INTERNATIONAL STANDARD

ISO 14802

First edition
2012-07-15

Corrosion of metals and alloys — Guidelines for applying statistics to analysis of corrosion data

Corrosion des métaux et alliages — Lignes directrices pour l'application des statistiques à l'analyse des données de corrosion

Reference number
ISO 14802:2012(E)

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

Published in Switzerland

Contents

Page

Foreword

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

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

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

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

ISO 14802 was prepared by Technical Committee ISO/TC 156, *Corrosion of metals and alloys.*

Corrosion of metals and alloys — Guidelines for applying statistics to analysis of corrosion data

1 Scope

This International Standard gives guidance on some generally accepted methods of statistical analysis which are useful in the interpretation of corrosion test results. This International Standard does not cover detailed calculations and methods, but rather considers a range of approaches which have applications in corrosion testing. Only those statistical methods that have wide acceptance in corrosion testing have been considered in this International Standard.

2 Significance and use

Corrosion test results often show more scatter than many other types of tests because of a variety of factors, including the fact that minor impurities often play a decisive role in controlling corrosion rates. Statistical analysis can be very helpful in allowing investigators to interpret such results, especially in determining when test results differ from one another significantly. This can be a difficult task when a variety of materials are under test, but statistical methods provide a rational approach to this problem.

Modern data reduction programs in combination with computers have allowed sophisticated statistical analyses to be made on data sets with relative ease. This capability permits investigators to determine whether associations exist between different variables and, if so, to develop quantitative expressions relating the variables.

Statistical evaluation is a necessary step in the analysis of results from any procedure which provides quantitative information. This analysis allows confidence intervals to be estimated from the measured results.

3 Scatter of data

3.1 Distributions

When measuring values associated with the corrosion of metals, a variety of factors act to produce measured values that deviate from expected values for the conditions that are present. Usually the factors which contribute to the scatter of measured values act in a more or less random way so that the average of several values approximates the expected value better than a single measurement. The pattern in which data are scattered is called its distribution, and a variety of distributions such as the normal, log–normal, bi-nominal, Poisson distribution, and extreme-value distribution (including the Gumbel and Weibull distribution) are observed in corrosion work.

3.2 Histograms

A bar graph, called a histogram, may be used to display the scatter of data. A histogram is constructed by dividing the range of data values into equal intervals on the abscissa and then placing a bar over each interval of a height equal to the number of data points within that interval.

The number of intervals, k, can be calculated using the following equation:

$$k = 1 + (3{,}32) \log n \tag{1}$$

where

n is the total number of data.

3.3 Normal distribution

Many statistical techniques are based on the normal distribution. This distribution is bell-shaped and symmetrical. Use of analysis techniques developed for the normal distribution on data distributed in another manner can lead to grossly erroneous conclusions. Thus, before attempting data analysis, the data should either be verified as being scattered like a normal distribution or a transformation should be used to obtain a data set which is approximately normally distributed. Transformed data may be analysed statistically and the results transformed back to give the desired results, although the process of transforming the data back can create problems in terms of not having symmetrical confidence intervals.

3.4 Normal probability paper

3.4.1 If the histogram is not confirmatory in terms of the shape of the distribution, the data may be examined further to see if it is normally distributed by constructing a normal probability plot as follows (see Reference [2]).

3.4.2 It is easiest to construct a normal probability plot if normal probability paper is available. This paper has one linear axis and one axis which is arranged to reflect the shape of the cumulative area under the normal distribution. In practice, the “probability” axis has 0,5 or 50 % at the centre, a number approaching 0 % at one end, and a number approaching 1,0 or 100 % at the other end. The scale divisions are spaced close in the centre and wider at both ends. A normal probability plot may be constructed as follows with normal probability paper.

NOTE Data that plot approximately on a straight line on the probability plot may be considered to be normally distributed. Deviations from a normal distribution may be recognized by the presence of deviations from a straight line, usually most noticeable at the extreme ends of the data.

3.4.2.1 Rearrange the data in order of magnitude from the smallest to the largest and number them as 1,2, … i, … n, which are called the rank of the points.

3.4.2.2 In order to plot the ith ranked data on the normal probability paper, calculate the ”midpoint” plotting position, $F(x_i)$, defined by the following equation:

$$F(x_i) = \frac{100(i - ½)}{n} \tag{2}$$

3.4.2.3 The data points [x_i, $F(x_i)$] can be plotted on the normal probability paper.

NOTE Occasionally, two or more identical values are obtained in a set of results. In this case, each point may be plotted, or a composite point may be located at the average of the plotting positions for all identical values.

It is recommended that probability plotting be used because it is a powerful tool for providing a better understanding of the population than traditional statements made only about the mean and standard deviation.

3.5 Other probability paper

If the histogram is not symmetrical and bell-shaped, or if the probability plot shows non-linearity, a transformation may be used to obtain a new, transformed data set that may be normally distributed. Although it is sometimes possible to guess the type of distribution by looking at the histogram, and thus determine the exact transformation to be used, it is usually just as easy to use a computer to calculate a number of different transformations and to check each for the normality of the transformed data. Some transformations based on known non-normal distributions, or that have been found to work in some situations, are listed as follows:

$y = \log x$ $\quad$ $y = \exp x$

$y = x^{0,5}$ $\quad$ $y = x^2$

$y = 1/x$ $\quad$ $y = \sin^{-1}(x/n)^{0,5}$

where

y is the transformed datum;

x is the original datum;

n is the number of data points.

Time to failure in stress corrosion cracking is often fitted with a log x transformation (see References [3][4]).

Once a set of transformed data is found that yields an approximately straight line on a probability plot, the statistical procedures of interest can be carried out on the transformed data. It is essential that results, such as predicted data values or confidence intervals, be transformed back using the reverse transformation.

3.6 Unknown distribution

3.6.1 General

If there are insufficient data points or if, for any other reason, the distribution type of the data cannot be determined, then two possibilities exist for analysis.

3.6.1.1 A distribution type may be hypothesized, based on the behaviour of similar types of data. If this distribution is not normal, a transformation may be sought which will normalize that particular distribution. See 3.5 for suggestions. Analysis may then be conducted on the transformed data.

3.6.1.2 Statistical analysis procedures that do not require any specific data distribution type, known as non-parametric methods, may be used to analyse the data. Non-parametric tests do not use the data as efficiently.

3.7 Extreme value analysis

If determining the probability of perforation by a pitting or cracking mechanism, the usual descriptive statistics for the normal distribution are not the most useful. Extreme value statistics should be used instead (see Reference [5]).

3.8 Significant digits

The proper number of significant digits should be used when reporting numerical results.

3.9 Propagation of variance

If a calculated value is a function of several independent variables and those variables have errors associated with them, the error of the calculated value can be estimated by a propagation of variance technique. See References [6][7] for details.

3.10 Mistakes

Mistakes when carrying out an experiment or in the calculations are not a characteristic of the population and can preclude statistical treatment of data or lead to erroneous conclusions if included in the analysis. Sometimes mistakes can be identified by statistical methods by recognizing that the probability of obtaining a particular result is very low. In this way, outlying observations can be identified and dealt with.

4 Central measures

4.1 Average

It is accepted practice to employ several independent (replicate) measurements of any experimental quantity to improve the estimate of precision and to reduce the variance of the average value. If it is assumed that the

processes operating to create error in the measurement are random in nature and are as likely to overestimate the true unknown value as to underestimate it, then the average value is the best estimate of the unknown value in question. The average value is usually indicated by placing a bar over the symbol representing the measured variable and calculated by

$$\bar{x} = \sum \frac{x_i}{n} \tag{3}$$

NOTE In this International Standard, the term “mean” is reserved to describe a central measure of a population, while “average” refers to a sample.

4.2 Median

If processes operate to exaggerate the magnitude of the error, either in overestimating or underestimating the correct measurement, then the median value is usually a better estimate. The median value, x_m, is defined as the value in the middle of all data and can be determined from the m-th ranked data.

$$x_m = \begin{cases} x_{n/2} & \text{for an even number, } n, \text{ of data points} \\ x_{(n+1)/2} & \text{for an odd number, } n, \text{ of data points} \end{cases} \tag{4}$$

4.3 Which to use

If the processes operating to create error affect both the probability and magnitude of the error, then other approaches are required to find the best estimation procedure. A qualified statistician should be consulted in this case.

In corrosion testing, it is generally observed that average values are useful in characterizing corrosion rates. In cases of penetration from pitting and cracking, failure is often defined as the first through-penetration and average penetration rates or times are of little value. Extreme value analysis has been used in these instances.

When the average value is calculated and reported as the only result in experiments where several replicate runs were made, information on the scatter of data is lost.

5 Variability measures

5.1 General

Several measures of distribution variability are available, which can be useful in estimating confidence intervals and making predictions from the observed data. In the case of normal distribution, a number of procedures are available and can be handled by computer programs. These measures include the following: variance, standard deviation, and coefficient of variation. The range is a useful non-parametric estimate of variability and can be used with both normal and other distributions.

5.2 Variance

Variance, σ^2, may be estimated for an experimental data set of n observations by computing the sample estimated variance, S^2, assuming that all observations are subject to the same errors:

$$S^2 = \frac{\sum d^2}{(n-1)} = \frac{\sum (\bar{x} - x_i)^2}{(n-1)} \tag{5}$$

where

d is the difference between the average and the measured value;

$n - 1$ is the number of degrees of freedom available.

Variance is a useful measure because it is additive in systems that can be described by a normal distribution, but the dimensions of variance are the square of units. A procedure known as analysis of variance (ANOVA) has been developed for data sets involving several factors at different levels in order to estimate the effects of these factors.

5.3 Standard deviation

Standard deviation, σ, is defined as the square root of the variance. It has the property of having the same dimensions as the average value and the original measurements from which it was calculated, and is generally used to describe the scatter of the observations.

The standard deviation of an average is different from the standard deviation of a single measured value, but the two standard deviations are related as in the following equation:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} \tag{6}$$

where

n is the total number of measurements which were used to calculate the average value.

When reporting standard deviation calculations, it is important to note clearly whether the value reported is the standard deviation of the average or of a single value. In either case, the number of measurements should also be reported. The sample estimate of the standard deviation is S.

5.4 Coefficient of variation

The population coefficient of variation is defined as the standard deviation divided by the mean. The sample coefficient of variation may be calculated as S/x and is usually reported as a percentage. This measure of variability is particularly useful in cases where the size of the errors is proportional to the magnitude of the measured value, so that the coefficient of variation is approximately constant over a wide range of values.

5.5 Range

The range, w, is defined as the difference between the maximum, x_{max}, and minimum, x_{min}, values in a set of replicate data values. The range is non-parametric in nature, i.e. its calculation makes no assumption about the distribution of error.

$$w = x_{max} - x_{min} \tag{7}$$

In cases when small numbers of replicate values are involved and the data are normally distributed, the range can be used to estimate the standard deviation by the relationship:

$$S = \frac{w}{\sqrt{n}} \tag{8}$$

where

S is the estimated sample standard deviation;

w is the range;

n is the number of observations.

The range has the same dimensions as the standard deviation.

5.6 Precision

5.6.1 General

Precision is the closeness of agreement between randomly selected individual measurements or test results. The standard deviation of the error of measurement may be used as a measure of imprecision.

5.6.1.1 One aspect of precision concerns the ability of one investigator or laboratory to reproduce a measurement previously made at the same location with the same method. This aspect is sometimes called repeatability.

5.6.2.1 Another aspect of precision concerns the ability of different investigators and laboratories to reproduce a measurement. This aspect is sometimes called reproducibility.

5.7 Bias

5.7.1 General

Bias is the closeness of agreement between an observed value and an accepted reference value. When applied to individual observations, bias includes a combination of a random component and a component due to systematic error. Under these circumstances, accuracy contains elements of both precision and bias. Bias refers to the tendency of a measurement technique to consistently underestimate or overestimate. In cases where a specific quantity such as corrosion rate is being estimated, a quantitative bias may be determined.

5.7.1.1 Corrosion test methods which are intended to simulate service conditions, for example natural environments, often produce a different severity of corrosion and relative ranking of performance of materials, as compared to severity and ranking under the conditions which the test is simulating. This is particularly true for test procedures which produce damage rapidly as compared to the service experience. In such cases, it is important to establish the correspondence between results from the service environment and test results for the class of material in question. Bias in this case refers to the variation in the acceleration of corrosion for different materials.

5.7.1.2 Another type of corrosion test method measures a characteristic that is related to the tendency of a material to suffer a form of corrosion damage, for example pitting potential. Bias in this type of test refers to the inability of the test to properly rank the materials to which the test applies as compared to service results.

6 Statistical tests

6.1 Null hypothesis

Null-hypothesis statistical tests are usually carried out by postulating a hypothesis of the form: the distribution of data under test is not significantly different from some postulated distribution. It is necessary to establish a probability that will be acceptable for rejecting the null hypothesis. In experimental work, it is conventional to use probabilities of 0,05 or 0,01 to reject the null hypothesis.

6.1.1 Type I errors occur when the null hypothesis is rejected falsely. The probability of rejecting the null hypothesis falsely is described as the significance level and is often designated as α.

6.1.2 Type II errors occur when the null hypothesis is accepted falsely. If the significance level is set too low, the probability of a Type II error, β, becomes larger. When a value of α is set, the value of β is also set. With a fixed value of β, it is possible to decrease β only by increasing the sample size, assuming that no other factors can be changed to improve the test.

6.2 Degrees of freedom

The number of degrees of freedom of a statistical test refers to the number of independent measurements that are available for the calculation.

6.3 t-Test

The t-statistic may be written in the form:

$$t = \frac{|\bar{x} - \mu|}{S(\bar{x})} \tag{9}$$

where

$\bar{x}$ is the sample average;

μ is the population mean;

$S(\bar{x})$ is the estimated standard deviation of the sample average.

The t-distribution is usually tabulated in terms of significance levels and degrees of freedom.

6.3.1 The t-test may be used to test the null hypothesis:

$$m = \mu$$

For example, the value m is not significantly different from μ, the population mean. The t-test is then:

$$t = \frac{|\bar{x} - m|}{S(x)\sqrt{\frac{1}{n}}} \tag{10}$$

The calculated value of t may be compared to the value of t for the number of degrees of freedom, n, and the significance level.

6.3.2 The t-statistic may be used to obtain a confidence interval for an unknown value, for example a corrosion-rate value calculated from several independent measurements:

$$\left(\bar{x} - tS(\bar{x})\right) < \mu < \left(\bar{x} + tS(\bar{x})\right) \tag{11}$$

where $tS(\bar{x})$ represents the one-half width confidence interval associated with the significance level chosen.

6.3.3 The t-test is often used to test whether there is a significant difference between two sample averages. In this case, the expression becomes:

$$t = \frac{|\overline{x_1} - \overline{x_2}|}{S(x)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{12}$$

where

$\overline{x_1}$ and $\overline{x_2}$ are the sample averages;

n_1 and n_2 are the number of measurements used in calculating x_1 and x_2;

$S(x)$ is the pooled estimate of the standard deviation from both sets of data.

i.e.

$$S(x) = \sqrt{\frac{(n_1 - 1)S^2(x_1) + (n_2 - 1)S^2(x_2)}{n_1 + n_2 - 2}} \tag{13}$$

6.3.4 One-sided t-test. The t-function is symmetrical and can have negative as well as positive values. In the above examples, only absolute values of the differences were discussed. In some cases, a null hypothesis of the form:

$$\mu > m$$

or

$$\mu < m$$

may be desired. This is known as a one-sided t-test and the significance level associated with this t-value is half of that for a two-sided t.

6.4 F-test

The F-test is used to test whether the variance associated with a variable, x_1, is significantly different from a variance associated with a variable x_2. The F-statistic is then:

$$Fx_1x_2 = S^2(x_2) \tag{14}$$

The F-test is an important component in the analysis of variance used in experimental designs. Values of F are tabulated for significance levels and degrees of freedom for both variables. In cases where the data are not normally distributed, the F-test approach may falsely show a significant effect because of the non-normal distribution rather than an actual difference in variances being compared.

6.5 Correlation coefficient

The correlation coefficient, r, is a measure of a linear association between two random variables. Correlation coefficients vary between −1 and +1 and the closer they are to either −1 or +1, the better the correlation. The sign of the correlation coefficient simply indicates whether the correlation is positive (y increases with x) or negative (y decreases as x increases). The correlation coefficient, r, is given by:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\left[\sum(x_i - \bar{x})\sum(y_i - \bar{y})\right]^{1/2}} = \frac{\sum x_i y_i - n\overline{xy}}{\left[\left(\sum {x_i}^2 - n\bar{x}^2\right)\sum({y_i}^2 - n\bar{y}^2)\right]^{1/2}} \tag{15}$$

where

x_i are the observed values of random variable x;

y_i are the observed values of random variable y;

$\bar{x}$ is the average value of x;

$\bar{y}$ is the average value of y;

n is the number of observations.

Generally, r^2 values are preferred because they avoid the problem of signs and relate directly to variance. Values of r or r^2 have been tabulated for different significance levels and degrees of freedom. In general, it is desirable to report values of r or r^2 presenting correlations and regression analyses.

NOTE The procedure for calculating correlation coefficients does not require that the x and y variables be random and, consequently, some investigators have used the correlation coefficient as an indication of the goodness of fit of data in a regression analysis.

However, the significance test using the correlation coefficient requires that the x and y values be independent variables of a population measured on randomly selected samples.

6.6 Sign test

The sign test is a non-parametric test used in sets or in paired data to determine if one component of the pair is consistently larger than the other (see Reference [9]). In this test method, the values of the data pairs are compared, and, if the first entry is larger than the second, a plus sign is recorded. If the second term is larger, then a minus sign is recorded. If both are equal, then no sign is recorded. The total number of plus signs, P, and minus signs, N, is computed. Significance is determined by the following test:

$$|P - N| > k\sqrt{P + N} \tag{16}$$

where

k is a function of significance level as follows:

k	Significance level
1,6	0,10
2,0	0,05
2,6	0,01

The sign test does not depend on the magnitude of the difference and so can be used in cases where normal statistics would be inappropriate or impossible to apply.

6.7 Outside count

The outside count test is a useful non-parametric technique to evaluate whether the magnitude of one of two data sets of approximately the same number of values is significantly larger than the other. The details of the procedure may be found elsewhere (see Reference [9]).

7 Curve fitting — Method of least squares

7.1 Minimizing variance

It is often desirable to determine the best algebraic expression to fit a data set with the assumption that a normally distributed random error is operating. In this case, the best fit will be obtained when the condition of minimum variance between the measured value and the calculated value is obtained for the data set. The procedures used to determine equations of best fit are based on this concept. Software is available for computer calculation of regression equations, including linear, polynomial and multiple-variable regression equations.

7.2 Linear regression — 2 variables

Linear regression is used to fit data to a linear relationship of the following form:

$$y = mx + b \tag{17}$$

In this case, the best fit is given by:

$$m = \left(n\sum xy - \sum x \sum y\right) / \left[n\sum x^2 - (\sum x)^2\right] \tag{18}$$

$$b = \frac{1}{n}\left[\sum x - m\sum y\right] \tag{19}$$

where

y is the dependent variable;

x is the independent variable;

m is the slope of the estimated line;

b is the y intercept of the estimated line;

Σx is the sum of x values, etc.;

n is the number of observations of x and y.

This standard deviation of m and the standard error of the expression are often of interest and can be calculated easily (see References [6][8][10]). One problem with linear regression is that all the errors are assumed to be associated with the dependent variable, y, and this might not be a reasonable assumption. A variation of the linear regression approach is available, assuming that the fitting equation passes through the origin. In this case, only one adjustable parameter will result from the fit. It is possible to use statistical tests, such as the F-test, to compare the goodness of fit between this approach and the two adjustable parameter fits described above.

7.3 Polynomial regression

Polynomial regression analysis is used to fit data to a polynomial equation of the following form:

$$y = a + bx + cx^2 + dx^3 + \ldots \tag{20}$$

where

a, b, c, d are the adjustable constants used to fit the data set;

x is the observed independent variable;

y is the observed dependent variable.

The equations required to carry out the calculation of the best-fit constants are complex and best handled by a computer. It is usually desirable to run a series of expressions and compute the residual variance for each expression to find the simplest expression fitting the data.

7.4 Multiple regression

Multiple-regression analysis is used when data sets involving more than one independent variable are encountered. An expression of the following form is desired in a multiple linear regression:

$$y = a + b_1x_1 + b_2x_2 + b_4x_4 + \ldots \tag{21}$$

where

$a, b_1, b_2, b_4, \ldots$ are the adjustable constants used to obtain the best fit of the data set;

$x_1, x_2, x_4, \ldots$ are the observed independent variables;

y is the observed dependent variable.

Because of the complexity of this problem, it is generally handled with the help of a computer. One strategy is to compute the value of all the "b's," together with standard deviation for each "b". It is usually necessary to

run several regression analyses, dropping variables, to establish the relative importance of the independent variables under consideration.

8 Analysis of variance

8.1 Comparison of effects

Analysis of variance is useful to determine the effect of a number of variables on a measured value when a small number of discrete levels of each independent variable is studied (see References [6], [8], [10], [11] and [12]). This is best handled by using a factorial or similar experimental design to establish the magnitude of the effects associated with each variable and the magnitude of the interactions between the variables.

8.2 The two-level factorial design

8.2.1 The two-level factorial design experiment is an excellent method for determining which variables have an effect on the outcome.

8.2.2 Each time an additional variable is to be studied, twice as many experiments are required to complete the two-level factorial design. When many variables are involved, the number of experiments becomes prohibitive.

8.2.3 Fractional replication can be used to reduce the amount of testing. When this is done, the amount of information that can be obtained from the experiment is also reduced.

9 Extreme value statistics

9.1 Scope of this clause

9.1.1 Extreme-value statistics provide a powerful method for analysing localized corrosion data, and especially for estimating pit depth. The maximum pit depth is more important than the average pit depth because perforation is caused by the deepest pit (see References [5][13][16][18]).

9.1.2 The normal (Gausian), Poisson, binomial, exponential and log–normal distributions are often observed in engineering data. The largest or the smallest values from these original distributions make another group of distributions, called extreme-value distributions. There are three types of extreme value distributions which are asymptotic limiting forms of the original distribution for large samples. Type I for the largest value is called the Gumbel, or doubly exponential, distribution and is often observed for the distribution of the deepest pits. Type III for the smallest value is called the Weibull distribution which is widely used for analysing failure life data in the field of reliability engineering. The procedure for estimating the parameters of the Gumbel distribution presented here can also be applied to the Weibull distribution.

9.1.3 This method allows for the estimation of the parameters of the Gumbel distribution. The maximum pit-depth perforation probability for a large area of given thickness can be estimated from the Gumbel distribution observed for small areas.