Standard Guide for Application of Basic Statistical Methods to Weathering Tests[1]

This standard is issued under the fixed designation G169; 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 elementary statistical methods for the analysis of data common to weathering experiments. The methods are for decision making, in which the experiments are designed to test a hypothesis on a single response variable. The methods work for either natural or laboratory weathering.

1.2 Only basic statistical methods are presented. There are many additional methods which may or may not be applicable to weathering tests that are not covered in this guide.

1.3 This guide is not intended to be a manual on statistics, and therefore some general knowledge of basic and intermediate statistics is necessary. The text books referenced at the end of this guide are useful for basic training.

1.4 This guide does not provide a rigorous treatment of the material. It is intended to be a reference tool for the application of practical statistical methods to real-world problems that arise in the field of durability and weathering. The focus is on the interpretation of results. Many books have been written on introductory statistical concepts and statistical formulas and tables. The reader is referred to these for more detailed information. Examples of the various methods are included. The examples show typical weathering data for illustrative purposes, and are not intended to be representative of specific materials or exposures.

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E41 Terminology Relating To Conditioning

G113 Terminology Relating to Natural and Artificial Weathering Tests of Nonmetallic Materials

G141 Guide for Addressing Variability in Exposure Testing of Nonmetallic Materials

2.2 *ISO Documents:*

ISO 3534/1 Vocabulary and Symbols – Part 1: Probability and General Statistical Terms[3]

ISO 3534/3 Vocabulary and Symbols – Part 3: Design of Experiments[3]

3. Terminology

3.1 *Definitions*—See Terminology G113 for terms relating to weathering, Terminology E41 for terms relating to conditioning and handling, ISO 3534/1 for terminology relating to statistics, and ISO 3534/3 for terms relating to design of experiments.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *arithmetic mean; average*—the sum of values divided by the number of values. **ISO 3534/1**

3.2.2 *blocking variable*—a variable that is not under the control of the experimenter, (for example, temperature and precipitation in exterior exposure), and is dealt with by exposing all samples to the same effects

3.2.2.1 *Discussion*—The term "block" originated in agricultural experiments in which a field was divided into sections or blocks having common conditions such as wind, proximity to underground water, or thickness of the cultivatable layer. **ISO 3534/3**

3.2.3 *correlation*—in weathering, the relative agreement of results from one test method to another, or of one test specimen to another.

3.2.4 *median*—the midpoint of ranked sample values. In samples with an odd number of data, this is simply the middle value, otherwise it is the arithmetic average of the two middle values.

3.2.5 *nonparametric method*—a statistical method that does not require a known or assumed sample distribution in order to support or reject a hypothesis.

3.2.6 *normalization*—a mathematical transformation made to data to create a common baseline.

[1] This guide is under the jurisdiction of ASTM Committee G03 on Weathering and Durability and is the direct responsibility of Subcommittee G03.93 on Statistics.

Current edition approved June 1, 2013. Published June 2013. Originally approved in 2001. Last previous edition approved in 2008 as G169 – 01 (2008)$^{\varepsilon 1}$. DOI: 10.1520/G0169-01R13.

[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.

[3] Available from American National Standards Institute, 11 W. 42nd St., 13th Floor, New York, NY 10036.

3.2.7 *predictor variable (independent variable)*— a variable contributing to change in a response variable, and essentially under the control of the experimenter. **ISO 3534/3**

3.2.8 *probability distribution (of a random variable)*—a function giving the probability that a random variable takes any given value or belongs to a given set of values. **ISO 3534/1**

3.2.9 *random variable*—a variable that may take any of the values of a specified set of values and with which is associated a probability distribution.

3.2.9.1 *Discussion*—A random variable that may take only isolated values is said to be "discrete." A random variable which may take any value within a finite or infinite interval is said to be "continuous." **ISO 3534/1**

3.2.10 *replicates*—test specimens with nominally identical composition, form, and structure.

3.2.11 *response variable (dependent variable)*— a random variable whose value depends on other variables (factors). Response variables within the context of this guide are usually property measurements (for example, tensile strength, gloss, color, and so forth). **ISO 3534/3**

4. Significance and Use

4.1 The correct use of statistics as part of a weathering program can greatly increase the usefulness of results. A basic understanding of statistics is required for the study of weathering performance data. Proper experimental design and statistical analysis strongly enhances decision-making ability. In weathering, there are many uncertainties brought about by exposure variability, method precision and bias, measurement error, and material variability. Statistical analysis is used to help decide which products are better, which test methods are most appropriate to gauge end use performance, and how reliable the results are.

4.2 Results from weathering exposures can show differences between products or between repeated testing. These results may show differences which are not statistically significant. The correct use of statistics on weathering data can increase the probability that valid conclusions are derived.

5. Test Program Development

5.1 *Hypothesis Formulation:*

5.1.1 All of the statistical methods in this guide are designed to test hypotheses. In order to apply the statistics, it is necessary to formulate a hypothesis. Generally, the testing is designed to compare things, with the customary comparison being:

Do the predictor variables significantly affect the response variable?

Taking this comparison into consideration, it is possible to formulate a default hypothesis that the predictor variables do not have a significant effect on the response variable. This default hypothesis is usually called H_o, or the Null Hypothesis.

5.1.2 The objective of the experimental design and statistical analysis is to test this hypothesis within a desired level of significance, usually an alpha level (α). The alpha level is the probability below which we reject the null hypothesis. It can be thought of as the probability of rejecting the null hypothesis when it is really true (that is, the chance of making such an error). Thus, a very small alpha level reduces the chance in making this kind of an error in judgment. Typical alpha levels are 5 % (0.05) and 1 % (0.01). The *x*-axis value on a plot of the distribution corresponding to the chosen alpha level is generally called the critical value (cv).

5.1.3 The probability that a random variable *X* is greater than the critical value for a given distribution is written *P(X>cv)*. This probability is often called the "*p*-value." In this notation, the null hypothesis can be rejected if

$$P(X>cv) < \alpha$$

5.2 *Experimental Design*—The next step in setting up a weathering test is to design the weathering experiment. The experimental design will depend on the type and number of predictor variables, and the expected variability in the sample population, exposure conditions, and measurements. The experimental design will determine the amount of replication, specimen positioning, and appropriate statistical methods for analyzing the data.

5.2.1 *Response Variable*—The methods covered in this guide work for a single response variable. In weathering and durability testing, the response variable will usually be a quantitative property measurement such as gloss, color, tensile strength, modulus, and others. Sometimes, qualitative data such as a visual rating make up the response variable, in which case nonparametric statistical methods may be more appropriate.

5.2.1.1 If the response variable is "time to failure," or a counting process such as "the number of failures over a time interval," then reliability-based methods should be used.

5.2.1.2 Here are the key considerations regarding the response variable:

(1) **What is the response variable?**

(2) **Will the data represent quantitative or qualitative measurements?**

Qualitative data may be best analyzed with a nonparametric method.

(3) **What is the expected variability in the measurement?**

When there is a high amount of measurement variability, then more replication of test specimens is needed.

(4) **What is the expected variability in the sample population?**

More variability means more replication.

(5) **Is the comparison relative (ranked) or a direct comparison of sample statistics (for example, means)?**

Ranked data is best handled with nonparametric methods.

5.2.1.3 It is important to recognize that variability in exposure conditions will induce variability in the response variable. Variability in both outdoor and laboratory exposures has been well-documented (for example, see Guide G141). Excessive variability in exposure conditions will necessitate more replication. See 5.2.2 for additional information.

5.2.2 *Predictor Variables*—The objective of most of the methods in this guide is to determine whether or not the predictor variables had a significant effect on the response variable. The variables will be a mixture of the things that are controllable (predictor variables – the items of interest), things that are uncontrolled (blocking variables), or even worse, things that are not anticipated.

5.2.2.1 The most common variables in weather and durability testing are the applied environmental stresses. These can be controlled, for example, temperature, irradiance, humidity level in a laboratory device, or uncontrolled, that is, an arbitrary outdoor exposure.

NOTE 1—Even controlled environmental factors typically exhibit variability, which must be accounted for (see Guide G141). The controlled variables are the essence of the weathering experiment. They can take on discrete or continuous values.

5.2.2.2 Some examples of discrete predictor variables are:

Polymer	A, B, C
Ingredient	A, B, C, D
Exposure location	A versus B (for example, Ohio to Florida, or, Laboratory 1, Laboratory 2, and Laboratory 3)

5.2.2.3 Some examples of continuous predictor variables are:

Ingredient level (for example, 0.1 %, 0.2 %, 0.4 %, 0.8 %)
Exposure temperature (for example, 40, 50, 60, 70°C)
Processing stress level (for example, temperature)

5.2.2.4 It is also possible to have predictor variables of each type within one experiment. One key consideration for each predictor variable is: Is it continuous or discrete? In addition, there are other important features to be considered:

(1) If discrete, how many possible states can it take on?

(2) If continuous, how much variability is expected in the values? If the variability is high, the number of replicates should be increased.

5.2.2.5 The exposure stresses are extremely important factors in any weathering test. If the exposure stresses are expected to be variable across the exposure area, then one of two approaches to experimental design should be taken:

(1) Reposition the test specimens over the course of the exposure to reduce this variability. This will reduce the amount of replication required in the design.

(2) Consider a block design, where the specimen positions are randomized. A block design will help make sure that variability in exposure stresses are portioned out over the sample population evenly. Position may also then be treated as a predictor variable.

5.2.3 *Experimental Matrix*—It is traditional to summarize the response and predictor variables in a matrix format. Each column represents a variable, and each row represents the result for the combination of predictor variables across the row. In a full factorial design, every possible combination of all of the levels for each predictor variable is tested (the rows of the matrix). In addition, each combination may be tested more than once (replication).

5.2.3.1 Table 1 illustrates an experiment with two factors, one with three possible states (Predictor Variable 2), the other with two (Predictor Variable 1), and two replicates per combination.

5.2.3.2 In general, it is not necessary to have identical numbers of replicates for each factor combination, nor is it always necessary to test every combination. A good rule of thumb is to test all combinations of levels that are expected to

TABLE 1 EXAMPLE EXPERIMENT

Response Variable	Predictor Variance 1	Predictor Variance 2
$^{x}AA_1$	A	A
$^{x}AA_2$	A	A
$^{x}AB_1$	A	B
$^{x}AB_2$	A	B
$^{x}AC_1$	A	C
$^{x}AC_2$	A	C
$^{x}BA_1$	B	A
$^{x}BA_2$	B	A
$^{x}BB_1$	B	B
$^{x}BB_2$	B	B
$^{x}BC_1$	B	C
$^{x}BC_2$	B	C

be important, and a few of the combinations at the more extreme levels for some of the factors. A detailed treatment of experimental designs other than the full factorial approach involves a model for the response variable behavior and is beyond the scope of this guide.

5.2.4 *Selecting a Statistical Method*—The final step in setting up the weathering experiment is to select an appropriate method to analyze the results. Fig. 1 uses information from the previous steps to choose some applicable methods:

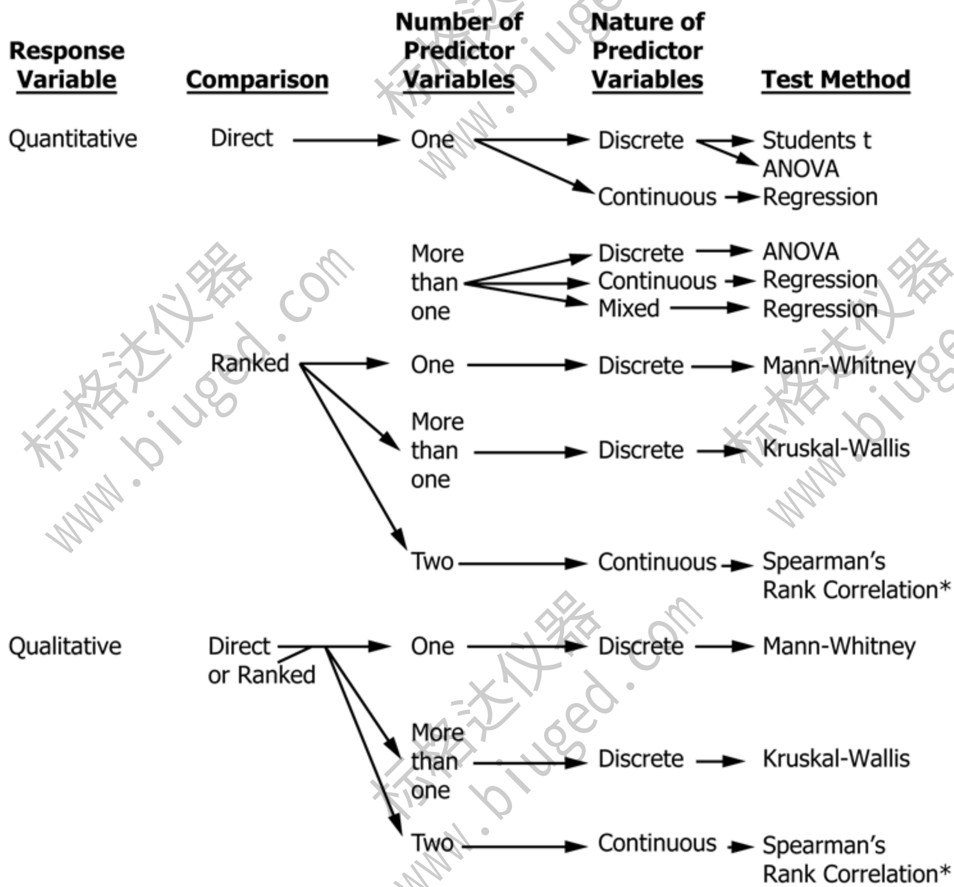
5.3 *Other Issues:*

5.3.1 *Determining the Frequency of Measurements*—In general, the faster the materials degrade when exposed, the more frequent the evaluations should be. If something is known about the durability of a material in advance of a test, that information should be used to plan the test frequency. If very little is known about the material's durability, it may be helpful to adopt a variable length approach in which frequent inspections are scheduled early on, with fewer later (according to the observed rate of change in the material).

5.3.1.1 If the materials under investigation exhibit sudden failures, or if the failure mechanisms are not detectable until a certain threshold is reached, it may be necessary to continue frequent inspections until failure. In this case, the frequent evaluations might be cursory, for example a visual inspection, rather than a full-blown analytical measurement. Another option, if available, is to automate detection of failure, allowing continuous inspection.

5.3.2 *Determining the Evaluation Timing and Duration of Testing*—If the service life of a product is of interest, it is usually necessary to test until at least some of the sample has failed. Failure is typically a predetermined level of property change, or the point at which the material can no longer perform its intended function. It is recommended that materials be tested until they fail, or at least until they exhibit significant change. When comparing the relative performance of two or more materials, it is recommended that testing continue until a statistically significant spread is observed in their performance. The more rapidly (across a time interval) a material changes in a response variable, the shorter the interval between observations must be to detect changes.

5.3.2.1 Sudden changes in a response variable at any time over the course of an exposure increase the uncertainty of the relationship between the predictor and response variables. In these cases, it is often a good idea to conduct multiple exposures (over time) and exposures in different environments.



*Spearman's rank correlation (q.v.) essentially works on two factors, one of which can only contain two levels (to be correlated).

FIG. 1 Selecting a Method

6. Statistical Methods

6.1 Use the step-by-step approach in Section 5 to arrive at one of the statistical methods. More than one method may apply to a particular experiment, in which case it does not hurt to try several approaches. A brief description of each method follows, along with a small example application.

6.2 *Student's t-Test:*

6.2.1 The Student's *t*-Test can be used to compare the means of two independent samples (random variables). This is the simplest comparison that can be made: there is only one factor with two possible states (by default discrete). Since it is such a direct and limited comparison, replication must be used, typically with at least three replicates in each sample. See Table 2.

6.2.2 The *t*-Test assumes that the data are close to normally distributed, although the test is fairly robust. The distributions of each sample need not be equal, however. For large sample sizes, the *t*-Distribution approaches the normal distribution. If you have reason to suspect that the data are not normally distributed, an alternate method like Mann-Whitney may be more appropriate.

6.2.3 Often, physical property measurements are close to normally distributed. The following is an example problem and analysis. The analysis was calculated two ways: assuming that the populations had equal variance, and not making such an assumption. In either case, the resulting probability values indicate that there is a significant difference in the sample means (assuming an alpha level of 0.05).

TABLE 2 STUDENT'S *t*-TEST EXAMPLE

Color Change	Formula
1.000	A
1.200	A
1.100	A
0.900	A
1.100	A
1.300	B
1.400	B
1.200	B

Predictor samples *t*-test on COLOR CHANGE grouped by FORMULA:

Formula	N	Mean	Standard Deviation
A	5	1.060	0.114
B	3	1.300	0.100

Analysis Method	*t* Value	Degrees of Freedom	P(X>cv)
Separate variances	3.116	4.9	0.036
Pooled variances	3.000	6.0	0.024

P(X>cv) indicates the probability that a Student's *t*-distributed random variable is greater than the cv, that is, the area under the tail of the *t*-distribution to the right of Point *t*. Since this value in either case is below a pre-chosen alpha level of 0.05, the result is significant. Note that this result would not be significant at an alpha level of 0.01.

6.3 *ANOVA:*

6.3.1 Analysis of Variance (ANOVA) performs comparisons like the *t*-Test, but for an arbitrary number of predictor variables, each of which can have an arbitrary number of levels. Furthermore, each predictor variable combination can have any number of replicates. Like all the methods in this guide, ANOVA works on a single response variable. The predictor variables must be discrete. See Table 3.

6.3.2 The ANOVA can be thought of in a practical sense as an extension of the *t*-Test to an arbitrary number of factors and levels. It can also be thought of as a linear regression model whose predictor variables are restricted to a discrete set. Here is the example cited in the *t*-Test, extended to include an additional formula, and another factor. The new factor is to test whether the resulting formulation is affected by the technician who prepared it. There are two technicians and three formulas under consideration.

6.3.3 This example also illustrates that one need not have identical numbers of replicates for each sample. In this example, there are two replicates per factor combination for Formula A, but no replication appears for the other formulas.

Analysis of Variance
Response variable: COLOR CHANGE

Source	Sum of Squares	Degrees of Freedom	Mean square	*F* Ratio	P(X>cv)
Formula	0.483	2	0.241	16.096	0.025
Technician	0.005	1	0.005	0.333	0.604
Error	0.045	3	0.015	-	-

6.3.4 Assuming an alpha level of 0.05, the analysis indicates that the formula resulted in a significant difference in color change means, but the technician did not. This is evident from the probability values in the final column. Values below the alpha level allow rejection of the null hypothesis.

6.4 *Linear Regression:*

6.4.1 Linear regression is essentially an ANOVA in which the factors can take on continuous values. Since discrete factors can be set up as belonging to a subset of some larger continuous set, linear regression is a more general method. It is in fact the most general method considered in this guide. See Table 4.

6.4.2 The most elementary form of linear regression is easy to visualize. It is the case in which we have one predictor variable and one response variable. The easy way to think of the predictor variable is as an *x*-axis value of a two dimensional plot. For each predictor variable level, we can plot the corresponding measurement (response variable) as a value on the ordinate axis. The idea is to see how well we can fit a line to the points on the plot. See Table 5.

6.4.3 For example, the following experiment looks at the effect of an impact modifying ingredient level on impact strength after one year of outdoor weathering in Arizona.

6.4.4 The plot of ingredient level versus retained impact strength shown with a linear fit and 95 % confidence bands looks like: (See Fig. 2)

6.4.5 This example illustrates the use of replicates at one of the levels. It is a good idea to test replicates at the levels that are thought to be important or desirable. The analysis indicates a good linear fit. We see this from the R^2 value (squared multiple R) of 0.976. The R^2 value is the fraction of the variability of the response variable explained by the regression model, indicates the degree of fit to the model.

6.4.6 The analysis of variance indicates a significant relationship between modifier level and retained impact strength in this test (the probability level is well below an alpha level of 5 %).

Linear Regression Analysis
Response Variable: Impact Retention (%)
Number of Observations: 7
Multiple R: 0.988
Squared Multiple R: 0.976

Source	Degrees of Freedom	Sum of Squares	*F* Ratio	P(X>cv)
Regression	1	0.0464	205.1	less than 0.0001
Residual	5	0.0011	-	-

6.4.7 Regression can be easily generalized to more than one factor, although the data gets difficult to visualize since each factor adds an axis to the plot (it is not so easy to view multidimensional data sets). It can also be adapted to nonlinear models. A common technique for achieving this is to transform data so that it is linear. Another way is to use nonlinear least squares methods, which are beyond the scope of this guide. Regression can also be extended to cover mixed continuous

TABLE 3 ANOVA EXAMPLE

Color Change	Formula	Technician
1.000	A	Elmo
1.100	A	Elmo
1.100	A	Homer
0.900	A	Homer
1.300	B	Elmo
1.400	B	Judd
1.200	B	Homer
0.700	C	Elmo
0.600	C	Homer

TABLE 4 REGRESSION EXAMPLE

Modifier Level	Impact Retention After Exposure
0.005	0.535
0.01	0.6
0.02	0.635
0.02	0.62
0.03	0.68
0.04	0.754
0.05	0.79

TABLE 5 PATHOLOGICAL LINEAR REGRESSION EXAMPLE

x	v
0.01	0.029979
0.02	0.054338
0.03	0.088581
0.04	0.082415
0.05	0.126631
0.06	0.073464
0.07	0.123222
0.08	0.097003
0.09	0.099728
0.75	0.805909
0.86	0.865667

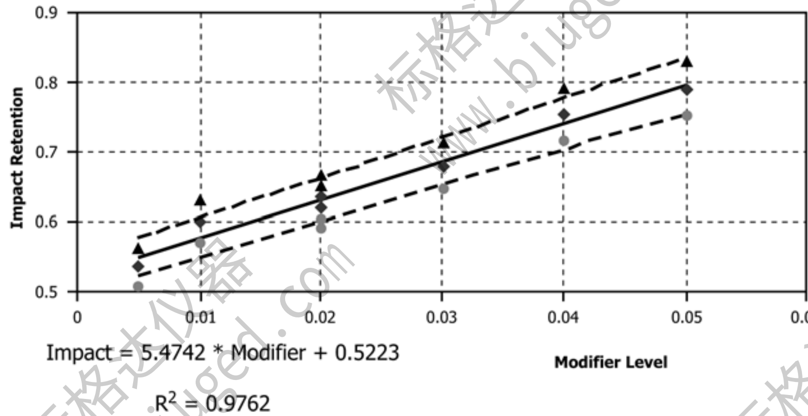


FIG. 2 Linear Regression Fit

and discrete factors. It should be noted that most spreadsheet and elementary data analysis applications can perform fairly sophisticated regression analysis.

6.4.8 Another use of regression is to compare two predictor random variables at a number of levels for each. For example, results from one exposure test can be plotted against the results from another exposure. If the points fall on a line, then one could conclude that the tests are "in agreement." This is called correlation. The usual statistic in a linear correlation analysis is R^2, which is a measure of deviation from the model (a straight line). The R^2 values near one indicate good agreement with the model, while those near zero indicate poor agreement. This type of analysis is different from the approaches suggested above which were constructed to test whether one random variable depended somehow on others. It should be noted, however, that correlation can always be phrased in ANOVA-like terms. The correlation example included for the Spearman rank correlation method illustrates this. The observations then make up a response random variable. Correlation on absolute results is not recommended in weathering testing. Instead, relative data (ranked data) often provide more meaningful analysis (see Spearman's rank correlation).

6.4.9 Regression/correlation can lead to misleadingly high R^2 values when the *x*-axis values are not well-spaced. Consider the following example, which contains a cluster of data that does not exhibit a good linear fit, along with a few outliers. Due to the large spread in the *x*-axis values, the clustered data appears almost as a single data point, resulting in a high R^2 value. (See Fig. 3).

Linear Regression Analysis
Number of Observations: 11
Multiple R: 0.997
Squared Multiple R: 0.994

Source	Degrees of Freedom	Sum of Squares	*F* Ratio	P(X>cv)
Regression	1	0.9235	1509	less than 0.0001
Residual	9	0.0055	-	-

6.4.10 Even though the analysis indicates a good fit to a linear model, the cluster of data does not fit a linear model well at all without the outliers. If the objective of this analysis were correlation, a ranked method like Spearman's (see 6.7) would provide a more reliable analysis.

6.5 *Mann-Whitney:*

6.5.1 The Mann-Whitney test is the nonparametric analog to the Student's *t*-Test. It is used to test for difference in two populations. This test is also known as the Rank-Sum test, the U-test, and the Wilcoxon Test. This test works by ranking the combined data from each population. It is important to look for repeats of the data (these are known as "ties") Ties are treated as follows: the rank is equivalent to the sum of the ranking values normally assigned for that value of the response variable divided by the number of repeats for that value of the response variable. (See the following example.) The ranks are then

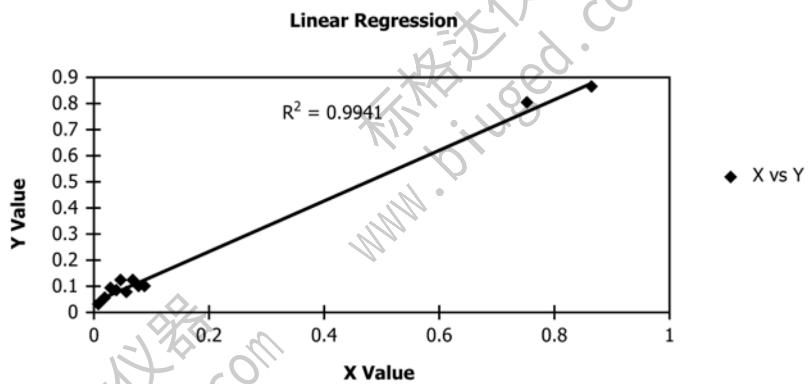


FIG. 3 Pathological Linear Regression Example

summed for one of the groups. This rank sum is normally distributed for a sufficient number of observations, with the following mean and standard deviation:

$$mean = \frac{n_A\left(n_A + n_B + 1\right)}{2}$$

$$standard\ deviation\ (SD) =$$

$$\sqrt{\frac{n_A n_B (n_A + n_B + 1)}{12} - \frac{n_A n_B \sum_{i=1} \left(t_i^3 - t_i\right)}{12(n_A + n_B)(n_A + n_B - 1)}}$$

where

n_A = number of data points in Sample A, and
n_B = number of specimens in Sample B.

If there are no ties in the data (see 6.5.1), the formula for *standard deviation* can be considerably simplified, because the second term under the radical (beginning with the minus sign) evaluates to zero.

6.5.2 The rank sum can be standardized by means of the transformation:

$$(rank\ sum - mean)/SD$$

This value can be compared with a table of *z*-values for the normal distribution to test for significance. (For small numbers of data points, the Student's *t*-distribution is more appropriate.) For example, consider the same data set that appears in the Student's *t*-Test section. Table 6 indicates a significant difference in sample means, since the standardized value is below the value of a normally distributed random variable at an alpha level of 0.05. This is the same conclusion as the *t*-Test.

Mann-Whitney Analysis:

$$mean = \frac{(5)(5+3+1)}{2} = 22.5$$

$$standard\ deviation\ (SD) =$$

$$\sqrt{\frac{(5)(3)(5+3+1)}{12} - \frac{(5)(3)\left((2^3-2)+(2^3-2)\right)}{(12)(5+3)(5+3-1)}} = 3.3139$$

Total Number of Observations: 8
Rank sum for Formula A = 1 + 2 + 3.5 + 3.5 + 5.5 = 15.5
Rank sum - mean = 15.5 – 22.5 = –7.0
Standardized value = –7.0/3.3139 = –2.11
Compare with an alpha level of 0.05 for a normal random variable, –1.96 to 1.96

6.6 *Kruskal-Wallis:*

6.6.1 The Kruskal-Wallis method is a nonparametric analog of single-factor ANOVA. This method compares the medians of three or more groups of samples. To carry out the Kruskal-Wallis method, the data are ranked just as in the Mann-Whitney Method.

TABLE 6 MANN-WHITNEY EXAMPLE

Color Change	Formula	Rank Order	
		Normal	Correlation for Ties
0.9	A	1	1
1	A	2	2
1.1	A	3	3.5
1.1	A	4	3.5
1.2	A	5	5.5
1.2	B	6	5.5
1.3	B	7	7
1.4	B	8	8

6.6.2 Unlike Mann-Whitney, the sampling distribution is arranged so that it follows the chi-square distribution, in which:

$$chi-square = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

And, if there are ties, the following correction must be applied:

$$chi-square(corrected) = \frac{chi-square}{1 - \frac{\sum t(t^2-1)}{N(N^2-1)}}$$

where:

N = total number of observations,
k = number of groups,
n_I = sample size of the *i*th group,
R_i = rank sum of *i*th group, and
t = count of a particular tie.

6.6.3 This statistic is compared against the chi-square distribution with $k - 1$ degrees of freedom (see Table X2.1 if needed), and if it exceeds the value corresponding to the alpha level, the null hypothesis is rejected, which means that the median of the response variable of one or more of the sample sets is different from the others. See Table 7

Kruskal-Wallis Analysis:

TABLE 7 KRUSKAL-WALLIS EXAMPLE

Formula	Gloss	Rank Order	
		Normal	Correlation for Ties
A	6	1	1
A	8	3	3
A	10	5	5
A	11	6	7
A	12	9	9
A	14	11	11.5
A	14	12	11.5
A	15	13	13
A	18	16	16
A	20	19	19
A	21	20	20
A	24	23	23
B	7	2	2
B	9	4	4
B	11	7	7
B	11	8	7
B	16	14	14
B	17	15	15
B	19	17	17.5
B	22	21	21
B	23	22	22
B	26	25	26.5
B	27	29	29.5
B	31	34	34.5
C	13	10	10
C	19	18	17.5
C	25	24	24
C	26	26	26.5
C	26	27	26.5
C	26	28	26.5
C	27	30	29.5
C	28	31	31
C	29	32	32
C	30	33	33
C	31	35	34.5
C	32	36	36

$$chi-square = \left(\frac{12}{(36)(36+1)}\right) * \left(\frac{139^2}{12} + \frac{200^2}{12} + \frac{327^2}{12}\right) - 3(36+1)$$
$$= 13.813$$

Since there are ties, the corrected chi-square must be calculated:

$$chi-square(corr.) = \frac{13.813}{1 - \frac{(3)(3^2-1)+(2)(2^2-1)+(2)(2^2-1)+(4)(4^2-1)+(2)(2^2-1)+(2)(2^2-1)}{(36)(37^2-1)}} = 13.84$$

Degrees of freedom = 3-1 = 2
From chi-square table – at an alpha level of 0.05 and 2 degrees of freedom – cv = 5.99
Since 13.84 > 5.99, the null hypothesis is rejected.

6.7 *Spearman's Rank Correlation:*

6.7.1 Spearman rank correlation is a nonparametric analog of correlation analysis as stated in 6.4 on linear regression. Like regression, it can be applied to compare two predictor random variables, each at several levels (which may be discrete or continuous). Unlike regression, Spearman's rank correlation works on ranked (relative) data, rather than directly on the data itself.

6.7.2 Like the R^2 value produced by regression, the Spearman's r_s coefficient indicates agreement. A value of r_s near one indicates good agreement; a value near zero, poor agreement. Of course, as a nonparametric method, the Spearman rank correlation does not make any assumptions about the normality of the distributions of the underlying data.

6.7.3 Spearman's method works by assigning a rank to each observation in each group separately (contrast this to the previous rank-sum methods in which the ranks are pooled). Ties are still ranked as in Mann-Whitney or Kruskal-Wallis, but the actual calculation does not have to be corrected. The Spearman's correlation is calculated according to the following formula:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

where:

n = number of observations, and
d_i = difference between the ranks of a pair.

7. Application

7.1 To illustrate the Spearman's test and bring together some common ideas between the test methods in this guide, we will consider an example that can be analyzed many ways. Suppose we are interested in a new laboratory test and how it compares with a specific outdoor exposure (Arizona, for example). There are ten different color specimens, and the durability measure is percent of gloss retained after exposure. We can think of this as a correlation test between the exposure conditions, or as a two-factor ANOVA-like test with gloss as the response variable, color as one predictor variable (10 levels), and exposure condition as another predictor variable (2 levels). See Table 8 for the data, along with rankings for use in the Spearman's calculation. Data analysis according to Spearman's method appears as follows, along with some other methods of comparison:

Spearman's Rank Correlation Analysis:
Dependent Variable: 60° Gloss Retention (%)
Grouped by Exposure Type
Number of Observations: 10

$$r_s = 1 - \frac{(6)[(2-2)^2+(1-1)^2+(10-10)^2+(9-9)^2+(7-6)^2+(4-3)^2+(3-4)^2+(8-8)^2+(5-5)^2+(6-7)^2]}{(10)(10^2-1)}$$

TABLE 8 Correlation Example

Gloss Retention	Color	Exposure Type	Rank
0.57	1	600 Hours laboratory	2
0.54	2	600 Hours laboratory	1
0.95	3	600 Hours laboratory	10
0.91	4	600 Hours laboratory	9
0.90	5	600 Hours laboratory	7
0.73	6	600 Hours laboratory	4
0.71	7	600 Hours laboratory	3
0.91	8	600 Hours laboratory	8
0.74	9	600 Hours laboratory	5
0.90	10	600 Hours laboratory	6
0.19	1	12 Months AZ direct	2
0.18	2	12 Months AZ direct	1
0.85	3	12 Months AZ direct	10
0.83	4	12 Months AZ direct	9
0.57	5	12 Months AZ direct	6
0.25	6	12 Months AZ direct	3
0.33	7	12 Months AZ direct	4
0.72	8	12 Months AZ direct	8
0.41	9	12 Months AZ direct	5
0.65	10	12 Months AZ direct	7

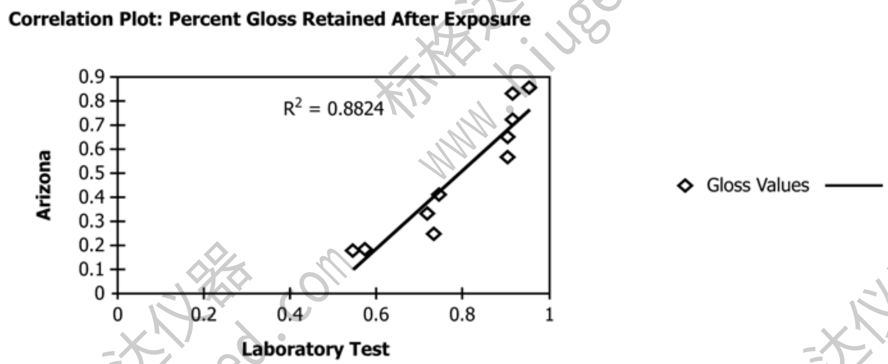


FIG. 4 Correlation Example

$r_s = 0.975758$

Linear Regression Analysis (Correlation):
Dependent Variable: 60° Gloss Retention (%)
Number of Observations: 10
Multiple R: 0.9394
Squared Multiple R: 0.8824
(See Fig. 4.)

Analysis of Variance:
Dependent Variable: 60° Gloss Retention (%)

Source	Sum of squares	Degrees of Freedom	Mean Square	*F* Ratio	*P*-value
Color	0.733641	9	0.081516	9.39231	0.001323
Exposure	0.416793	1	0.416793	48.02333	6.84E05
Error	0.078111	9	0.008679	-	-

8. Summary of Results

8.1 The Spearman's method indicates good agreement in material durability rankings between the exposures. Linear regression indicates a good fit to a linear model.

8.2 The correlation plot illustrates this graphically. However, from the plot, we see that the Arizona exposure resulted in lower retained gloss overall. We also see that there is a wide spread in durability for the 10 different colors.

8.3 ANOVA detects the differences in harshness between exposures, and indicates that they are significantly different. ANOVA also detects the differences in retained gloss across the ten colors, indicating that in this example, color is a significant factor.

9. Keywords

9.1 experimental design; statistics; weathering

APPENDIXES

(Nonmandatory Information)

X1. RESOURCES

Downie, N. M., and Heath, R. W., *Basic Statistical Methods*, 4th ed., Harper & Row Publishers, New York, 1974.

Freund, J. E., *Modern Elementary Statistics*, 4th ed., Prentice Hall, 1974.

Simon, L. E., *An Engineer's Manual of Statistical Methods*, John Wiley & Sons, New York, 1941.

Sheskin, David J., *Handbook of parametric and Nonparametric Statistical Procedures*, CRC Press, New York, 1997.

Gonick, Larry, and Smith, Woolcott, *The Cartoon Guide to Statistics*, Harper Collins, New York, 1993.

X2. CHI-SQUARE TABLE

TABLE X2.1 Critical Values for α

df	0.05	0.01	0.001
1	3.84	6.64	10.83
2	5.99	9.21	13.82
3	7.82	11.35	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.52
6	12.59	16.81	22.46
7	14.07	18.48	24.32
8	15.51	20.09	26.13
9	16.92	21.67	27.88
10	18.31	23.21	29.59
11	19.68	24.73	31.26
12	21.03	26.22	32.91
13	22.36	27.69	34.53
14	23.69	29.14	36.12
15	25	30.58	37.7
16	26.3	32	39.25
17	27.59	33.41	40.79
18	28.87	34.81	42.31
19	30.14	36.19	43.82
20	31.41	37.57	45.32
21	32.67	38.93	46.8
22	33.92	40.29	48.27
23	35.17	41.64	49.73
24	36.42	42.98	51.18
25	37.65	44.31	52.62
26	38.89	45.64	54.05
27	40.11	46.96	55.48
28	41.34	48.28	56.89
29	42.56	49.59	58.3
30	43.77	50.89	59.7
40	55.76	63.69	73.41
50	67.51	76.15	86.66
60	79.08	88.38	99.62
70	90.53	100.42	112.31
80	101.88	112.33	124.84
90	113.15	124.12	137.19
100	124.34	135.81	149.48