnitude is expected to lie with a certain specified degree of significance (100-*p*) %. These limits are called the *confidence limits*, the interval between them is called the *confidence interval*, and the value of *p* is the *level of significance*.

If n measurements have been made on a sample from a population having a normal distribution, and the mean value $\bar{x}$ and the standard deviation s have been calculated, it can be shown that the standard deviation associated with the mean s_m is given by

$$s_m = \frac{s}{\sqrt{n}} \qquad [12]$$

The (100-p) % confidence limits of the mean can then be expressed by the formula:

$$\pm \frac{t_p \cdot s}{\sqrt{n}} \qquad [13]$$

where t_p is a value related to the normal distribution which gives the probability that the true value lies within the given limits.

The appropriate value of t_p is obtained from a table of *Student's t distribution*. This value is dependent both upon the selected level of significance, p, and upon the number of measurements, N, on which the calculation of s is based. The number of degrees of freedom, f, is given by $f = n - 1$. Normally $N = n$.

The level of significance can be chosen at will, the levels most frequently used being those given in *Table 2*, *viz:* 5 %, 1 % and 0,1 %. Usually the 5 % level is recommended.

In this case, the expression

$$U = \pm \frac{t_p \cdot s}{\sqrt{n}} \qquad [14]$$

is also the *expanded uncertainty* associated with the mean, where t_p is the coverage factor ($p < 0,05$).

It should be noted here that the confidence limits can be made narrower, i.e. the precision of the estimate of the mean value can be increased, by increasing the number of measurements n, but this does not of course affect the variation in the material indicated by the standard deviation s.

12 Required number of measurements

With the aid of equation [14] for the uncertainty or confidence interval, it is possible to calculate how many measurements will be required in order to obtain an estimate of the mean for a certain variable with a given degree of precision. A condition for this, however, is that an estimate, s, of the standard deviation , σ , is available.

Assuming that the standard deviation, s, has been calculated from earlier measurements, and that we wish to estimate the mean of the population with a precision of ± a and (100 - p) % confidence, (in other words, we require that the (100 - p) % confidence interval for the mean value shall have a width of 2a), the approximate number of observations, n, required to fulfil these conditions is given by:

$$n = \left(\frac{t_p \cdot s}{a} \right)^2 \qquad [15]$$

where the appropriate value of t_p is taken from *Table 2*.

Table 2 Student's t-distribution

Degrees of freedom, f	p		
	5%	1%	0,1%
4	2,78	4,60	8,61
5	2,57	4,03	6,86
6	2,45	3,71	5,96
7	2,36	3,50	5,40
8	2,31	3,36	5,04
9	2,26	3,25	4,78
10	2,23	3,17	4,59
11	2,20	3,11	4,44
12	2,18	3,06	4,32
13	2,16	3,01	4,22
14	2,14	2,98	4,14
15	2,13	2,95	4,07
16	2,12	2,92	4,02
17	2,11	2,90	3,97
18	2,10	2,88	3,92
19	2,09	2,86	3,88
20	2,09	2,84	3,85
21	2,08	2,83	3,82
22	2,07	2,82	3,79
23	2,07	2,81	3,77
24	2,06	2,80	3,74
25	2,06	2,79	3,73
26	2,06	2,78	3,71
27	2,05	2,77	3,69
28	2,05	2,76	3,67
29	2,05	2,76	3,66
30	2,04	2,75	3,65
40	2,02	2,70	3,55
60	2,00	2,66	3,46
120	1,98	2,62	3,37
∞	1,96	2,58	3,29

13 Repeatability limits / reproducibility limits

It is not always the precision of the mean that is the matter of greatest concern. Sometimes, it is important to use the statistical data available to provide an estimate of the maximum expected difference, with a given degree of probability, between two items taken at random from a population.

In this case, if the expanded uncertainty associated with a single item is $U = t_p \cdot s$, then the maximum expected difference is equal to:

$$\Delta = U \cdot \sqrt{2} \qquad [16]$$

This equation can also be used to assess the repeatability limits, i.e. the maximum expected difference between the means of two sets of measurements made on material from the same population.

$$\Delta = \frac{\sqrt{2} \cdot t_p \cdot s}{\sqrt{n}} \qquad [17]$$

For measurements made under reproducibility conditions in different laboratories, the maximum expected difference between two sets of measurements is equal to

$$\Delta = \sqrt{2} \cdot t_p \cdot s^* \qquad [18]$$

where s^* is an appropriately calculated standard deviation derived from a consideration of the various sources of variance in different instruments, different laboratories etc.. This is discussed more fully in ISO/TR 24498 and in Clause 17.

14 Comparison of two means (t-test)

An associated task, often encountered in practice, is to compare the means obtained from two different series of observations, and to assess whether or not, at a given probability level, they can be considered to come from the same population, e.g, whether two batches of nominally the same material differ significantly in their properties.

Consider the case where $\bar{x}_1$ and $\bar{x}_2$ are the means of the series 1 and 2, n_1 and n_2 are the numbers of individual measurements in each of the two series, and s_1 and s_2 are the calculated standard deviations.

If there is no great difference between s_1 and s_2 and there is no cause for assuming that there is any essential difference in the dispersions of the populations, a combined measure of the standard deviation based on all the measurement results, s, can be calculated according to the formula:

$$s = \sqrt{\frac{\sum(x_1 - \bar{x}_1)^2 + \sum(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}} \qquad [19]$$

The statistic t is then calculated according to:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad [20]$$

This value of t is then compared with the value t_p (*Table 2*) corresponding to the p % level of significance and the number of degrees of freedom f given by:

$$f = n_1 + n_2 - 2 \qquad [21]$$

If the value of t is greater than the value t_p, there is a significant difference between $\bar{x}_1$ and $\bar{x}_2$ and the two samples cannot be considered to have come from the same population.

> *Note* – It can never be established that the two samples do in fact come from the same population; only that if they come from different populations then the difference between the two populations is less than $|x_1 - x_2|$.

15 Additivity of variances

The assessment of the significance of a test result or the uncertainty associated with the result is often more complicated than a mere calculation of the mean and standard deviation of the measurement data. If the uncertainty is to be assessed in relation to the results of other tests in the same laboratory or of tests in different laboratories, other sources of variation and uncertainty must be considered. The basic principle to be observed in such cases is that, provided the different sources of variance are independent and uncorrelated, the variances are additive, i.e.:

$$s_{total}^2 = s_1^2 + s_2^2 + \ldots \qquad [22]$$

In this context, it is also important to note that if a mean standard deviation is required this must always be calculated as the root mean square:

$$s_{mean} = \sqrt{\frac{\left(s_1^2 + s_2^2 + \ldots + s_j^2\right)}{j}} \qquad [23]$$

where j is the number of items in the series. It is not correct merely to calculate the mean of the standard deviations.

16 Repeatability limits

When two independent tests are carried out under repeatability conditions, i.e. within the same laboratory by the same operator using the same equipment on the same occasion, the 95% probability that the two test results will not differ by an amount greater than Δ is

$$\Delta = \frac{t_p \cdot \sqrt{2} \cdot s}{\sqrt{n}} \quad [24]$$

provided that the only source of variation is in the material. In equation [24], s is an overall standard deviation calculated in accordance with equation [19].

This is not however a realistic situation. In general, under normal laboratory conditions within a paper mill, there are also other sources of variation between tests, which are associated with a standard deviation between tests, s_{bt}, so that the total within-laboratory standard deviation, s_{wl}, is given by:

$$s_{wl}^2 = s_{bt}^2 + \frac{s_{wt}^2}{n} \quad [25]$$

where s_{wt} is the within-test standard deviation and n is the number of measurements in each test. It is possible to analyse the results and to determine separately the value of s_{bt}, but this is often not necessary.

The repeatability standard deviation, s_r, is equal to s_{wl} and the repeatability limits, r, are calculated as:

$$r = 1{,}96 \cdot \sqrt{2} \cdot s_r \quad [26]$$

which can be simplified to the expression:

$$r = 2{,}77 \cdot s_r \quad [27]$$

> *Note 1* – The standard deviation calculated directly from the results of different tests carried out in this manner is the repeatability standard deviation. It is *not* the between-test standard deviation which is defined as the component of the repeatability deviation which is independent of variations in the material.

> *Note 2* – In SCAN-G 6, the quantity here referred to as s_{wl} has the designation s_{bts}.

In SCAN-G 6, this discussion is extended to include other sources of variation within a laboratory where the test is destructive and where tests are no longer carried out under repeatability conditions. In this case, if repeated tests are carried out on homogeneous material from the same batch on different occasions, it is possible to calculate the "long-term within-laboratory reproducibility limits" in an analogous manner if the total standard deviation is calculated.

17 Reproducibility limits

Similarly, if similar tests are carried out in different laboratories with different items of equipment, other sources of variation will introduce additional uncertainties. In this case, the total variance is called the reproducibility variance and it may be expressed as:

$$s_R^2 = s_{bl}^2 + s_{bt}^2 + \frac{s^2}{n} \quad [28]$$

and the reproducibility limits, R, are given by:

$$R = 1{,}96 \cdot \sqrt{2} \cdot s_R \quad [29]$$

> *Note* – s_R is not equal to the between-laboratory standard deviation s_{bl}, which is only one component of the total reproducibility uncertainty.

18 Within-laboratory standard deviation outliers

When a comparison is made involving different laboratories, it is sometimes necessary to exclude a laboratory that shows an unreasonably high within-laboratory standard deviation.

In order to determine whether an abnormally large standard deviation is statistically an outlier, Cochran's test according to ISO 5725, with a rejection level of 1 %, can be applied. The test shall be applied only to the laboratory having the highest deviation.

Cochran's statistic, C, is calculated as:

$$C = \frac{s_{i.\max}^2}{\left[\sum_{i=1}^{P} s_i^2\right]} \quad [30]$$

where P is the number of laboratories in the comparison.

Critical values for Cochran's test are given in *Table 3*. A laboratory should be excluded from the comparison if the value of C obtained is higher than the value given in the table.

Any laboratory excluded on the basis of too high a standard deviation shall be completely excluded from the subsequent analysis.

Table 3. Critical values for Cochran's test (p<0,01)

No of labs	No of measurements in each test, *n*	
	5	10
5	0,633	0,485
6	0,564	0,423
7	0,508	0,375
8	0,463	0,337
9	0,425	0,307
10	0,393	0,281
11	0,366	
12	0,343	0,242
13	0,322	
14	0,304	
15	0,288	0,200
16	0,274	
17	0,261	
18	0,249	
19	0,239	
20	0,229	0,157
21		
22		
23		
24	0,197	0,134
25		

19 Reporting test results

The result of a series of measurements of a property should generally be reported by:

(a) the mean;
(b) the number of measurements, *n;*
(c) the standard deviation, *s;*
(d) the 95 % confidence interval of the mean, or the expanded uncertainty of the mean;
(e) whether any outliers have been rejected and the criterion of rejection that has been applied.

Note 1 – If the data have been transformed so that the confidence interval has become asymmetrical in relation to the mean (e.g. as in the case of a skew distribution), the mean value, the two confidence limits and the number of measurements should be reported, as well as the mode of calculation used.

Note 2 – In certain cases, such as with skew distributions, it may be suitable to report the median value, the range of the variation, and the number of measurements.

20 Literature

1. SCAN-G 6:00 Paper, board and pulp – Uncertainty of results from physical testing

2. ISO/TR 24498:2006 Paper, board and pulps – Estimation of uncertainty of test methods

3. ISO 5725-1:1994 Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions

4. ISO 5725-2:1994 Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method

Annex – Numerical examples

A.0. General

The following numerical examples are given as an added illustration of the applications indicated in the text. They also provide data which can be used to check the accuracy of calculation programmes from other sources or those written internally.

A.1. A sample from a population with a normal distribution

Fifteen individual measurements of a variable with a normal distribution have been obtained in a test as shown in Table A.1.

Table A.1

4,10	4,37	4,51
4,24	4,45	4,59
4,28	4,44	4,66
4,31	4,47	4,70
4,36	4,50	4,75

Calculate in turn:

1. $\sum x$ .... = 66,73
2. $\bar{x} = \frac{\sum x}{n}$ .... = 4,449
3. $\sum x^2$ .... = 297,3119
4. $s^2 = \frac{\sum (x-\bar{x})^2}{(n-1)}$ .... = 0,0323
5. s .... = 0,180
6. $CoV = \frac{100 \cdot s}{\bar{x}}$ .... = 4,0 %
7. $f = n-1$ .... = 14
8. t_5 for $f = 14$ .... = 2,14
9. $U = \frac{t_5 \cdot s}{\sqrt{n}}$ .... = 0,0997

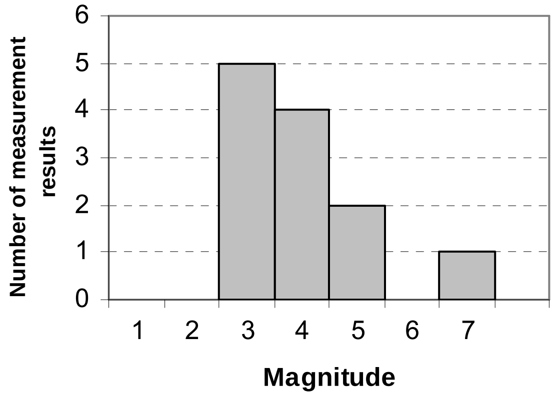
A.2. A population with a skew distribution

The following twelve individual measurements of a variable with an unknown distribution have been obtained in a test, as shown in Table A.2.

Table A.2

2,07	2,79	3,37	4,40
2,37	2,84	3,41	4,84
2,64	3,23	3,78	6,30

Draw a histogram in order to obtain an idea of the type of distribution.



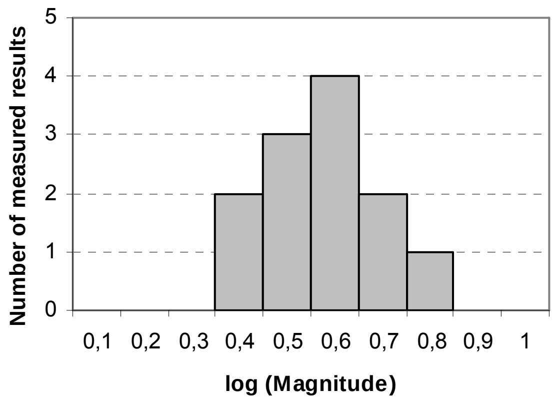
The histogram indicates a skew distribution. An ordinary calculation of the mean and the dispersion may then be inappropriate.

Try a transformation, e.g. by taking the base-ten logarithms, as in Table A.3.

Table A.3

0,316	0,446	0,528	0,642
0,375	0,453	0,533	0,685
0,422	0,509	0,577	0,799

Draw another histogram for the transformed variable.



The distribution obtained by this transformation appears to be approximately normal, and the computation can therefore be continued with the logarithmic values:

1. $\sum x$ .... = 6,286
2. $\bar{x}$ .... = 0,5238

3. $\sum x^2$ .... = 3,498648
4. $s^2 = \frac{\sum(x-\bar{x})^2}{(n-1)}$ .... = 0,01871
5. s .... = 0,137
6. $f = n-1$ .... = 11
7. t_5 for $f = 11$ .... = 2,20
8. $U = \frac{t_5 \cdot s}{\sqrt{n}}$ .... = 0,0777

Reverting to the original scale, we obtain
$\bar{x}' = (anti\log 0,5238)$ .... = 3,34
The lower 95 % confidence limit =
$anti\log(0,5238 - 0,0777)$ .... = 2,79
The upper 95 % confidence limit =
$anti\log(0,5238 + 0,0777)$ .... = 4,00

The confidence interval is thus asymmetrical around this weighted mean, $\bar{x}'$. In this case, the coefficient of variation is difficult to interpret, and hence the following is reported:

$\bar{x}' = 3,34 \quad (n = 12)$
95 % confidence interval 2,79 to 4,00.

Note – The calculation of values that had not been converted to logarithms would have given a mean of 3,50, and the confidence limits 2,77 and 4,23, i.e. a shift towards higher values which in reality are less representative of the distribution.

A.3. Required number of measurements

If it is assumed that the data obtained are related to a population with a normal distribution, it is possible to calculate the number of measurements required in order to reduce the uncertainty of the estimated mean of the population to less than a given value.

Consider the case where it is required to assess the mean of the population with a precision of ± 0,10 with 95 % confidence. Preliminary tests show that the standard deviation is approximately 0,25. How many test pieces shall be measured?

Table 2 gives, for
$f = \infty$
$t_5 = 1,96$
Calculate
$n = \left(\frac{t_5 \cdot s}{a}\right)^2 = \left(\frac{1,96 \cdot 0,25}{0,10}\right)^2$ .... = 24,01

Approximately 25 measurements should thus be made.

A.4. The significance of the difference between two means

Sets of 10 measurements of a normally distributed variable have been made on each of two samples of paper, as in Table A.4.

Table A.4

Measurement No	Sample 1	Sample 2
1	30	26
2	23	22
3	26	25
4	24	24
5	28	26
6	24	24
7	25	26
8	28	19
9	24	23
10	28	27
Mean	26,0	24,2
Std.dev.	2,357	2,394

The task is to determine whether these two samples can be regarded as having been drawn from the same population, i.e. whether the difference between the two means may be due to chance, or whether there is a statistically significant difference between them.

The calculated values of the means and standard deviations are shown in the table. Since the two standard deviations are very similar, the combined standard deviation can be used.

Calculate:

$s = \sqrt{\frac{\sum(x_1-\bar{x}_1)^2 + \sum(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$ .... = 2,376

Calculate:

$t_5 = \frac{(x_1 - x_2)}{s \cdot \sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ .... = 1,694

The value of t_5 given in Table 2 for $f = (n + n_2 - 2) = 18$ is 2,10.

Since the calculated value of t_5 is less than this value, the difference between the two means is not significant and the two samples can very well have been taken from the same population.

A.5. Rejection of extreme values

Twelve laboratories have each carried out a test on material supplied from a single batch. Each test consists of ten measurements and the results are shown in Table

A.5, as the mean and standard deviation for each laboratory.

Table A.5

Laboratory	Mean	Std.dev.
1	52,6	3,5
2	54,4	3,7
3	54,8	3,3
4	55,6	5,2
5	56,2	3,8
6	56,8	3,0
7	57,2	3,6
8	57,4	3,2
9	58,6	3,6
10	60,0	3,5
11	62,2	3,8
12	75,8	3,4

In the third column of this table, the value of the standard deviation reported from laboratory No 4 appears to be unusually high. To check whether this is a statistical outlier, calculate Cochran's statistic C

$$C = \frac{(5,2)^2}{\sum s^2} \quad \text{..........} = 0,167$$

According to Table 3, the critical value for 12 laboratories with 10 measurements in each test is 0,242. Since the value obtained is less than this critical value, there is no need to consider this laboratory to be an outlier.

In the second column of Table A.5, the value of the mean reported by Laboratory No 12 seems to be unusually high. To check whether this is a statistical outlier, calculate Grubb's statistic G

$$G = \frac{(75,8 - \bar{x})}{s} \quad \text{..........} = 2,874$$

According to Table 1, the critical value for an outlier for 12 laboratories is 2,636. Since the value obtained is higher than this critical value, this laboratory is a statistical outlier and should be eliminated from subsequent calculations.

A.6 Calculation of reproducibility limits

After elimination of laboratory 12 as indicated above, the data shown in Table A.6 are obtained.

Table A.6

Laboratory	Mean	Std.dev.
1	52,6	3,5
2	54,4	3,7
3	54,8	3,3
4	55,6	5,2
5	56,2	3,8
6	56,8	3,0
7	57,2	3,6
8	57,4	3,2
9	58,6	3,6
10	60,0	3,5
11	62,2	3,8
Mean	**56,89**	**3.695**
Std.dev.	**2,688**	

In this table, the mean is calculated as:

$$\frac{\sum x_i}{n} \quad \text{..........} = 56,89$$

the mean standard deviation is calculated as:

$$\sqrt{\frac{\sum s_i^2}{n}} \quad \text{..........} = 3,695$$

and the standard deviation of the mean is calculated as

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad \text{..........} = 2,688$$

The reproducibility standard deviation is thus 2,69, and the reproducibility limits are calculated as

$$R = 1,96 \cdot \sqrt{2} \cdot s \quad \text{..........} = 7,45$$

Note – The reproducibility limits include a contribution from the variation within the material. If it is of interest to calculate the standard deviation between laboratories disregarding the contribution from the test material, this can be calculated as:

$$s_{bl} = \sqrt{\left[2,688^2 - \left(\frac{3,695^2}{10}\right)\right]} \quad \text{..........} = 2,42$$

SCAN-test Methods are issued and recommended by KCL, PFI and STFI-Packforsk for the pulp, paper and board industries in Finland, Norway and Sweden. Distribution: Secretariat, Scandinavian Pulp, Paper and Board Testing Committee, Box 5604, SE-114 86 Stockholm, Sweden