Designation: C1067 – 12

Standard Practice for Conducting a Ruggedness Evaluation or Screening Program for Test Methods for Construction Materials[1]

This standard is issued under the fixed designation C1067; 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 practice covers a procedure for evaluating the ruggedness of a test method by determining the effects of different experimental factors on the variation of test results. The procedure is intended for use during the development of a test method before the interlaboratory study is executed, such as those described in Practices C802 and E691.

1.2 This practice covers, in general terms, techniques for planning, collecting data, and analyzing results from a few laboratories. Appendix X1 provides the details of the procedure with an example and Appendix X2 provides additional information on the methodology.

1.3 The practice is not intended to give information pertinent to estimating multilaboratory precision.

1.4 The system of units for this practice is not specified. Dimensional quantities in the practice are presented only in illustrations of calculation methods.

1.5 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

C670 Practice for Preparing Precision and Bias Statements for Test Methods for Construction Materials

C802 Practice for Conducting an Interlaboratory Test Program to Determine the Precision of Test Methods for Construction Materials

E456 Terminology Relating to Quality and Statistics

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

E1169 Practice for Conducting Ruggedness Tests

3. Terminology

3.1 *Definitions:*

3.1.1 For definitions of statistical terms used in this standard, refer to Terminology E456.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *determination, n*—numerical value of a characteristic of a test specimen measured in accordance with the given the test method.

3.2.2 *effect, n—of a factor*, the difference in the measured characteristics at each level of a factor averaged over all levels of other factors in the experiment.

3.2.3 *factor, n*—a condition or element in the test procedure or laboratory environment that can be controlled and that is a potential source of variation of determinations.

3.2.4 *level, n*—the value or setting of a factor associated with a determination.

3.2.5 *replication, n*—the act of obtaining, under specified conditions, two or more determinations on identical specimens.

3.2.5.1 *Discussion*—Replicate determinations are typically required to be obtained by the same operator, using the same apparatus, on specimens that are similar as possible, and during a short time interval.

3.2.6 *ruggedness, n*—the characteristic of a test method such that determinations are not influenced to a statistically significant degree by small changes in the testing procedure or environment.

3.2.6.1 *Discussion*—Statistical significance is evaluated by comparing the observed variation due to a factor to the expected variation due to chance alone.

3.2.7 *screening, n*—a planned experiment using a low number of determinations to detect among many factors those that have a significant effect on variation of determinations compared with chance variation.

3.2.7.1 *Discussion*—In this practice, the influence of seven factors is evaluated using a replicated set of eight determinations, that is, a total of 16 determinations.

[1] This practice is under the jurisdiction of ASTM Committee C09 on Concrete and Concrete Aggregates and is the direct responsibility of Subcommittee C09.94 on Evaluation of Data (Joint C09 and C01).

Current edition approved July 1, 2012. Published September 2012. Originally approved in 1987. Last previous edition approved in 2007 as C1067 – 00 (2007). DOI: 10.1520/C1067-12.

[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.

***A Summary of Changes section appears at the end of this standard**

4. Summary of Practice

4.1 The practice requires that the user develop, from theoretical or practical knowledge, or both, a list of factors that plausibly would cause significant variation in test results (determinations) if the factors were not controlled. The technique is limited to the analysis of the effects of seven factors and requires 1/8 of the determinations that would be required to evaluate seven factors in a full factorial study. Procedures exist for analysis of smaller and larger numbers of factors (see Guide E1169), but seven is a convenient number for many test methods for construction materials. The seven-factor analysis requires 16 determinations by each laboratory. The procedure can be executed usefully by a single laboratory, but sometimes additional information can be obtained if it is repeated in one or two additional laboratories.

4.2 The procedure requires that two levels of each factor be identified, and 16 determinations be obtained with prescribed combinations of factor levels. The levels assigned to a factor may be quantitative or qualitative (for example, 20°C versus 25°C or brass versus steel).

4.3 After data are acquired, a statistical procedure is applied to establish which of the factors under study have a statistically significant effect on test results.

5. Significance and Use

5.1 The purpose of a ruggedness evaluation, or screening program, is to determine the sensitivity of the test method to changes in levels of pertinent operating factors using a small number of tests. Normally, operating conditions for a test method are defined along with allowable tolerances. A ruggedness analysis determines the effect of "worst-case" variation in operating conditions within the specified tolerances. If the ruggedness evaluation indicates high variation (poor precision), the method can be revised with smaller tolerances on operating conditions to improve the precision.

5.2 This practice evaluates the effects of seven factors using eight treatments. The disadvantage of this approach is that it only estimates the main effects of the factors and does not detect the effects of interactions among factors. For this reason, this is a screening program and additional investigation is required to investigate whether there are interaction effects.

5.3 A major reason for poor precision in test methods is the lack of adequate control over the sources of variation in testing procedures or testing environments. These sources of variation often are not controlled adequately because they were not identified during the development of the test procedures as having a large effect on the determinations. This practice provides a systematic procedure to establish the required degree of control for different testing parameters.

5.4 All new test methods must be subjected to an interlaboratory program to develop a precision and bias statement. These programs can be expensive and lengthy, and the result may show that the method is too variable and should not be published without further revision. Interlaboratory studies may give the subcommittee an indication that the method is too variable, but they do not usually give a clear picture of the causes of the variation. Application of this practice using one or two laboratories before finalizing the test method and conducting the interlaboratory study is an economical way to determine these causes.

5.5 Many existing test methods were developed before there was a requirement for precision and bias statements. Since this became a requirement, most of these test methods have developed precision and bias statements, and the result is that many have been found to suffer from relatively large amount of variation. This practice provides a relatively simple and economical way to investigate the causes of variation in test methods, so that a subcommittee will have some guidance as to which parts of the test method need to be revised.

5.6 The procedure can be used for a screening program within a single laboratory, but involvement of at least three laboratories is recommended, particularly if the single laboratory were to be the one that developed the test method. This is particularly important for new test methods. The originating laboratory is so much a part of the development of the test method that it is difficult for it to be objective in spotting any problems in the clarity of the test method directions. Two additional laboratories will probably contribute fresh critical review of the validity of the test method and provide assistance in clarifying the instructions of the test method when needed. This practice, however, is not intended to provide information on multilaboratory precision, but it does provide some information on single-operator precision, which could be used to develop a temporary repeatability statement until the interlaboratory study is completed.

6. Materials

6.1 The number and types of material shall cover the range of material properties to which the test method is applicable. The test method may not apply to material types or property values outside the range evaluated. Three to five materials with different properties will usually be sufficient.

6.1.1 Some preliminary testing may help the laboratories involved determine the materials that will be used in the screening program.

7. Procedure

7.1 Determine the number of laboratories that will participate in the screening program and which materials each will use. The maximum amount of information is obtained if all laboratories include all materials in their part of the program, however, cost can be reduced if each laboratory uses a different material. In this case, caution must be exercised in interpreting the results because laboratory-dependent effects cannot be separated from material-dependent effects.

7.2 Factors that are likely to have the greatest effect on the variability of the determinations are selected for study. Levels of these factors are determined by selecting the minimum and maximum levels that would plausibly occur in the execution of the test method if there were no particular efforts to control them. Levels often represent quantitative factors, such as temperature or pressure, but they may also represent qualitative factors, such as old versus new or wet versus dry. Only two levels are allowed for each factor. In this practice, factors are

TABLE 1 Pattern of Assigning Levels[A] to Seven Factors

Determination Number	Factor						
	A	B	C	D	E	F	G
1 (9)[B]	a	b	c	D	E	F	g
2 (10)	a	b	C	D	e	f	G
3 (11)	a	B	c	d	E	f	G
4 (12)	a	B	C	d	e	F	g
5 (13)	A	b	c	d	e	F	G
6 (14)	A	b	C	d	E	f	g
7 (15)	A	B	c	D	e	f	g
8 (16)	A	B	C	D	E	F	G

[A] Lower case letter indicates one level for the factor and upper case letter indicates the other level.
[B] The numbers in parentheses refer to the determinations in replicate set 2.

assigned letter designations, *A* through *G*, and the two levels of each factor are designated with upper and lower cases of these letters, as shown in Table 1.

NOTE 1—In textbooks dealing with design of experiments, factor levels are often denoted with plus (+) and minus (-) signs.

7.3 Assign combinations of factor levels to each determination according to Table 1. The eight determinations will be replicated; therefore, the full study on each material will require 16 determinations. Run the 16 determinations in random order.

7.4 To analyze the results, construct a 16 row by 16 column results matrix composed of ±1 values as shown in Table 2. The values in row 1 are all +1. The values in rows 2 to 8 for each replicate set correspond to the high and low settings of the factors as given in Table 1. The pattern in rows 1 to 8 of the first replicate set is repeated for rows 9 to 16 of the second replicate set. For rows 9 to 16 of the second replicate set, the signs are reversed from those in the first set. The various combinations of plus and minus values in Table 2 are applied to the values of the 16 determinations and various sums of the signed determinations are calculated. For each row of Table 2, calculate the *Z* and *W* statistics using Eq 1 and 2.

$$Z_r = \sum_{i=1}^{16} \alpha_{ri}\, d_i \tag{1}$$

$$W_r = \frac{Z_r^2}{16} \tag{2}$$

where:

r = row number as shown in Table 2, where r = 1 to 16,
i = determination number ranging from 1 to 16,
α_{ri} = +1 or -1 as defined in Table 2 for each row number and determination number, and
d_i = determination number i as defined in Table 1.

7.5 The *Z*-statistic for row 1 (Z_1) represents the sum of the 16 determinations and $Z_1/16$ is the overall average of the 16 determinations. The *Z*-statistics for rows 2 through 8 (Z_2 through Z_8) are related to the effects of each of the seven factors (see Note 2). These values of *Z* represent the differences between the sum of the determinations at the high level of the factor and the sum of the determinations at the low level of the factor. The *Z*-values are divided by eight to obtain the effect of each factor averaged of over the levels of the other factors. For example, $Z_3/8$ is the average effect of factor *B* as it is varied from the low level to the high level.

NOTE 2—A positive value for an effect of a factor means that the response increases as the factor level is changed from its low level to its high level. The opposite is the case for a negative effect. Recall that an effect is the difference between the average of the determinations at the high setting minus the average at the low setting of the factor.

7.6 The *W* values are various mean squares. W_1 is the mean of the square of the sum of all determinations and is not used in this analysis. The values W_2 to W_8 are the mean squares for each factor and are compared with the random error (see Note 3). The *W* values for rows 9 through 16 (W_9 to W_{16}) are used to calculate the error variance (s^2) according to Eq 3 (see Note 4).

$$s^2 = \frac{\sum_{r=9}^{16} W_r}{8} \tag{3}$$

NOTE 3—Appendix X2 provides additional information of the meaning of the term "mean squares."

NOTE 4—The error variance s^2 is the pooled variance of the two replicate determinations for each of the eight conditions.

7.7 To establish whether a factor has a statistically significant effect on the results, compute the *F* statistic for each factor using Eq 4.

$$F_f = \frac{W_r}{s^2} \tag{4}$$

TABLE 2 Matrix of Signs to be Applied to 16 Determinations (d_1 to d_{16}) to Calculate *Z*- and *W*-Statistics

	Sign Applied to Each Determination in Computing Z_i																	
	Eight Determinations for Replicate Set 1								Eight Determinations for Replicate Set 2									
row	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	*Z*	*W*
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Z_1	W_1
2	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	Z_2	W_2
3	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	Z_3	W_3
4	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	Z_4	W_4
5	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	Z_5	W_5
6	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	Z_6	W_6
7	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	Z_7	W_7
8	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	Z_8	W_8
9	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	Z_9	W_9
10	-1	-1	-1	-1	1	1	1	1	1	1	1	1	-1	-1	-1	-1	Z_{10}	W_{10}
11	-1	-1	1	1	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	Z_{11}	W_{11}
12	-1	1	-1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	Z_{12}	W_{12}
13	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1	1	1	-1	-1	Z_{13}	W_{13}
14	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	1	-1	1	-1	Z_{14}	W_{14}
15	1	-1	-1	1	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	Z_{15}	W_{15}
16	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1	-1	1	1	-1	Z_{16}	W_{16}

TABLE 3 Summary of Statistics for Seven Factors and Random Error

Factor	W	F
A	W_2	$F_A = W_2/s^2$
B	W_3	$F_B = W_3/s^2$
C	W_4	$F_C = W_4/s^2$
D	W_5	$F_D = W_5/s^2$
E	W_6	$F_E = W_6/s^2$
F	W_7	$F_F = W_7/s^2$
G	W_8	$F_G = W_8/s^2$
	W_9	
	W_{10}	
	W_{11}	
	W_{12}	$s^2 = \frac{\sum_{r=9}^{16} W_r}{8}$
	W_{13}	
	W_{14}	
	W_{15}	
	W_{16}	

where:

F_f = value of F-statistic for factor f (A through G) for the corresponding row (2 through 8) of Table 2.

Table 3 summarizes the calculations given by Eq 3 and 4.

7.8 An F_f value that is ≥ 5.32 represents a statistically significant effect for factor f at a probability of not greater than 5 % for drawing an erroneous conclusion.

7.9 An example of an analysis of data representing results on 4 materials from 3 laboratories is presented in Appendix X1.

7.10 If desired, one of the alternative methods discussed in X2.5 of Appendix X2 is permitted for determining which factors have statistically significant effects.

8. Keywords

8.1 analysis of variance; precision; ruggedness; screening; test method; variation

APPENDIXES

(Nonmandatory Information)

X1. EXAMPLE OF A RUGGEDNESS PROGRAM

X1.1 This appendix describes the procedure for conducting a ruggedness evaluation using as an example a proposed test method for measuring the viscosity of asphalt. Three laboratories participated in the program.

X1.2 As the first step in the ruggedness evaluation, each of the laboratories critically examined the procedure in the proposed test method. The objectives of the examination were as follows:

1. To determine if the instructions were clear, concise, and complete,
2. To decide which factors were likely to influence test results and therefore should be included in the study,
3. To select materials that covered the range of the property of interest for the range of physical forms of the materials to be tested, and
4. To determine the proper levels to be evaluated for each of the chosen factors.

X1.3 In this example, representatives of the three laboratories, after familiarizing themselves with the proposed test method, met and tried to improve the instructions for the viscosity measurement. They selected factors and levels that they believed could affect the measured viscosities. In a preliminary investigation, one of the laboratories measured viscosity at 24°C, 25°C, and 26°C and found that there was about a 10 % variation with a change of 1°C. This was considered too large so 24.6 and 25.4°C were selected as the lower and upper temperature levels for the ruggedness evaluation. In the same manner, the effects of the other factors were examined and the two levels to be used for each factor were selected. The seven factors selected for the program and their levels are shown in Table X1.1. The levels of the factors were assigned to each of the eight determinations in accordance with Table 1 from the body of this practice. Table X1.2 shows the testing conditions (or treatments) for each of the eight replicated determinations.

X1.4 Four materials were selected to cover the range of viscosities to be measured by the test method. For each testing condition, the viscosities were determined by each of the three laboratories with one replication. Thus each laboratory conducted 16 determinations for each material, for a total of 64 determinations. For each material, the 16 determinations were acquired in random order. This is a critical part of the program to guard against systematic variations in the testing conditions. The tests results, grouped by laboratory, are shown in Table X1.3.

X1.5 After the data were obtained, the results for each laboratory-material combination were analyzed independently. Thus in this program, there are 12 analyses corresponding to each row of data in Table X1.3. To proceed with each analysis, the relevant row of data from Table X1.3 is copied into 16 rows to create a 16 by 16 matrix. Each column corresponds to a determination and the value of that determination is repeated 16 times. The numbers in the matrix are multiplied by the corresponding values of +1 or -1 given in Table 2 in the body of this practice. Table X1.4 is an example of the resulting matrix derived from the data for Material 1 and Laboratory 1 in Table X1.3.

X1.6 After the 16 by 16 matrix with the proper signs applied to each determination has been created, the next step is to calculate the sum of each row, with due regard to sign, to obtain 16 Z-values, which are identified as Z_1 to Z_{16}. Table X1.5 shows the resulting sums for Laboratory 1 and Material 1. The value Z_1 represents the sum of all viscosities and $Z_1/16$ is the overall average viscosity for the laboratory-material combination. The value Z_2 represents the difference between the results at the high level of factor A and at the low level. In this case factor A is temperature, so Z_2 measures the effect of temperature. In the same manner, Z_3 measures the effect of the factor B, the age of the viscometer. The value Z_4 measures the effect of factor C, the vacuum level. The value Z_5 measures the effect of factor D, whether or not the sample is stirred before filling the viscometer. The value Z_6 measures the effect of factor E, whether the viscometer is vertical or slanted slightly. The value Z_7 measures the effect of factor F, the variation of the height of the asphalt when the viscometer is filled. The value Z_8 measures the effect of factor G, the time that the viscometer is kept in the water bath before testing. Each of these Z-values comprises eight determinations at one level of the factor and eight determinations at the other level. Therefore, the effect of a factor is obtained by dividing the corresponding Z-value by eight.

X1.7 The next step in the analysis is to square the Z-values and divide the squares by 16. The resulting values, which are denoted W_1 to W_{16} are various kinds of "mean sum of squares." As far as the ruggedness evaluation is concerned, the values W_2 to W_8 are measures of the variance of the means associated with each factor level. For example W_2 is the variance associated with the average values of the determinations obtained at the high and low temperatures. See Appendix X2 for more discussion on the meaning of the W-values.

X1.8 To determine if a factor has a statistically significant effect, the values of W_2 to W_8 are compared with the error variance (also called the mean square error). The error variance is the within-test variance calculated from the replicate determinations for each of the eight conditions and it indicates the random error associated with the test method. The error variance is obtained by calculating the sum of W_9 to W_{16} and diving by 8 as indicated by Eq 3 in the body of this practice. The calculated values of s^2 for each laboratory-material combination are shown in Tables X1.5-X1.16. If there are duplicate determinations, as is the case in this program, the error variance can also be determined as follows:

$$s^2 = \frac{\sum_{1}^{k} \Delta^2}{2k} \qquad \text{(X1.1)}$$

where:

s^2 = error variance or the pooled within-test variance,
Δ = the difference between duplicate determinations, and
k = number of pairs of determinations (k=8 in this program).

X1.9 The final step in the analysis is to compute the F-values for each of the factors by dividing W_2 to W_8 by s^2 as indicated by Eq 4 in the body of the practice. The calculated F-values for each laboratory-material combination are shown in Tables X1.5-X1.16. These values are compared with the critical F-value at a significance level of 0.05 for 1 degree of freedom for the numerator and 8 degrees of freedom for the denominator. The critical value is 5.32. If the calculated F-value for a factor is ≥5.32, the factor has a statically significant effect with no more than a 5 % probability of making the incorrect inference.

X1.10 The calculated F-values that exceed the critical value are shown as bold numbers in Tables X1.5-X1.16. Table X1.17 summarizes the calculated F-values for all factors an all laboratory-material combinations. All F-values that are less than 5.32 are indicated in the table as NS to show that they are not statistically significant, and the corresponding factor does not have a statistically significant effect on the results. The effect of temperature (factor A) was found to be highly significant for every material and every laboratory indicating the importance of tight control of temperature. The effect of variation in the level of vacuum (factor C) showed five statistically significant F-values indicating a need for tight tolerance on the applied vacuum. The effect of the viscometer deviating from the vertical position (factor E) was statistically significant in six of the laboratory-material combinations indicating the need for tight tolerance on the alignment of the viscometer. The other factors showed a scattering of barely significant values, but these were not judged to be of sufficient importance to require tighter controls.

X1.11 Representatives of the three laboratories met after completion of the ruggedness evaluation. After discussion of the results, the decision was made that it was practical and desirable to control temperature, vacuum, and the angle of the viscosity tube to within the following limits:

Temperature: 25.0 ± 0.1°C
Vacuum: 300 ± 2 mmHg
Angle with horizontal: 90.0 ± 1.0°

X1.11.1 With these changes, an interlaboratory study was organized and carried out using the revised test method.

TABLE X1.1 Levels Assigned to Seven Factors

Factor	Level
A: Temperature	a = 24.6°C A = 25.4°C
B: Age of viscometer tube	b = New B = Old
C: Applied vacuum	c = 310 mmHg C = 290 mmHg
D: Stirring sample before charging viscometer	d = No stirring D = Stir for 1 minute
E: Angle of viscometer	e = 87° from horizontal E = 90° from horizontal
F: Height of filling	f = 6 mm (1 mm above line) F = 4 mm (1 mm bellow line)
G: Time viscometer held in bath	g = 40 min (10 min more than specified) G = 20 min (10 min less than specified)

TABLE X1.2 Conditions for Each Determination

Determination Number	Factor *A* Temperature	*B* Viscometer	*C* Vacuum	*D* Stirring	*E* Angle	*F* Fill Height	*G* Time in Bath
1 (9)	24.6°C	New	310 mmHg	1 min	90º	4 mm	40 min
2 (10)	24.6°C	New	290 mmHg	1 min	87º	6 mm	20 min
3 (11)	24.6°C	Old	310 mmHg	No	90º	6 mm	20 min
4 (12)	24.6°C	Old	290 mmHg	No	87º	4 mm	40 min
5 (13)	25.4°C	New	310 mmHg	No	87º	4 mm	20 min
6 (14)	25.4°C	New	290 mmHg	No	90º	6 mm	40 min
7 (15)	25.4°C	Old	310 mmHg	1 min	87º	6 mm	40 min
8 (16)	25.4°C	Old	290 mmHg	1 min	90º	4 mm	20 min

TABLE X1.3 Viscosity Data

Material	Viscosity															
	First Replicate Determination Number								Second Replicate Determination Number							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Laboratory 1															
1	2370	2258	2355	2185	1825	1845	1820	1830	2320	2275	2350	2380	1840	1850	1825	1820
2	520	495	519	480	401	404	398	402	492	516	490	522	390	408	402	395
3	4205	4006	4191	3846	3212	3284	3185	3221	4200	4160	4130	4020	3218	3180	3280	3280
4	1075	1061	1060	961	803	793	801	805	1050	1070	1015	1000	808	790	795	805
	Laboratory 2															
1	2350	2240	2335	2165	1805	1825	1800	1810	2280	2310	2400	2120	1825	1806	1809	1812
2	540	515	539	500	421	424	418	422	518	545	524	492	410	425	430	420
3	4235	4036	4121	3876	3242	3314	3117	3250	4250	4142	3960	4205	3310	3112	3240	3117
4	1102	1040	1085	980	820	811	824	828	1110	1125	1040	1050	825	804	816	835
	Laboratory 3															
1	2390	2278	2375	2205	1845	1865	1840	1850	2400	2268	2350	2250	1860	1850	1870	1845
2	510	485	509	470	391	394	388	392	505	482	510	480	395	390	385	392
3	4200	3975	4160	3816	3190	3246	3150	3200	4180	3990	4140	3890	3200	3180	3220	3195
4	1050	990	1035	930	786	766	775	780	1040	980	1050	970	780	760	785	782

TABLE X1.4 Analysis Matrix Based on Applying Signs in Table 1 to Data for Laboratory 1 and Material 1

NOTE 1—The data in Tables X1.5-X1.16 are derived from matrices constructed as illustrated by this table for each of the remaining eleven laboratory-material combinations from Table X1.3.

Row	Replicate 1								Replicate 2							
	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}	d_{16}
1	2370	2258	2355	2185	1825	1845	1820	1830	2320	2275	2350	2380	1840	1850	1825	1820
2	-2370	-2258	-2355	-2185	1825	1845	1820	1830	-2320	-2275	-2350	-2380	1840	1850	1825	1820
3	-2370	-2258	2355	2185	-1825	-1845	1820	1830	-2320	-2275	2350	2380	-1840	-1850	1825	1820
4	-2370	2258	-2355	2185	-1825	1845	-1820	1830	-2320	2275	-2350	2380	-1840	1850	-1825	1820
5	2370	2258	-2355	-2185	-1825	-1845	1820	1830	2320	2275	-2350	-2380	-1840	-1850	1825	1820
6	2370	-2258	2355	-2185	-1825	1845	-1820	1830	2320	-2275	2350	-2380	-1840	1850	-1825	1820
7	2370	-2258	-2355	2185	1825	-1845	-1820	1830	2320	-2275	-2350	2380	1840	-1850	-1825	1820
8	-2370	2258	2355	-2185	1825	-1845	-1820	1830	-2320	2275	2350	-2380	1840	-1850	-1825	1820
9	2370	2258	2355	2185	1825	1845	1820	1830	-2320	-2275	-2350	-2380	-1840	-1850	-1825	-1820
10	-2370	-2258	-2355	-2185	1825	1845	1820	1830	2320	2275	2350	2380	-1840	-1850	-1825	-1820
11	-2370	-2258	2355	2185	-1825	-1845	1820	1830	2320	2275	-2350	-2380	1840	1850	-1825	-1820
12	-2370	2258	-2355	2185	-1825	1845	-1820	1830	2320	-2275	2350	-2380	1840	-1850	1825	-1820
13	2370	2258	-2355	-2185	-1825	-1845	1820	1830	-2320	-2275	2350	2380	1840	1850	-1825	-1820
14	2370	-2258	2355	-2185	-1825	1845	-1820	1830	-2320	2275	-2350	2380	1840	-1850	1825	-1820
15	2370	-2258	-2355	2185	1825	-1845	-1820	1830	-2320	2275	2350	-2380	-1840	1850	1825	-1820
16	-2370	2258	2355	-2185	1825	-1845	-1820	1830	2320	-2275	-2350	2380	-1840	1850	1825	-1820

TABLE X1.5 Results of Calculations for Laboratory 1 and Material 1

Z_1 = 33148	W_1 = 68,674,369.00	Avg. = 2071.8
Z_2 = -3838	W_2 = 920,640.25	F_A = **357.41**[A]
Z_3 = -18	W_3 = 20.25	F_B = 0.01
Z_4 = -262	W_4 = 4,290.25	F_C = 1.67
Z_5 = -112	W_5 = 784.00	F_D = 0.30
Z_6 = 332	W_6 = 6,889.00	F_E = 2.67
Z_7 = -8	W_7 = 4.00	F_F = 0.00
Z_8 = -42	W_8 = 110.25	F_G = 0.04
Z_9 = -172	W_9 = 1,849.00	
Z_{10} = 142	W_{10} = 1,260.25	s^2 = 2575.88
Z_{11} = -198	W_{11} = 2,450.25	s = 50.75
Z_{12} = -242	W_{12} = 3,660.25	
Z_{13} = 248	W_{13} = 3,844.00	
Z_{14} = 292	W_{14} = 5,329.00	
Z_{15} = -128	W_{15} = 1,024.00	
Z_{16} = 138	W_{16} = 1,190.25	

[A] Bold numbers in Tables X1.5-X1.16 indicate statistically significant values.

TABLE X1.6 Results of Calculations for Laboratory 1 and Material 2

Z_1 = 7234	W_1 = 3,270,672.25	Avg. = 452.1
Z_2 = -834	W_2 = 43,472.25	F_A = **172.51**
Z_3 = -18	W_3 = 20.25	F_B = 0.08
Z_4 = 10	W_4 = 6.25	F_C = 0.02
Z_5 = 6	W_5 = 2.25	F_D = 0.01
Z_6 = 26	W_6 = 42.25	F_E = 0.17
Z_7 = -30	W_7 = 56.25	F_F = 0.22
Z_8 = -18	W_8 = 20.25	F_G = 0.08
Z_9 = 4	W_9 = 1.00	
Z_{10} = 16	W_{10} = 16.00	s^2 = 252.00
Z_{11} = -24	W_{11} = 36.00	s = 15.87
Z_{12} = -124	W_{12} = 961.00	
Z_{13} = 16	W_{13} = 16.00	
Z_{14} = 116	W_{14} = 841.00	
Z_{15} = 4	W_{15} = 1.00	
Z_{16} = 48	W_{16} = 144.00	

TABLE X1.7 Results of Calculations for Laboratory 1 and Material 3

Z_1 = 58618	W_1 = 214,754,370.25	Avg. = 3663.6
Z_2 = -6898	W_2 = 2,973,900.25	F_A = **586.74**
Z_3 = -312	W_3 = 6,084.00	F_B = 1.20
Z_4 = -624	W_4 = 24,336.00	F_C = 4.80
Z_5 = 456	W_5 = 12,996.00	F_D = 2.56
Z_6 = 764	W_6 = 36,481.00	F_E = **7.20**
Z_7 = -214	W_7 = 2,862.25	F_F = 0.56
Z_8 = 218	W_8 = 2,970.25	F_G = 0.59
Z_9 = -318	W_9 = 6,320.25	
Z_{10} = 206	W_{10} = 2,652.25	s^2 = 5068.50
Z_{11} = -216	W_{11} = 2,916.00	s = 71.19
Z_{12} = -248	W_{12} = 3,844.00	
Z_{13} = -288	W_{13} = 5,184.00	
Z_{14} = 540	W_{14} = 18,225.00	
Z_{15} = -150	W_{15} = 1,406.25	
Z_{16} = 2	W_{16} = 0.25	

TABLE X1.8 Results of Calculations for Laboratory 1 and Material 4

Z_1 = 14692	W_1 = 13,490,929.00	Avg. = 918.3
Z_2 = -1892	W_2 = 223,729.00	F_A = **828.24**
Z_3 = -208	W_3 = 2,704.00	F_B = **10.01**
Z_4 = -122	W_4 = 930.25	F_C = 3.44
Z_5 = 232	W_5 = 3,364.00	F_D = **12.45**
Z_6 = 94	W_6 = 552.25	F_E = 2.04
Z_7 = -78	W_7 = 380.25	F_F = 1.41
Z_8 = 162	W_8 = 1,640.25	F_G = **6.07**
Z_9 = 26	W_9 = 42.25	
Z_{10} = -18	W_{10} = 20.25	s^2 = 270.13
Z_{11} = -2	W_{11} = 0.25	s = 16.44
Z_{12} = -116	W_{12} = 841.00	
Z_{13} = 18	W_{13} = 20.25	
Z_{14} = 120	W_{14} = 900.00	
Z_{15} = -64	W_{15} = 256.00	
Z_{16} = 36	W_{16} = 81.00	

TABLE X1.9 Results of Calculations for Laboratory 2 and Material 1

Z_1 = 32692	W_1 = 66,797,929.00	Avg. = 2043.3
Z_2 = -3708	W_2 = 859,329.00	F_A = **813.76**
Z_3 = -190	W_3 = 2,256.25	F_B = 2.14
Z_4 = -516	W_4 = 16,641.00	F_C = **15.76**
Z_5 = 130	W_5 = 1,056.25	F_D = 1.00
Z_6 = 544	W_6 = 18,496.00	F_E = **17.52**
Z_7 = -358	W_7 = 8,010.25	F_F = **7.59**
Z_8 = 382	W_8 = 9,120.25	F_G = **8.64**
Z_9 = -32	W_9 = 64.00	
Z_{10} = 8	W_{10} = 4.00	s^2 = 1056.00
Z_{11} = -30	W_{11} = 56.25	s = 32.50
Z_{12} = 16	W_{12} = 16.00	
Z_{13} = 10	W_{13} = 6.25	
Z_{14} = 76	W_{14} = 361.00	
Z_{15} = 218	W_{15} = 2,970.25	
Z_{16} = -282	W_{16} = 4,970.25	

TABLE X1.10 Results of Calculations for Laboratory 2 and Material 2

Z_1 = 7543	W_1 = 3,556,053.06	Avg. = 471.4
Z_2 = -803	W_2 = 40,300.56	F_A = **331.86**
Z_3 = -53	W_3 = 175.56	F_B = 1.45
Z_4 = -57	W_4 = 203.06	F_C = 1.67
Z_5 = 73	W_5 = 333.06	F_D = 2.74
Z_6 = 81	W_6 = 410.06	F_E = 3.38
Z_7 = -97	W_7 = 588.06	F_F = 4.84
Z_8 = 49	W_8 = 150.06	F_G = 1.24
Z_9 = 15	W_9 = 14.06	
Z_{10} = -15	W_{10} = 14.06	s^2 = 121.44
Z_{11} = 11	W_{11} = 7.56	s = 11.02
Z_{12} = -57	W_{12} = 203.06	
Z_{13} = -51	W_{13} = 162.56	
Z_{14} = 61	W_{14} = 232.56	
Z_{15} = 71	W_{15} = 315.06	
Z_{16} = -19	W_{16} = 22.56	

TABLE X1.11 Results of Calculations for Laboratory 2 and Material 3

Z_1 = 58527.00	W_1 = 214,088,108.06	Avg. = 3657.9
Z_2 = -7123.00	W_2 = 3,171,070.56	F_A = **226.64**
Z_3 = -755.00	W_3 = 35,626.56	F_B = 2.55
Z_4 = -423.00	W_4 = 11,183.06	F_C = 0.80
Z_5 = 247.00	W_5 = 3,813.06	F_D = 0.27
Z_6 = 191.00	W_6 = 2,280.06	F_E = 0.16
Z_7 = 443.00	W_7 = 12,265.56	F_F = 0.88
Z_8 = -171.00	W_8 = 1,827.56	F_G = 0.13
Z_9 = -145.00	W_9 = 1,314.06	
Z_{10} = 433.00	W_{10} = 11,718.06	s^2 = 13991.81
Z_{11} = -171.00	W_{11} = 1,827.56	s = 118.29
Z_{12} = -55.00	W_{12} = 189.06	
Z_{13} = -77.00	W_{13} = 370.56	
Z_{14} = 1107.00	W_{14} = 76,590.56	
Z_{15} = -413.00	W_{15} = 10,660.56	
Z_{16} = 385.00	W_{16} = 9,264.06	

TABLE X1.12 Results of Calculations for Laboratory 2 and Material 4

Z_1 = 15095	W_1 = 14,241,189.06	Avg. = 943.4
Z_2 = -1969	W_2 = 242,310.06	F_A = **269.21**
Z_3 = -179	W_3 = 2,002.56	F_B = 2.22
Z_4 = -149	W_4 = 1,387.56	F_C = 1.54
Z_5 = 265	W_5 = 4,389.06	F_D = 4.88
Z_6 = 135	W_6 = 1,139.06	F_E = 1.27
Z_7 = 5	W_7 = 1.56	F_F = 0.00
Z_8 = 101	W_8 = 637.56	F_G = 0.71
Z_9 = -115	W_9 = 826.56	
Z_{10} = 121	W_{10} = 915.06	s^2 = 900.06
Z_{11} = 67	W_{11} = 280.56	s = 30.00
Z_{12} = -195	W_{12} = 2,376.56	
Z_{13} = -69	W_{13} = 297.56	
Z_{14} = 189	W_{14} = 2,232.56	
Z_{15} = -65	W_{15} = 264.06	
Z_{16} = 11	W_{16} = 7.56	

TABLE X1.13 Results of Calculations for Laboratory 3 and Material 1

Z_1 = 33341	W_1 = 69,476,392.56	Avg. = 2083.8
Z_2 = -3691	W_2 = 851,467.56	F_A = **3224.49**
Z_3 = -171	W_3 = 1,827.56	F_B = **6.92**
Z_4 = -519	W_4 = 16,835.06	F_C = **63.75**
Z_5 = 141	W_5 = 1,242.56	F_D = 4.71
Z_6 = 509	W_6 = 16,192.56	F_E = **61.32**
Z_7 = -51	W_7 = 162.56	F_F = 0.62
Z_8 = 1	W_8 = 0.06	F_G = 0.00
Z_9 = -45	W_9 = 126.56	
Z_{10} = -5	W_{10} = 1.56	s^2 = 264.06
Z_{11} = -45	W_{11} = 126.56	s = 16.25
Z_{12} = 15	W_{12} = 14.06	
Z_{13} = -5	W_{13} = 1.56	
Z_{14} = 115	W_{14} = 826.56	
Z_{15} = -85	W_{15} = 451.56	
Z_{16} = 95	W_{16} = 564.06	

TABLE X1.14 Results of Calculations for Laboratory 3 and Material 2

Z_1 = 7078	W_1 = 3,131,130.25	Avg. = 442.4
Z_2 = -824	W_2 = 42,436.00	F_A = **3857.82**
Z_3 = -26	W_3 = 42.25	F_B = 3.84
Z_4 = -108	W_4 = 729.00	F_C = **66.27**
Z_5 = 0	W_5 = 0.00	F_D = 0.00
Z_6 = 126	W_6 = 992.25	F_E = **90.20**
Z_7 = -8	W_7 = 4.00	F_F = 0.36
Z_8 = 34	W_8 = 72.25	F_G = **6.57**
Z_9 = 0	W_9 = 0.00	
Z_{10} = 6	W_{10} = 2.25	s^2 = 11.00
Z_{11} = -16	W_{11} = 16.00	s = 3.32
Z_{12} = -6	W_{12} = 2.25	
Z_{13} = 22	W_{13} = 30.25	
Z_{14} = 16	W_{14} = 16.00	
Z_{15} = -18	W_{15} = 20.25	
Z_{16} = -4	W_{16} = 1.00	

TABLE X1.15 Results of Calculations for Laboratory 3 and Material 3

Z_1 = 57932	W_1 = 209,757,289.00	Avg. = 3620.8
Z_2 = -6770	W_2 = 2,864,556.25	F_A = **2885.84**
Z_3 = -390	W_3 = 9,506.25	F_B = **9.58**
Z_4 = -948	W_4 = 56,169.00	F_C = **56.59**
Z_5 = 288	W_5 = 5,184.00	F_D = 5.22
Z_6 = 1070	W_6 = 71,556.25	F_E = **72.09**
Z_7 = -190	W_7 = 2,256.25	F_F = 2.27
Z_8 = 168	W_8 = 1,764.00	F_G = 1.78
Z_9 = -58	W_9 = 210.25	
Z_{10} = 40	W_{10} = 100.00	s^2 = 992.63
Z_{11} = -180	W_{11} = 2,025.00	s = 31.51
Z_{12} = 22	W_{12} = 30.25	
Z_{13} = -62	W_{13} = 240.25	
Z_{14} = 280	W_{14} = 4,900.00	
Z_{15} = -60	W_{15} = 225.00	
Z_{16} = 58	W_{16} = 210.25	

TABLE X1.16 Results of Calculations for Laboratory 3 and Material 4

Z_1 = 14259	W_1 = 12,707,442.56	Avg. = 891.2
Z_2 = -1831	W_2 = 209,535.06	F_A = **1523.20**
Z_3 = -45	W_3 = 126.56	F_B = 0.92
Z_4 = -343	W_4 = 7,353.06	F_C = **53.45**
Z_5 = 105	W_5 = 689.06	F_D = 5.01
Z_6 = 267	W_6 = 4,455.56	F_E = **32.39**
Z_7 = -23	W_7 = 33.06	F_F = 0.24
Z_8 = 107	W_8 = 715.56	F_G = 5.20
Z_9 = -35	W_9 = 76.56	
Z_{10} = 35	W_{10} = 76.56	s^2 = 137.56
Z_{11} = -99	W_{11} = 612.56	s = 11.73
Z_{12} = -17	W_{12} = 18.06	
Z_{13} = 51	W_{13} = 162.56	
Z_{14} = 33	W_{14} = 68.06	
Z_{15} = -17	W_{15} = 18.06	
Z_{16} = 33	W_{16} = 68.06	

TABLE X1.17 Summary of *F* Values[A] for All Laboratories, All Materials, and All Factors

Laboratory	Material	Average Viscosity	Table	Temperature, F_A	Age, F_B	Vacuum, F_C	Stir, F_D	Angle, F_E	Fill Height, F_F	Bath Time F_G
1	1	2071.8	X1.5	357.41	NS	NS	NS	NS	NS	NS
	2	452.1	X1.6	172.51	NS	NS	NS	NS	NS	NS
	3	3663.6	X1.7	586.74	NS	NS	NS	7.20	NS	NS
	4	918.3	X1.8	828.24	10.01	NS	12.45	NS	NS	6.07
2	1	2043.3	X1.9	813.76	NS	15.76	NS	17.52	7.59	8.64
	2	471.4	X1.10	331.86	NS	NS	NS	NS	NS	NS
	3	3657.9	X1.11	226.64	NS	NS	NS	NS	NS	NS
	4	943.4	X1.12	269.21	NS	NS	NS	NS	NS	NS
3	1	2083.8	X1.13	3224.49	6.92	63.75	NS	61.32	NS	NS
	2	442.4	X1.14	3857.82	NS	66.27	NS	90.20	NS	6.57
	3	3620.8	X1.15	2885.84	9.58	56.59	NS	72.09	NS	NS
	4	891.2	X1.16	1523.20	NS	53.45	NS	32.39	NS	NS

[A] The entry NS indicates that the calculated *F*-value is less than the critical value (5.32) and the effect of the factor is not statistically significant.

X2. THEORY OF THE RUGGEDNESS ANALYSIS

X2.1 Introduction

X2.1.1 Any statistical analysis depends on assumptions. Because a ruggedness or screening program is usually run on a new test method, there is little history or experience to validate the necessary assumptions. An extensive study could yield the experience to validate the assumptions, but it would also increase the cost of the ruggedness program to the point that few such programs could be undertaken. This practice seeks to balance these risks by making plausible assumptions to make the practice practical and useful.

X2.1.2 A ruggedness program attempts to identify the important factors that cause variability of results obtained using the test method. It is important that all of the major factors be included in the study, because if one is left out, the study will not be able to identify its significance. The procedure in this practice is set up to evaluate the effects of seven factors using as few tests as possible. This is usually sufficient to cover the major sources of variability. Designs for both fewer and more factors are given in statistical texts for use when needed (**1**, **2**, **3**).[3]

X2.1.3 It is unusual to need to investigate more than seven factors and it is typical to have at least five factors that can be varied. When only five factors are considered to be potentially significant, two other factors can nearly always be selected about which there may be some doubt. A seven factor analysis is usually suitable for most screening programs to evaluate the ruggedness of a test method.

X2.2 Factorial Designs

X2.2.1 This practice is based on a standard two-level, fractional factorial design (**3**). A full factorial experiment for seven factors at two levels would require 2^7 or 128 determinations for each laboratory and each material with no replication. If this design had been followed for the example in Appendix X1, 128 determinations times 3 laboratories times 4 materials times 2 replications equals 3072 determinations for the program instead of the 192 determinations that were used in the example. The fractional factorial design used in this practice permits the evaluation of the effects of 7 factors with 1/16 of the determinations that would be run in a full factorial design. There are, however, limitations to fractional factorial designs.

X2.2.2 A full factorial experiment identifies not only significant effects of the main factors but also those of the interactions of factors. An interaction is a source of variability due to the combination of main factors that cannot be explained by the individual effects of the main factors. For the fractional factorial design used in this practice, main factor effects are confounded with two-factor interactions. This means that when a factor effect is calculated (see X2.4), it is not known if two-factor interactions contribute to the calculated effect. There are instances where the effect of an interaction is greater than the sum of the effects of the main factors that create the interaction. Interactions are, however, usually regarded as being smaller sources of variation compared with main effects, particularly three- and four-factor interactions. In this practice, the effects of all interactions are assumed to be negligible. This is done to permit testing of the significance of a large number of main effects while holding the size of the experiment down to manageable levels. The effect of interactions will not always be negligible and there are times when an estimate of a main effect will include an interaction. If there is concern about certain two-factor interactions, additional factorial experiments can be run to evaluate their significance.

X2.2.3 In this practice, each factor is assigned one of two levels. Three or more levels could have been used. The higher number of levels would give information about the shape of the relationship between the measured response and factor level. An increase in the number of levels would also increase the number of runs or determinations. For example, seven factors at three levels would require 2187 runs for a full factorial design compared with 128 runs when two levels are used.

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

Because each of the factors in the screening program will vary typically by a small amount, the relationships between measured response and factor level would be substantially linear and there would be little gained by using more than two levels of a factor.

X2.3 Error Variance

X2.3.1 The statistical analysis in this practice is based on the assumption that the random errors are normally distributed. Because a new method is being evaluated, there would be few data available to show that the errors are normally distributed. The assumption, however, appears reasonable based on experience with other ASTM test methods.

X2.3.2 In order to evaluate whether variations in test results due to changes in factor levels are statistically significant, it is necessary to have an estimate of the random error associated with the test method. The random error is typically called the single-operator precision (see Practice C670). An estimate of random error can be obtained by using duplicate determinations for each testing condition. It is important in evaluations of this kind to keep the amount of work and therefore the cost as low as possible while obtaining valid results. The amount of work would be halved if the tests by each laboratory were regarded as a replication of the experiment. Experience in ASTM interlaboratory studies shows that there usually are significant differences in test results between laboratories. Therefore, in this practice each laboratory conducts duplicate tests under the same testing conditions, that is, for the same factor levels defined in Table 1. There are several techniques for manipulating the data from the duplicate testing to obtain the pooled estimate of error variance. In this practice, the calculation is based on summing the determination values after they have been multiplied by the appropriate ± 1 values given in rows 9 to 16 of Table 2 (see Eq 1). The resulting sums Z_9 to Z_{16} are squared and mean squares W_9 to W_{16} are calculated according to Eq 2. The average of the mean squares is the pooled error variance (Eq 3). This procedure can be implemented readily using an electronic spreadsheet. If the error variance is computed using another method (see Eq X1.1), rows 9 to 16 are not needed in the analysis matrix (Table 2). In this case it is only necessary to calculate Z_1 to Z_8 and the corresponding values of W_1 to W_8. The procedure for calculating the F-values remains the same.

X2.4 Effects of Factors

X2.4.1 In this practice, because there are duplicate tests, the factor effects are obtained by dividing each of the values Z_2 to Z_8 by eight. For each factor, the effect is the difference between the average of the determinations when the factor is at a high level and the average of the determinations when the factor is at a low level. Two-level factorial experiments are designed such that for all determinations used to calculate the average at one level of the factor, the other factors have the same number at a high level and at a low level. As a result, the effect of a factor is freed of the effects of the levels of the other factors.

X2.4.2 An effect can have a plus or minus sign. A plus sign means that the average measured value increases as the factor setting changes from the low level to the high level. A negative sign means the measured value decreases for the same change in factor levels. For quantitative factors, keep in mind that "low level" and "high level" do not necessarily coincide with the numerical value for the setting level. The level that is considered as the "low level" is defined in the design of the experiment. As shown in Table X1.1, in the example in Appendix X1 several of the "low levels" actually have higher numerical values for the factor setting. This has no effect on the results of the analysis, only in how the effects are interpreted. See X2.5 for additional discussion.

X2.5 Determining Significant Effects

X2.5.1 In a two-level factorial screening program, such as used in this practice, there are three types of means (or averages) that are used in analyzing whether the effect of any factor is statistically significant. These means are:

1. The grand mean (overall average) of all determinations.
2. The means of the determinations at the low and high levels of the factor (treatment averages).
3. The means of the duplicate determinations for each of the factor combinations (cell averages).

X2.5.1.1 The method of analysis in this practice is known formally as analysis of variance (ANOVA). It involves forming the ratio of the variance associated with changing the level of a factor to the error variance of the test method, which is estimated from the duplicate determinations. The basic principle is that if a factor has no effect, these two variances should be close to each other because both are estimates of the true error variance. If a factor has a large effect, the variance due to the factor levels will be larger and the ratio will be significantly greater than one. The method used to judge whether the ratio is significantly greater than one is discussed below.

X2.5.1.2 The determination of the error variance is discussed in X2.3. In terms of the three types of averages mentioned above, the error variance is based on the scatter of the replicate determinations about their average (cell average). For the case of duplicate measurements, as used in this practice, the error variance may be calculated using Eq X1.1. The corresponding degrees of freedom for this pooled variance is eight, because each variance estimated from duplicate determinations has one degree of freedom and eight estimates are pooled together. In ANOVA, the error variance is typically called the *mean square error (MSE)*.

X2.5.1.3 The variance associated with the change in factor levels is called the *mean square of treatments (MST)*. For this practice, *MST* is defined by the following equation:

$$MST = \frac{8\left(\left(\bar{Y}_+ - \bar{\bar{Y}}\right)^2 + \left(\bar{Y}_- - \bar{\bar{Y}}\right)^2\right)}{2-1} \qquad \text{(X2.1)}$$

where:

$\bar{Y}_+$ = average of determinations at the high level of the factor,
$\bar{Y}_-$ = average of determinations at the low level of the factor, and
$\bar{\bar{Y}}$ = overall average of all determinations.

X2.5.1.4 In Eq X2.1, the number 8 arises because eight values are used to calculate the averages for the high and low

factor levels. The denominator is the degrees of freedom associated with this variance, which is the number of factor levels (2) minus one.

X2.5.1.5 To determine whether the factor effect is statistically significant, the *F*-statistic is calculated as follows:

$$F = \frac{MST}{MSE} \quad \text{(X2.2)}$$

X2.5.1.6 If there is no factor effect, *MST* and *MSE* are both estimates of the error variance and the ratio *F* should follow an *F*-distribution for 1 degree of freedom in the numerator and 8 degrees of freedom in the denominator. For a risk level of 5 %, the critical *F*-value for 1 and 8 degrees of freedom is 5.32. In this screening program, if the calculated *F*-value for a factor is less than 5.32, it can be concluded that there is more than a 5 % probability that the difference between *MST* and *MSE* is due only to chance and the factor effect is not statistically significant. If the *F*-value is greater than 5.32, there is less than a 5 % probability that such a high value is due to only to chance, and it can be concluded with confidence that there is a statistically significant factor effect.

X2.5.2 *Example*—To illustrate an analysis of variance, consider the effect of the level of vacuum using the data for Laboratory 2 and Material 1 as shown in Table X1.3. These data are plotted in Fig. X2.1. For clarity, the determinations are plotted using slightly different values of vacuum rather than the actual values of 290 and 310 mmHg. The replicate determinations at the two levels of vacuum are also shown in Table X2.1. In this case, a vacuum of 310 mmHg is defined as the "low" level (see Table X1.1). The fifth column in Table X2.1 is the difference between the duplicate determinations, and the sixth column is the square of these differences. The sum of the squares of the differences is shown at the bottom of the column. The sum is divided by 16 to obtain the *MSE* (see Eq X1.1). The overall average of the 16 determinations, the average of the eight determinations at the high level of vacuum, and the average of the eight determinations at the low level of vacuum are shown at the bottom of the table and are plotted in Fig. X2.1. The value of *MST* is calculated according to Eq X2.1. The calculated *F*-value is 15.76. Because the *F*-value exceeds 5.32, there is less than a 5 % probability that this high value is due to chance alone. Thus we conclude that the level of vacuum has a statistically significant effect.

X2.5.3 *Comparison*—The results in Table X2.1 are compared with the results in Table X1.9. The value of *MST* is identical to W_4 and the value of *MSE* is identical to s^2. Thus the

TABLE X2.1 Data for Laboratory 2 and Material 1 and Analysis of Variance

Run	Vacuum, mmHg	Viscosity		Difference, Δ	Δ^2
		Replicate 1	Replicate 2		
1	310	2350	2280	70	4900
2	290	2240	2310	-70	4900
3	310	2335	2400	-65	4225
4	290	2165	2120	45	2025
5	310	1805	1825	-20	400
6	290	1825	1806	19	361
7	310	1800	1809	-9	81
8	290	1810	1812	-2	4
Overall Avg =2043.25				Sum Δ^2 =	16896
Avg_{290} =2011				s^2 =	16896/(16)=1056
Avg_{310} =2075.5					
Effect =2011 – 2075.5 = -64.5					
$MST_C = W_4 = 8 *[(2011-2043.25)^2 + (2075.5-2043.25)^2]/1 = 16,641$					
F_C =16,641/1056 =15.76					

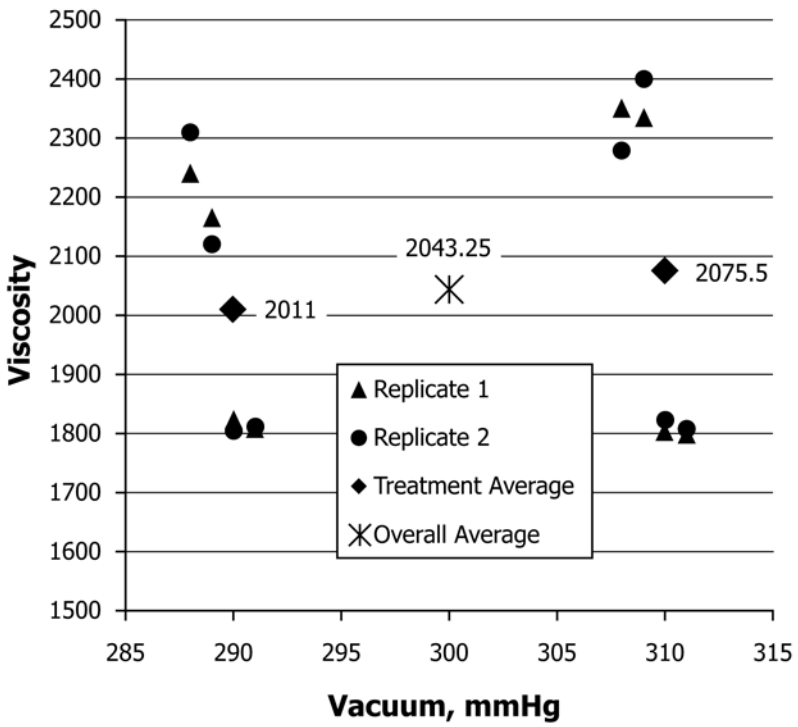

FIG. X2.1 Data for Laboratory 2 and Material 1 Plotted Versus the Level of Vacuum (Factor C)

analysis method adopted in this practice is in fact the same as a routine analysis of variance used to detect whether there are statistically significant differences among means.

X2.5.4 *ANOVA Table*—The analysis method used in this practice, as described by Table 2 and Eq 1 to Eq 4, is intended for application by users with limited background in statistical methods. The method can be implemented readily using an electronic spreadsheet, and the example in Appendix X1 provides the user a means to check that the spreadsheet is set up correctly. For users with knowledge of statistical tools, the use of a computer program that performs a multi-factor ANOVA is an alternative method for analyzing the data generated by a screening program. This requires that the data be arranged in a particular format depending on the program that is used. Table X2.2 shows data for Laboratory 2 and Material 1 arranged as input for a particular statistical analysis program. The ±1 values in columns X_A to X_G correspond to the factor levels for each of the 16 determinations (See Table 1). The output of the analysis is an ANOVA table as shown in Table X2.3. Comparison to Table X1.9 shows that the "Mean Squares" for rows *A* through *G* in Table X2.3 are identical to the values W_2 to W_8 and the value of *MSE* equals s^2. Therefore, the *F*-values for the seven factors are also identical. The last column of the ANOVA table gives the probability of observing the calculated value of *F* due to chance alone. In this practice, if an *F*-value has a probability ≤ 0.05, the corresponding factor is declared to have a statistically significant effect.

X2.5.5 *Linear Regression*—Another way of analyzing the results of the screening program is by carrying out a multivariate regression analysis of the data. In this practice, it is assumed that only the main effects of the seven factors affect the results. It is possible to express the expected result of a determination as a linear combination of the factor effects according to Eq X2.3.

TABLE X2.2 Data for Laboratory 2 and Material 1 Arranged for Input for ANOVA Analysis

X_A	X_B	X_C	X_D	X_E	X_F	X_G	*d*
-1	-1	-1	1	1	1	-1	2350
-1	-1	1	1	-1	-1	1	2240
-1	1	-1	-1	1	-1	1	2335
-1	1	1	-1	-1	1	-1	2165
1	-1	-1	-1	-1	1	1	1805
1	-1	1	-1	1	-1	-1	1825
1	1	-1	1	-1	-1	-1	1800
1	1	1	1	1	1	1	1810
-1	-1	-1	1	1	1	-1	2280
-1	-1	1	1	-1	-1	1	2310
-1	1	-1	-1	1	-1	1	2400
-1	1	1	-1	-1	1	-1	2120
1	-1	-1	-1	-1	1	1	1825
1	-1	1	-1	1	-1	-1	1806
1	1	-1	1	-1	-1	-1	1809
1	1	1	1	1	1	1	1812

TABLE X2.3 ANOVA Table for Laboratory 2 and Material 1

Source	df	Sums of Squares	Mean Square	F-ratio	Probability
Constant	1	6.68E+07	6.68E+07	63256	>0.0001
A	1	859329	859329	**813.76**	>0.0001
B	1	2256.25	2256.25	2.1366	0.182
C	1	16641	16641	**15.759**	0.0041
D	1	1056.25	1056.25	1.0002	0.3465
E	1	18496	18496	**17.515**	0.0031
F	1	8010.25	8010.25	**7.5855**	0.0249
G	1	9120.25	9120.25	**8.6366**	0.0187
Error	8	8448	1056		
Total	15	923357			

$$d = \bar{\bar{Y}} + \sum_{f=A}^{G} H_f X_f + \varepsilon \qquad \text{(X2.3)}$$

where:

d = expected property value for the treatment (factor settings),
$\bar{\bar{Y}}$ = overall average of all determinations,
H_f = half-effect for factor f,
X_f = factor level (±1) for each factor for the treatment, and
ε = random error.

X2.5.5.1 In this model, the coefficients H_f are "half-effects" because the variable *X* goes from -1 to +1, for a change of 2 units. If the *X* values were chosen to be 0 and 1 for the low and high levels, the coefficients would be the full effects. A multivariate regression analysis can be performed using an electronic spreadsheet. The data are arranged the same way as for an ANOVA as shown in Table X2.2. The regression analysis determines the overall average, the half-effects, and the random error. In addition, it determines the probability of obtaining each half-effect value due only to chance. Table X2.4 shows the results of a regression analysis of the data in Table X2.2. For this particular spreadsheet program, the output is divided into three parts. The upper part gives various measures of correlation, the random error, and the number of data points. The middle part gives the mean square error (the square of the random error) and other factors to evaluate the statistical significance of the linear model. The bottom part is the important part as far as evaluating which factors have a statistically significant effect. The column labeled "Coefficients" shows the overall mean (intercept) and values of the half-effects for factors *A* through *G*. The last column indicates the probability of obtaining half-effect values as large as those calculated by the analysis but due only to chance. Comparing the probabilities from the regression analysis to the probabilities in the ANOVA table in Table X2.3 shows that they are identical. Thus the same conclusion is reached with regard to which factors are statistically significant. As was stated in X2.4, the factor effects are equal to the values of Z_2 to Z_8 divided by eight. It can be shown that if the values of Z_2 to Z_8 in Table X1.9 are divided by 16, the results are identical to the half-effect coefficients in Table X2.4.

TABLE X2.4 Results of Regression Analysis for Laboratory 2 and Material 1

Regression Statistics	
Multiple R	0.9954
R Square	0.9909
Adjusted R Square	0.9828
Standard Error	32.5
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	914909	130701.3	123.77	1.61E-07
Residual	8	8448	1056		
Total	15	923357			

	Coefficients	*Standard Error*	*tStat*	*P-value*
Intercept	2043.25	8.12	251.51	6.992E-17
A	-231.75	8.12	-28.53	**2.466E-09**
B	-11.875	8.12	-1.46	0.1820
C	-32.25	8.12	-3.97	**0.0041**
D	8.125	8.12	1.00	0.3465
E	34	8.12	4.19	**0.0031**
F	-22.375	8.12	-2.75	**0.0249**
G	23.875	8.12	2.94	**0.0187**

REFERENCES

(1) Kempthorne, O., *Design and Analysis of Experiments*, John Wiley and Sons, Inc., New York, NY, 1952.
(2) Cochran, W. G., and Cox, G. M., *Experimental Designs*, John Wiley and Sons, Inc., New York, NY, 1957.
(3) Box, G. E. P., Hunter, W .G., and Hunter, J. S., *Statistics for Experimenters*, John Wiley and Sons, Inc., New York, NY, 1978.

SUMMARY OF CHANGES

Committee C09 has identified the location of selected changes to this practice since the last issue, C1067 – 00 (2007), that may impact the use of this practice. (Approved July 1, 2012.)

(1) Made extensive revisions to Sections 1 – 7.
(2) Changed the supplementary information to nonmandatory Appendixes.
(3) Added information on using ANOVA and linear regression analysis to analyze data resulting from this practice.