Designation: G16 – 13

Standard Guide for Applying Statistics to Analysis of Corrosion Data[1]

This standard is issued under the fixed designation G16; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This guide covers and presents briefly some generally accepted methods of statistical analyses which are useful in the interpretation of corrosion test results.

1.2 This guide does not cover detailed calculations and methods, but rather covers a range of approaches which have found application in corrosion testing.

1.3 Only those statistical methods that have found wide acceptance in corrosion testing have been considered in this guide.

1.4 The values stated in SI units are to be regarded as standard. No other units of measurement are included in this standard.

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E178 Practice for Dealing With Outlying Observations

E691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method

G46 Guide for Examination and Evaluation of Pitting Corrosion

IEEE/ASTM SI 10 American National Standard for Use of the International System of Units (SI): The Modern Metric System

3. Significance and Use

3.1 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.

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

3.3 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.

4. Errors

4.1 *Distributions*—In the measurement of 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 error 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 are seen in corrosion work.

4.2 *Histograms*—A bar graph called a histogram may be used to display the scatter of the data. A histogram is constructed by dividing the range of data values into equal intervals on the abscissa axis 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 should be few enough so that almost all intervals contain at least three points; however, there should be a sufficient number of intervals to facilitate visualization of the shape and symmetry of the bar heights. Twenty intervals are usually recommended for a histogram. Because so many points are required to construct a histogram, it is unusual to find data sets in corrosion work that lend themselves to this type of analysis.

4.3 *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

[1] This guide is under the jurisdiction of ASTM Committee G01 on Corrosion of Metals and is the direct responsibility of Subcommittee G01.05 on Laboratory Corrosion Tests.

Current edition approved Dec. 1, 2013. Published December 2013. Originally approved in 1971. Last previous edition approved in 2010 as G16–95 (2010). DOI: 10.1520/G0016-13.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

should be used to obtain a data set which is approximately normally distributed. Transformed data may be analyzed 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.

4.4 *Normal Probability Paper*—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 described as follows (**1**).[3]

4.4.1 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 center, a number approaching 0 percent at one end, and a number approaching 1.0 or 100 % at the other end. The marks are spaced far apart in the center and close together at the ends. A normal probability plot may be constructed as follows with normal probability paper.

NOTE 1—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.

4.4.1.1 Number the data points starting at the largest negative value and proceeding to the largest positive value. The numbers of the data points thus obtained are called the ranks of the points.

4.4.1.2 Plot each point on the normal probability paper such that when the data are arranged in order: y (1), y (2), y (3), ..., these values are called the order statistics; the linear axis reflects the value of the data, while the probability axis location is calculated by subtracting 0.5 from the number (rank) of that point and dividing by the total number of points in the data set.

NOTE 2—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 the identical values.

4.4.2 If normal probability paper is not available, the location of each point on the probability plot may be determined as follows:

4.4.2.1 Mark the probability axis using linear graduations from 0.0 to 1.0.

4.4.2.2 For each point, subtract 0.5 from the rank and divide the result by the total number of points in the data set. This is the area to the left of that value under the standardized normal distribution. The cumulative distribution function is the number, always between 0 and 1, that is plotted on the probability axis.

4.4.2.3 The value of the data point defines its location on the other axis of the graph.

4.5 *Other Probability Paper*—If the histogram is not symmetrical and bell-shaped, or if the probability plot shows nonlinearity, a transformation may be used to obtain a new, transformed data set that may be normally distributed. Although it is sometimes possible to guess at 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:

$$y = \log x \qquad y = \exp x$$
$$y = \sqrt{x} \qquad y = x^2$$
$$y = 1/x \qquad y = \sin^{-1}\sqrt{x/n}$$

where:

y = transformed datum,
x = original datum, and
n = number of data points.

Time to failure in stress corrosion cracking usually is best fitted with a log x transformation (**2, 3**).

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. Results, such as predicted data values or confidence intervals, must be transformed back using the reverse transformation.

4.6 *Unknown Distribution*—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:

4.6.1 A distribution type may be hypothesized based on the behavior of similar types of data. If this distribution is not normal, a transformation may be sought which will normalize that particular distribution. See 4.5 above for suggestions. Analysis may then be conducted on the transformed data.

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

4.7 *Extreme Value Analysis*—In the case of determining the probability of perforation by a pitting or cracking mechanism, the usual descriptive statistics for the normal distribution are not the most useful. In this case, Guide G46 should be consulted for the procedure (**4**).

4.8 *Significant Digits*—IEEE/ASTM SI 10 should be followed to determine the proper number of significant digits when reporting numerical results.

4.9 *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 Refs (**5**) and (**6**) for details.

4.10 *Mistakes*—Mistakes either in carrying out an experiment or in 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

[3] The boldface numbers in parentheses refer to a list of references at the end of this standard.

mistakes can be identified by statistical methods by recognizing that the probability of obtaining a particular result is very low.

4.11 *Outlying Observations*—See Practice E178 for procedures for dealing with outlying observations.

5. Central Measures

5.1 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.

NOTE 3—In this standard, the term "mean" is reserved to describe a central measure of a population, while average refers to a sample.

5.2 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.

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

5.4 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 in these cases, average penetration rates or times are of little value. Extreme value analysis has been used in these cases, see Guide G46.

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

6. Variability Measures

6.1 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 with 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.

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

$$S^2 = \frac{\sum d^2}{n-1} \quad (1)$$

where:

d = the difference between the average and the measured value,
$n-1$ = the degrees of freedom available.

Variance is a useful measure because it is additive in systems that can be described by a normal distribution; however, the dimensions of variance are 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. (See Section 9.)

6.3 *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.

6.3.1 *Standard Deviation of the Average*—The standard deviation of an average, $S\bar{x}$, is different from the standard deviation of a single measured value, but the two standard deviations are related as in (Eq 2):

$$S\bar{x} = \frac{S}{\sqrt{n}} \quad (2)$$

where:

n = 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.

6.4 *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 $S/\bar{x}$ and is usually reported in percent. 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.

6.5 *Range*—The range is defined as the difference between the maximum and minimum values in a set of replicate data values. The range is non-parametric in nature, that is, its calculation makes no assumption about the distribution of error. In cases when small numbers of replicate values are involved and the data are normally distributed, the range, w, can be used to estimate the standard deviation by the relationship:

$$S \simeq \frac{w}{\sqrt{n}}, \; n < 12 \quad (3)$$

where:

S = the estimated sample standard deviation,
w = the range, and
n = the number of observations.

The range has the same dimensions as standard deviation. A tabulation of the relationship between σ and w is given in Ref (7).

6.6 *Precision*—Precision is 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.

6.6.1 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.

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

6.7 *Bias*—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 under- or overestimate. In cases where a specific quantity such as corrosion rate is being estimated, a quantitative bias may be determined.

6.7.1 Corrosion test methods which are intended to simulate service conditions, for example, natural environments, often are more severe on some materials than others, as compared to 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.

6.7.2 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. Ranking may also be used as a qualitative estimate of bias in the test method types described in 6.7.1.

7. Statistical Tests

7.1 *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.

7.1.1 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.

7.1.2 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 α, it is possible to decrease β only by increasing the sample size assuming no other factors can be changed to improve the test.

7.2 *Degrees of Freedom*—The degrees of freedom of a statistical test refer to the number of independent measurements that are available for the calculation.

7.3 *t Test*—The *t* statistic may be written in the form:

$$t = \frac{|\bar{x} - \mu|}{S(\bar{x})} \tag{4}$$

where:

$\bar{x}$ = the sample average,
μ = the population mean, and
$S(\bar{x})$ = estimated standard deviation of the sample average.

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

7.3.1 The *t* test may be used to test the null hypothesis:

$$m = \mu \tag{5}$$

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

$$t = \frac{|\bar{x} - m|}{S(x)\sqrt{\frac{1}{n}}} \tag{6}$$

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

7.3.2 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:

$$(\bar{x} - t\,S(\bar{x})) < \mu < (\bar{x} + t\,S(\bar{x})) \tag{7}$$

where:

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

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

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{S(x)\sqrt{1/n_1 + 1/n_2}} \tag{8}$$

where:

$\bar{x}_1$ *and* $\bar{x}_2$ = sample averages,
n_1 *and* n_2 = number of measurements used in calculating $\bar{x}_1$ and $\bar{x}_2$ respectively, and
$S(x)$ = pooled estimate of the standard deviation from both sets of data.

i.e.:

$$S(x) = \sqrt{\frac{(n_1 - 1)S^2(x_1) + (n_2 - 1)S^2(x_2)}{n_1 + n_2 - 2}} \tag{9}$$

7.3.4 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:

$$\mu > m \tag{10}$$

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.

7.4 *F Test*—Labeling the variable with the larger observed variance as x_1, the F statistic is used to test whether the variance associated with that variable is significantly larger than the variable associated with variable x_2. The F statistic is then:

$$Fx_1,\ x_2 = \frac{S^2(x_1)}{S^2(x_2)} \tag{11}$$

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.

7.5 *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 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:

$$r = \frac{\left[\sum (x_i - \bar{x})(y_i - \bar{y})\right]}{\left[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2\right]^{\frac{1}{2}}} = \frac{\left(\sum x_i\, y_i\right) - n\bar{x}\bar{y}}{\left\{\left[\sum (x_i^2) - n\,\bar{x}^2\right]\left[\sum (y_i^2) - n\,\bar{y}^2\right]\right\}^{\frac{1}{2}}} \tag{12}$$

where:

x_i = observed values of random variable x,
y_i = observed values of random variable y,
$\bar{x}$ = average value of x,
$\bar{y}$ = average value of y, and
n = number of observations.

Generally, r^2 values are preferred because they avoid the problem of sign and the r^2 values 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 when presenting correlations and regression analyses.

NOTE 4—The procedure for calculating correlation coefficient does not require that the x and y variables be random and consequently, some investigators have used the correlation coefficient as an indication of goodness of fit of data in a regression analysis. However, the significance test using correlation coefficient requires that the x and y values be independent variables of a population measured on randomly selected samples.

7.6 *Sign Test*—The sign test is a non-parametric test used in sets of paired data to determine if one component of the pair is consistently larger than the other (**8**). 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:

$$|P - N| > k\sqrt{P+N} \tag{13}$$

where k = 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.

7.7 *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 (**8**).

7.8 *Corner Count*—The corner count test is a non-parametric graphical technique for determining whether there is correlation between two variables. It is simpler to apply that the correlation coefficient, but requires a graphical presentation of the data. The detailed procedure may be found elsewhere (**8**).

8. Curve Fitting—Method of Least Squares

8.1 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.

8.2 *Linear Regression—2 Variables*—Linear regression is used to fit data to a linear relationship of the following form:

$$y = mx + b \tag{14}$$

In this case, the best fit is given by:

$$m = \left(n\sum xy - \sum x \sum y\right)/\left[n\sum x^2 - \left(\sum x\right)^2\right] \tag{15}$$

$$b = \frac{1}{n}\left[\sum x - m\sum y\right] \tag{16}$$

where:

y = the dependent variable
x = the independent variable,
m = the slope of the estimated line,
b = the y intercept of the estimated line,
$\sum x$ = the sum of x values and so forth, and
n = 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 may be calculated easily (**5, 7, 9**). One problem with linear regression is that all the errors are assumed to be associated with the dependent variable, y, and this may not be a reasonable assumption. A variation of the

linear regression approach is available, assuming 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.

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

$$y = a + bx + cx^2 + dx^3 \text{ and so forth} \quad (17)$$

where:

a, b, c, d = adjustable constants to be used to fit the data set,
x = the observed independent variable, and
y = 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.

8.4 *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:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 \text{ and so forth} \quad (18)$$

where:

a, b_1, b_2, b_3, *and so forth* = adjustable constants used to obtain the best fit of the data set
x_1, x_2, x_3, and so forth = the observed independent variables
y = 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 analysis, dropping variables, to establish the relative importance of the independent variables under consideration.

9. Comparison of Effects—Analysis of Variance

9.1 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 (**5**, **7**, **9**, **10**, **11**). 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.

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

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

9.2.2 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.

9.3 In the design and analysis of interlaboratory test programs, Practice E691 should be consulted.

10. Keywords

10.1 analysis of variance; corrosion data; curve fitting; statistical analysis; statistical tests

APPENDIX

(Nonmandatory Information)

X1. SAMPLE CALCULATIONS

X1.1 Calculation of Variance and Standard Deviation

X1.1.1 *Data*—The 27 values shown in Table X1.1 are calculated mass loss based corrosion rates for copper panels in a one year rural atmospheric exposure.

X1.1.2 *Calculation of Statistics:*

X1.1.2.1 Let x_i = corrosion rate of the i^{th} panel. The average corrosion rate of 27 panels, $\bar{x}$:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{54.43}{27} = 2.016 \quad (X1.1)$$

The variance estimate based on this sample, $s^2(x)$:

$$s^2(x) = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} = \quad (X1.2)$$

$$\frac{110.085 - 27 \times (2.016)^2}{26} = \frac{0.350}{26} = 0.0135$$

The standard deviation is:

$$s(x) = (0.0135)^{1/2} = 0.116 \quad (X1.3)$$

The coefficient of variation is:

$$\frac{0.116}{2.016} \times 100 = 5.75\,\% \quad (X1.4)$$

The standard deviation of the average is:

$$s(\bar{x}) = \frac{0.116}{(27)^{\frac{1}{2}}} = 0.0223 \quad (X1.5)$$

The range, *w*, is the difference between the largest and smallest values:

$$w = 2.21 - 1.70 = 0.41 \quad (X1.6)$$

The mid-range value is: