



# Standard Guide for Reducing Test Variability<sup>1</sup>

This standard is issued under the fixed designation D 4853; 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 ( $\epsilon$ ) indicates an editorial change since the last revision or reapproval.

## 1. Scope

1.1 This guide serves as an aid to subcommittees writing and maintaining test methods. It helps to (1) determine if it is possible to reduce test variability, and, if so, (2) provide a systematic approach to the reduction.

1.2 This guide includes the following topics:

| Topic Title                                     | Section Number |
|---|----------------|
| Scope   | 1              |
| Referenced Documents                            | 2              |
| Definitions                                     | 3              |
| Significance and Use                            | 4              |
| Measures of Test Variability                    | 5              |
| Unnecessary Test Variability                    | 6              |
| Identifying Probable Causes of Test Variability | 7              |
| Determining the Causes of Test Variability      | 8              |
| Averaging                                       | 9              |
| Calibration                                     | 10             |
| Keywords  | 11             |

1.3 The annexes include:

| Topic Title                                    | Annex Number |
|--|--------------|
| Statistical Test Selection                     | Annex A1     |
| Frequency Distribution Identification          | Annex A2     |
| Design of Ruggedness Tests                     | Annex A3     |
| Ruggedness Test Analysis:                      |              |
| Unknown or Undefined Distribution—Small Sample | Annex A4     |
| Unknown or Undefined Distribution—Large Sample | Annex A5     |
| Binomial Distribution                          | Annex A6     |
| Poisson Distribution                           | Annex A7     |
| Normal Distribution                            | Annex A8     |
| Design of a Randomized Block Experiment        | Annex A9     |
| Randomized Block Experiment Analysis:          |              |
| Unknown or Undefined Distribution—Small Sample | Annex A10    |
| Unknown or Undefined Distribution—Large Sample | Annex A11    |
| Binomial Distribution                          | Annex A12    |
| Poisson Distribution                           | Annex A13    |
| Normal Distribution                            | Annex A14    |
| Averaging:                                     |              |
| No Compositing                                 | Annex A15    |
| Compositing                                    | Annex A16    |

## 2. Referenced Documents

### 2.1 ASTM Standards:

D 123 Terminology Relating to Textiles<sup>2</sup>

<sup>1</sup> This guide is under the jurisdiction of ASTM Committee D13 on Textiles and is the direct responsibility of Subcommittee D13.93 on Statistics.

Current edition approved Sept. 10, 1997. Published August 1998. Originally published as D4853 – 88. Last previous edition D4853 – 91.

<sup>2</sup> Annual Book of ASTM Standards, Vol 07.01.

D 1907 Test Method for Yarn Number by the Skein Method<sup>2</sup>

D 2256 Test Method for Tensile Properties of Yarns by the Single-Strand Method<sup>2</sup>

D 2904 Practice for Interlaboratory Testing of a Textile Test Method that Produces Normally Distributed Data<sup>2</sup>

D 2906 Practice for Statements on Precision and Bias for Textiles<sup>2</sup>

D 3512 Test Method for Pilling Resistance and Other Related Surface Changes of Textile Fabrics: Random Tumble Pilling Tester Method<sup>3</sup>

D 3659 Test Method for Flammability of Apparel Fabrics by Semi-Restraint Method<sup>3</sup>

D 4356 Practice for Establishing Consistent Test Method Tolerances<sup>3,4</sup>

D 4467 Practice for Interlaboratory Testing of a Test Method that Produces Non-Normally Distributed Data<sup>3</sup>

D 4686 Guide for Identification of Frequency Distributions<sup>3</sup>

D 4854 Guide for Estimating the Magnitude of Variability from Expected Sources in Sampling Plans<sup>3</sup>

E 456 Terminology Relating to Quality and Statistics<sup>4</sup>

E 1169 Guide for Conducting Ruggedness Tests<sup>4</sup>

### 2.2 ASTM Adjuncts:

TEX-PAC<sup>5</sup>

NOTE 1—Tex-Pac is a group of PC programs on floppy disks, available through ASTM Headquarters, 100 Barr Harbor Drive, West Conshohocken, PA 19428, USA. The analysis described in Annex A4 can be conducted using one of these programs.

## 3. Terminology

### 3.1 Definitions:

3.1.1 *average, n—for a series of observations*, the total divided by the number of observations. (Syn. *arithmetic average, arithmetic mean, mean*)

3.1.2 *block, n—in experimenting*, a group of units that is relatively homogeneous within itself, but may differ from other similar groups.

3.1.3 *degrees of freedom, n—for a set*, the number of values that can be assigned arbitrarily and still get the same value for each of one or more statistics calculated from the set of data.

<sup>3</sup> Annual Book of ASTM Standards, Vol 07.02.

<sup>4</sup> Annual Book of ASTM Standards, Vol 14.02.

<sup>5</sup> PC programs on floppy disks are available through ASTM. For a 3½ inch disk request PCN:12-429040-18, for a 5½ inch disk request PCN:12-429041-18.

3.1.4 *duplicate, n*—in *experimenting or testing*, one of two or more runs with the same specified experimental or test conditions but with each experimental or test condition not being established independently of all previous runs. (Compare *replicate*)

3.1.5 *duplicate, vt*—in *experimenting or testing*, to repeat a run so as to produce a duplicate. (Compare *replicate*)

3.1.6 *error of the first kind,  $\alpha$ , n*—in a *statistical test*, the rejection of a statistical hypothesis when it is true. (Syn. *Type I error*)

3.1.7 *error of the second kind,  $\beta$ , n*—in a *statistical test*, the acceptance of a statistical hypothesis when it is false. (Syn. *Type II error*)

3.1.8 *experimental error, n*—variability attributable only to a test method itself.

3.1.9 *factor, n*—in *experimenting*, a condition or circumstance that is being investigated to determine if it has an effect upon the result of testing the property of interest.

3.1.10 *interaction, n*—the condition that exists among factors when a test result obtained at one level of a factor is dependent on the level of one or more additional factors.

3.1.11 *mean*—See *average*.

3.1.12 *median, n*—for a *series of observations*, after arranging them in order of magnitude, the value that falls in the middle when the number of observations is odd or the arithmetic mean of the two middle observations when the number of observations is even.

3.1.13 *mode, n*—the value of the variate for which the relative frequency in a series of observations reaches a local maximum.

3.1.14 *randomized block experiment, n*—a kind of experiment which compares the average of  $k$  different treatments that appear in random order in each of  $b$  blocks.

3.1.15 *replicate, n*—in *experimenting or testing*, one of two or more runs with the same specified experimental or test conditions and with each experimental or test condition being established independently of all previous runs. (Compare *duplicate*)

3.1.16 *replicate, vt*—in *experimenting or testing*, to repeat a run so as to produce a replicate. (Compare *duplicate*)

3.1.17 *ruggedness test, n*—an experiment in which environmental or test conditions are deliberately varied to evaluate the effect of such variations.

3.1.18 *run, n*—in *experimenting or testing*, a single performance or determination using one of a combination of experimental or test conditions.

3.1.19 *standard deviation,  $s$ , n*—of a *sample*, a measure of the dispersion of variates observed in a sample expressed as the positive square root of the sample variance.

3.1.20 *treatment combination, n*—in *experimenting*, one set of experimental conditions.

3.1.21 *Type I error*—See *error of the first kind*.

3.1.22 *Type II error*—See *error of the second kind*.

3.1.23 *variance,  $s^2$ , n*—of a *sample*, a measure of the dispersion of variates observed in a sample expressed as a function of the squared deviations from the sample average.

3.1.24 For definitions of textile terms, refer to Terminology D 123. For definitions of other statistical terms, refer to Terminology E 456.

#### 4. Significance and Use

4.1 This guide can be used at any point in the development or improvement of a test method, if it is desired to pursue reduction of its variability.

4.2 There are three circumstances in which a subcommittee responsible for a test method would want to reduce test variability:

4.2.1 During the development of a new test method, ruggedness testing might reveal factors which produce an unacceptable level of variability, but which can be satisfactorily controlled once the factors are identified.

4.2.2 Another is when analysis of data from an interlaboratory test of a test method shows significant differences between levels of factors or significant interactions which were not desired or expected. Such an occurrence is an indicator of lack of control which means that the precision of the test method is not predictable.

4.2.3 The third situation is when the method is in statistical control, but it is desired to improve its precision, perhaps because the precision is not good enough to detect practical differences with a reasonable number of specimens.

4.3 The techniques in this guide help to detect a statistical difference between test results. They do not directly answer questions about practical differences. A statistical difference is one which is not due to experimental error, that is, chance variation. Each statistical difference found by the use of this guide must be compared to a practical difference, the size of which is a matter of engineering judgment. For example, a change of one degree in temperature of water may show a statistically significant difference of 0.05 % in dimensional change, but 0.05 % may be of no importance in the use to which the test may be put.

#### 5. Measures of Test Variability

5.1 There are a number of measures of test variability, but this guide concerns itself with only two: one is the probability,  $p$ , that a test result will fall within a particular interval; the other is the positive square root of the variance which is called the standard deviation,  $s$ . The standard deviation is sometimes expressed as a percent of the average which is called the coefficient of variation,  $CV$  %. Test variability due to lack of statistical control is unpredictable and therefore cannot be measured.

#### 6. Unnecessary Test Variability

6.1 The following are some frequent causes of unnecessary test variability:

- 6.1.1 Inadequate instructions.
- 6.1.2 Miscalibration of instruments or standards.
- 6.1.3 Defective instruments.
- 6.1.4 Instrument differences.
- 6.1.5 Operator training.
- 6.1.6 Inattentive operator.
- 6.1.7 Reporting error.

- 6.1.8 False reporting.
- 6.1.9 Choice of measurement scale.
- 6.1.10 Measurement tolerances either unspecified or incorrect.
- 6.1.11 Inadequate specification of, or inadequate adherence to, tolerances of test method conditions. (For establishing consistent tolerances, see Practice D 4356.)
- 6.1.12 Incorrect identification of materials submitted for testing.
- 6.1.13 Damaged materials.

## 7. Identifying Probable Causes of Test Variability

7.1 Sometimes the causes of test variability will appear to be obvious. These should be investigated as probable causes, but the temptation should be avoided to ignore other possible causes.

7.2 The list contained in Section 6 should be reviewed to see if any of these items could be producing the observed test variability.

7.3 To aid in selecting the items to investigate in depth, plot frequency distributions and statistical quality control charts (1).<sup>6</sup> Make these plots for all the data and then for each level of the factors which may be causes of, or associated with, test variability.

7.4 In examining the patterns of the plots, there may be some hints about which factors are not consistent among their levels in their effect on test variability. These are the factors to pursue.

## 8. Determining the Causes of Test Variability

### 8.1 *Use of Statistical Tests:*

8.1.1 This section includes two statistical techniques to use to investigate the significance of the factors identified as directed in Section 7: ruggedness tests and randomized block experiments analyses. In using these techniques, it is advantageous to choose a model to describe the distribution from which the data come. Methods for identifying the distributions are contained in Annex A2 and Guide D 4686. For additional information about distribution identification, see Shapiro (2).

8.1.2 In order to assure being able to draw conclusions from ruggedness testing and components of variance analysis, it is essential to have sufficient available data. Not infrequently, the quantity of data is so small as to preclude significant differences being found if they exist.

### 8.2 *Ruggedness Tests:*

8.2.1 Use ruggedness testing to determine the method's sensitivity to variables which may need to be controlled to obtain an acceptable precision. Ruggedness tests are designed using only two levels of each of one or more factors being examined. For additional information see Guide E 1169.

8.2.2 Prepare a definitive statement of the type of information the task group expects to obtain from the ruggedness test. Include an example of the statistical analysis to be used, using hypothetical data.

8.2.3 Design, run, and analyze the ruggedness test as directed in Annex A3-Annex A8.

8.2.4 From a summary table obtained as directed in A3.6, the factors to which the test method is sensitive may become apparent. Some sensitivity is to be expected; it is usually desirable for a test method to detect differences between fabrics of different constructions, fiber contents, or finishes. Some sensitivities may be expected, but may be controllable; temperature is frequently such a factor.

8.2.5 If analysis shows that any test conditions have a significant effect, modify the test procedure to require the degree of control required to eliminate any significant effects.

### 8.3 *Randomized Block Experiment:*

8.3.1 When it is desired to investigate a test method's sensitivity to a factor at more than two levels, use a randomized block experiment. Such factors might be: specimen chambers within a machine, operators, shifts, or extractors. Analysis of the randomized block experiment will help to determine how much the factor levels contribute to the total variation of the test method results. Comparison of the factor level variation from factor to factor will identify the sources of large sums of squares in the total variation of the test results.

8.3.2 Prepare a definitive statement of the type of information the task group expects to obtain from the measurement of sums of squares. Include an example of the statistical analysis to be used, using hypothetical data.

8.3.3 Design, run, and analyze the randomized block experiment as directed in Annex A9-Annex A14.

8.3.4 From a summary table of results with blocks as rows and factor levels as columns, such as in A14.2.1, the levels of a factor to which the test method is sensitive may become apparent. Some sensitivity to level changes may be expected, but may be controllable; different operators are frequently such levels.

8.3.5 If the analysis shows any significant effects associated with changes in level of a factor, revise the test procedure to obviate the necessity for a level change. If this is not possible, give a warning, and explain how to minimize the effect of necessary level changes.

## 9. Averaging

9.1 *Variation*—Averages have less variation than individual measurements. The more measurements include in an average, the less its variation. Thus, the variation of test results can be reduced by averaging, but averaging will not improve the precision of a test method as measured by the variance of specimen selection and testing (Note 2).

NOTE 2—This section is applicable to all sampling plans producing variables data regardless of the kind of frequency distribution of these data because no estimations are made of any probabilities.

9.2 *Sampling Plans with No Composites*—Some test methods specify a sampling plan as part of the procedure. Selective increases or reductions in the number of lot and laboratory samples, and specimens specified can sometimes be made which will reduce test result variation and also reduce cost (Note 3). To investigate the possibility of making sampling plan revisions which will reduce variation proceed as directed in either Annex A15 or Annex A16.

<sup>6</sup> The boldface numbers in parentheses refer to the list of references at the end of this guide.

NOTE 3—The objective of sampling plan selection is to achieve acceptable variation with minimum cost. For calculation of sampling and testing costs, see A2.5 in Guide D 4854.

9.3 *Sampling Plan with Composites*—Some test methods specify compositing samples, a kind of averaging, as part of the sampling plan. Compositing is done to reduce the variation of the test results, and also to reduce the cost of testing. Composites are prepared by blending equal amounts of two or more individual sample units. Compositing cannot reduce the overall variation of a test method. Consider compositing when: (1) blending is possible, and when (2) the total of sampling variance of lot and laboratory is large compared with the specimen testing variance. It might be found that the cost of testing can be reduced and still obtain the same test variation (Notes 3 and 4). To investigate the consequences of compositing proceed as directed in Annex A16.

NOTE 4—When compositing is done, information about the variation among the sample units which were blended is lost. Compositing limits the utility of the results from the test method and reduces the quantity of data available to control the use of the test method.

9.4 *Different Types of Materials*—Sampling plan studies based on this guide are applicable only to material(s) on which the studies are made. Make separate studies on three or more kinds of materials of the type on which the test method may be used and which produce test results covering the range of interest. If similar results are not obtained, revise the sampling plans in the test method to take this into account.

## 10. Calibration

10.1 If the variability of a test method cannot be improved by use of any of the techniques previously described, biases

may be present that vary over short periods of time. The existence of such biases is discoverable by the use of the technique described in 8.3 or statistical quality control charts (1). To reduce the test variation due to such biases proceed as directed in 10.2.

10.2 Use a reference material to make a calibration each time a series of samples is tested. Adjust the sample test results in relation to the test result from the standard.

10.3 The best way to select and use a reference material is dependent on the test method of interest, but the following principles apply in all cases:

10.3.1 Prepare and test the reference material as directed in the test method.

10.3.2 Run tests on the reference material just before the samples are tested, and plot the results of the tests on statistical quality control charts (1). Adjust the test results, using only such reference material test results.

10.3.3 Ensure an adequate and homogeneous supply of the reference material.

10.3.4 Select a new supply of the reference material well in advance of depleting the supply of the old material. Test the old and the new material at the same time for approximately 20 or 30 tests before going to the new material. This practice is necessary to establish the level of the new reference material.

## 11. Keywords

11.1 developing test methods; interlaboratory testing; ruggedness test design; statistics; uniformity

## ANNEXES

### (Mandatory Information)

#### A1. STATISTICAL TEST SELECTION

A1.1 *Guide to Statistical Test Selection*—The statistical technique used to determine the significance of any differences due to different test conditions is dictated by the model chosen to describe the distribution from which the data come. The appropriate techniques are contained in the annexes. A guide to the appropriate statistical test to use in each situation is listed in Table A1.1.

**TABLE A1.1 Guide to Appropriate Statistical Tests**

| Distribution         | Sample Size        | Ruggedness Test <sup>A</sup>             | Randomized Block Experiment <sup>B</sup> |
|----------------------|--------------------|--|--|
| Unknown or Undefined | Small <sup>C</sup> | Wilcoxon Rank Sum, Annex A4              | Friedman Rank Sum, Annex A10             |
| Unknown or Undefined | Large <sup>C</sup> | Wilcoxon Rank Sum, Annex A5              | Friedman Rank Sum, Annex A11             |
| Binomial             | <sup>D</sup>       | Critical differences or z-test, Annex A6 | Friedman Rank Sum or ANOVA, Annex A12    |
| Poisson              | <sup>D</sup>       | Critical differences or z-test, Annex A7 | Friedman Rank Sum or ANOVA, Annex A13    |
| Normal               | <sup>D</sup>       | ANOVA, Annex A8                          | ANOVA, Annex A14                         |

<sup>A</sup> Use ruggedness tests for two levels per factor.

<sup>B</sup> Use randomized block experiments for more than two levels per factor.

<sup>C</sup> "Small" and "large" are defined in the applicable annexes.

<sup>D</sup> The applicable annexes specify requirements such as sample sizes.

## A2. FREQUENCY DISTRIBUTION IDENTIFICATION

A2.1 *Identification*—Use the procedures in Guide D 4686 to identify the underlying distribution.

A2.2 *Unknown or Undefined Distribution*—Sometimes raw data come from an underlying distribution which seems to produce non-normal individuals and sample averages. In some of these cases a transformation can be made so that the transformed data do behave as if they come from a normal distribution (see Shapiro (2)). The use of such transformed data

will allow statistical tests to be made, using the assumption of normality.

A2.2.1 If the data cannot be successfully transformed, there are available methods of analysis that are described in Annex A4, Annex A5, Annex A10, and Annex A11 which are free of assumptions about the type of distribution from which the data come.

## A3. DESIGN OF RUGGEDNESS TESTS

A3.1 Select the factors to be investigated, such as temperature, time or concentration. Choose an upper and a lower level for each factor. (The terms “upper” and “lower” used to designate factor levels may or may not have functional meaning, such as two different pieces of equipment.) Choose the two levels to be sufficiently different to show test method sensitivity to that factor, if such exists, but close enough to be within a range of values which can reasonably be controlled when the test method is run.

A3.2 After selecting the number of factors to be tested, calculate the number of distinct treatment combinations to investigate:

$$R = N + 1 \quad (\text{A3.1})$$

where:

$R$  = the number of distinct treatment combinations, and  
 $N$  = the number of factors.

A3.3 Determine the level of each factor in each treatment combination by generating a table of  $N$  rows and  $R$  columns, indicating lower and upper levels of each factor by zeroes (0) and ones (1) respectively. Put one in the second column (treatment combination 1) for each of the factors. Put values in each of the remaining cells in such a way that the sum,  $C$ , for each row and each column is the same, excluding the first column, and so that:

$$C = N/2 \quad \text{for } N \text{ even} \quad (\text{A3.2})$$

$$C = N/2 - 0.5 \quad \text{for } N \text{ odd} \quad (\text{A3.3})$$

where:

$C$  = column totals or row totals, exclusive of the column for treatment combination 1, and  
 $N$  = the number of factors.

A3.4 The format for this design is shown in Table A3.1. Examples are shown in Tables A4.3, A6.1, and A8.1.

A3.5 Randomize the order of all of the replications within the experiment, performing at least two replications of each treatment combination. Record the resulting data in a table having the format shown in Table A3.2.

**TABLE A3.1 Format for a Fractional Factorial Design**

| Factor | Treatment Combination |        |        |     |        | Sum |
|--------|-----------------------|--------|--------|-----|--------|-----|
|        | 1                     | 2      | 3      | ... | $R$    |     |
| A      | 1                     | 0 or 1 | 0 or 1 | ... | 0 or 1 | $C$ |
| B      | 1                     | 0 or 1 | 0 or 1 | ... | 0 or 1 | $C$ |
| C      | 1                     | 0 or 1 | 0 or 1 | ... | 0 or 1 | $C$ |
| .      | .                     | .      | .      | ... | .      | .   |
| N      | 1                     | 0 or 1 | 0 or 1 | ... | 0 or 1 | $C$ |
| Sum    | $N$                   | $C$    | $C$    | ... | $C$    |     |

**TABLE A3.2 Format for Recording Ruggedness Test Results Test Results for Test Method XXX**

| Replicate | Treatment Combination |     |     |     |     |
|-----------|-----------------------|-----|-----|-----|-----|
|           | 1                     | 2   | 3   | ... | $R$ |
| 1         | x                     | x   | x   | ... | x   |
| 2         | x                     | x   | x   | ... | x   |
| 3         | x                     | x   | x   | ... | x   |
| .         | .                     | .   | .   | ... | .   |
| .         | .                     | .   | .   | ... | .   |
| n         | x                     | x   | x   | ... | x   |
| Average   | $A$                   | $A$ | $A$ | ... | $A$ |

A3.5.1 The following are the minimum recommended sizes for ruggedness tests depending on the type of distribution chosen to model the data:

A3.5.1.1 *Unknown or Undefined Distribution*—There should be a minimum of six observations at each level of each factor. The experiment shown in Table A4.4 is a minimum size experiment because there are six observations at each factor level. Factor A is at the higher level in Treatment Combinations 1 and 4, and there are three replicates of each treatment combination. This produces six observations of Factor A at the higher level. Factor A is at the lower level in Treatment Combinations 2 and 3, and there are three replicates of each treatment combination. This produces six observations of Factor A at the lower level. Similar counts for the other two factors, give six observations at each level.

A3.5.1.2 *Binomial Distribution*—Section A6.1.1 gives criteria for experiment size when using a normal approximation

and a *z*-test. There is no generally accepted rule-of-thumb for smaller experiments. One method is for the experimenter to develop tables similar to Table A6.4 and Table A6.5 and use them to decide whether a specific experiment size will offer enough discrimination.

A3.5.1.3 *Poisson Distribution*—Section A7.1.1 gives criteria for experiment size when using a normal approximation and the *z*-test. For smaller tests the experimenter can use Table 6 in Practice D 2906 to determine the approximate number of total counts needed to give the desired discrimination.

A3.5.1.4 *Normal Distribution*—There should be a minimum of ten degrees of freedom in the estimate of experimental error.

A3.6 When using a design for investigating a normal distribution, it will not be possible to separate effects of different factor levels from interactions; it will be possible to obtain an estimate of the effect of changing a specific factor level. To do this, average the results of the replications of those treatment combinations when that factor was at the upper level, and average the results of the replications of those treatment combinations when that factor was at the lower level. Enter the two averages in a table. See Table A4.4 for an example.

### RUGGEDNESS TEST ANALYSIS

#### A4. UNKNOWN OR UNDEFINED DISTRIBUTION—SMALL SAMPLE

##### A4.1 Procedure:

A4.1.1 For data whose distribution is unknown or undefined, whose variates may be discrete or continuous, and the total number of observations is fewer than or equal to twenty, use the Wilcoxon Rank Sum Test described in Hollander and Wolf (3). Group the data for each factor by levels, and assign ranks to each observation within the factor. In the event of ties, assign the average rank of the tied observations. After assigning ranks, sum them for each level of each factor. For an example, see Table A4.1.

A4.1.2 To determine if the effects of the levels of the factors are significantly different, compare the greater rank sum for each factor with the values in Table A4.2, or use the table of probabilities for Wilcoxon's Rank Sum *W* Statistic shown in Hollander and Wolf (3).

##### A4.2 Example:

A4.2.1 An example is given to illustrate this procedure. A ruggedness test was performed on a method for pilling resistance determination (Test Method D 3512 – 82) to examine the suitability of a synthetic lining material to replace the cork liner specified for the random tumble pilling tester. For this test, some of the factors which could logically affect the test results are: material under test, number of 30-min tumbling cycles, and lining material.

TABLE A4.2 Critical Values of Wilcoxon's *W* Statistic—5 % Probability Level

| Observations in One Level of Factor | Observ. in Other Level of Factor |    |    |    |    |     |     |
|-------------------------------------|----------------------------------|----|----|----|----|-----|-----|
|                                     | 4                                | 5  | 6  | 7  | 8  | 9   | 10  |
| 2                                   | —                                | 13 | 15 | 17 | 18 | 20  | 22  |
| 3                                   | 18                               | 20 | 22 | 25 | 27 | 30  | 32  |
| 4                                   | 25                               | 28 | 31 | 34 | 37 | 40  | 43  |
| 5                                   | —                                | 36 | 40 | 44 | 47 | 51  | 55  |
| 6                                   | —                                | —  | 50 | 55 | 59 | 63  | 67  |
| 7                                   | —                                | —  | —  | 66 | 71 | 76  | 81  |
| 8                                   | —                                | —  | —  | —  | 85 | 90  | 96  |
| 9                                   | —                                | —  | —  | —  | —  | 105 | 111 |
| 10                                  | —                                | —  | —  | —  | —  | —   | 128 |

- A—Test Material
  - 1 (upper level)—double knit fabric
  - 0 (lower level)—satin fabric
- B—Number of Tumbling Cycles
  - 1 (upper level)—one cycle
  - 0 (lower level)—two cycles
- C—Lining Material
  - 1 (upper level)—synthetic liner
  - 0 (lower level)—cork liner

A4.2.2 Using Eq A3.1, the number of distinct treatment combinations for this example is:  $R = 3 + 1 = 4$ .

A4.2.3 Since there are three factors, *N* is odd, so Eq A3.3 is used to determine the column and row totals:  $C = 3/2 - 0.5 = 1$ .

TABLE A4.1 Data Arrangement for Wilcoxon Rank Sum Test

| A—Material |      |         |      | B—No. of Cycles |      |         |      | C—Liner Material |      |         |      |
|------------|------|---------|------|-----------------|------|---------|------|------------------|------|---------|------|
| Level 1    |      | Level 0 |      | Level 1         |      | Level 0 |      | Level 1          |      | Level 0 |      |
| Ob.        | Rank | Ob.     | Rank | Ob.             | Rank | Ob.     | Rank | Ob.              | Rank | Ob.     | Rank |
| 4.0        | 11   | 2.0     | 5    | 4.0             | 11   | 1.0     | 2    | 4.0              | 11   | 2.0     | 5    |
| 3.7        | 10   | 2.0     | 5    | 3.7             | 10   | 1.0     | 2    | 3.7              | 10   | 2.0     | 5    |
| 4.3        | 12   | 2.0     | 5    | 4.3             | 12   | 1.0     | 2    | 4.3              | 12   | 2.0     | 5    |
| 2.5        | 8    | 1.0     | 2    | 2.0             | 5    | 2.5     | 8    | 1.0              | 2    | 2.5     | 8    |
| 2.5        | 8    | 1.0     | 2    | 2.0             | 5    | 2.5     | 8    | 1.0              | 2    | 2.5     | 8    |
| 2.5        | 8    | 1.0     | 2    | 2.0             | 5    | 2.5     | 8    | 1.0              | 2    | 2.5     | 8    |
| Sum        | 57   |         | 21   |                 | 48   |         | 30   |                  | 39   |         | 39   |

A4.2.4 The corresponding design table is shown as Table A4.3.

A4.2.5 The four treatment combinations shown in Table A4.3 were used, with three replications of each treatment combination. The observations and their averages for each specific treatment combination are listed in Table A4.4. The replications were conducted in a randomized order over the entire experiment. A summary of results is shown in Table A4.5.

A4.2.6 Table A4.1 shows the data of Table A4.4 arranged by both levels of each factor with their ranks within each factor. The statistical significance of the difference of the levels in each factor is determined by use of the Wilcoxon Rank Sum Test (A4.1 and A4.2). Use the critical values shown either in (1) Table A4.2 or in (2) Hollander and Wolf (3). The procedures for doing this are as follows:

A4.2.6.1 *Evaluation Using Table A4.2*—If the two levels of a factor have the same effect on the test results for this example, then a rank sum of 39 would be expected. (Thirty nine is one half of the sum of the ranks one through twelve. Twelve is the total number of observations on each factor; six observations at each of the two levels.) Table A4.2 shows that for six observations at each of the two levels of a factor, a rank sum equal to or greater than fifty is significantly different from 39 at the 95 % probability level. The greater rank sum for the material tested was the only one which equalled or exceeded fifty, and it is therefore concluded that the test method is sensitive to the type of material, but not to the number of tumble cycles or type of lining material examined in this ruggedness test.

A4.2.6.2 *Evaluation Using Hollander and Wolf*—Using pp. 68–69 of Hollander and Wolf (3), compare the greater rank sum for each factor with the values in the table of probabilities for Wilcoxon’s Rank Sum *W* statistics as directed in A4.1. The

**TABLE A4.3 Pilling Resistance Determination—Ruggedness Test Design**

| Factor             | Treatment Combination |   |   |   |
|--------------------|-----------------------|---|---|---|
|                    | 1                     | 2 | 3 | 4 |
| A—Material         | 1                     | 0 | 0 | 1 |
| B—Number of Cycles | 1                     | 1 | 0 | 0 |
| C—Liner Material   | 1                     | 0 | 1 | 0 |

**TABLE A4.4 Pilling Resistance Rating**

| Replicate | Treatment Combination |     |     |     |
|-----------|-----------------------|-----|-----|-----|
|           | 1                     | 2   | 3   | 4   |
| 1         | 4.0                   | 2.0 | 1.0 | 2.5 |
| 2         | 3.7                   | 2.0 | 1.0 | 2.5 |
| 3         | 4.3                   | 2.0 | 1.0 | 2.5 |
| Average   | 4.0                   | 2.0 | 1.0 | 2.5 |

**TABLE A4.5 Summary of Pilling Resistance Determinations**

| Factor             | Average for |             | Difference Between Averages |
|--------------------|-------------|-------------|-----------------------------|
|                    | Upper Level | Lower Level |                             |
| A—Material         | 3.2         | 1.5         | 1.7                         |
| B—Number of Cycles | 3.0         | 1.8         | 1.2                         |
| C—Liner Material   | 2.5         | 2.2         | 0.3                         |

results of referring to the table for this example are contained in Table A4.6. The type of material tested was the only significant factor, since the probability of obtaining such a high rank sum of 57 or greater is only 0.001. If there was no difference between the two materials, this is a very unlikely occurrence (see Note A4.1). Thus it is concluded that the two materials have different pilling resistances. The probability of obtaining a rank sum of 48 or greater as was done for number of cycles is 0.090; therefore the effects of the different number of tumble cycles are not significant. The probability of a rank sum of 39 or greater occurring, as it did for lining material, is 0.531, and 39 happens to equal the expected rank sum; therefore the effects of the two lining materials are not significantly different.

NOTE A4.1—Throughout this guide, any probability of occurrence equal to or greater than 0.05 is considered large enough to conclude that no significant difference exists between or among estimates being compared.

**TABLE A4.6 Probabilities for Rank Sums**

| Factor             | Greater Rank Sum | Probability <sup>A</sup> |
|--------------------|------------------|--------------------------|
| A—Material         | 57               | 0.001                    |
| B—Number of Cycles | 48               | 0.090                    |
| C—Liner Material   | 39               | 0.531                    |

<sup>A</sup> Probability of occurrence of a rank sum this large or larger. Calculated as directed in A4.2.6.2.

## A5. UNKNOWN OR UNDEFINED DISTRIBUTION—LARGE SAMPLE

A5.1 If there are more than ten replications of each factor level, use the large sample approximation as given on pp. 68–69 of Hollander and Wolf (3).

A6. BINOMIAL DISTRIBUTION

A6.1 Procedure:

A6.1.1 Calculation of Significance—According to McClave and Dietrich (4), if the interval  $p \pm 3\sigma$ , where  $\sigma$  is defined as in A6.1.3, does not contain zero or one, then the number of observations is sufficient for assuming that a particular binomial distribution can be approximated by a normal distribution. If this requirement is met for each factor level, then perform a  $z$ -test as directed in A6.1.4 and A6.1.5. Otherwise calculate critical differences as directed in A6.1.6.

A6.1.2 Calculation of  $p$ —Calculate the proportion of specimens at each level of each factor that obtained a given result:

$$p = x/n \tag{A6.1}$$

where:

$x$  = number of specimens that had a specified attribute at a specific level of a factor, and

$n$  = number of specimens tested for a level of a factor.

A6.1.3 Calculation of  $s$ —Calculate the sample standard deviation, an estimate of  $\sigma$ , for each  $p$  as follows:

$$s = [p(1 - p)/n]^{1/2} \tag{A6.2}$$

A6.1.4 Calculation of Variance of Difference—Calculate the sample variance of the difference of the two  $p$ 's for the two levels of each factor:

$$s_d^2 = p_U(1 - p_U)/n_U + p_L(1 - p_L)/n_L \tag{A6.3}$$

where:

$s_d^2$  = the sample variance of the difference of the  $p$ 's,

$p_U$  = the proportion at the upper level,

$p_L$  = the proportion at the lower level,

$n_U$  = number of observations at the upper level, and

$n_L$  = number of observations at the lower level.

A6.1.5  $z$ -Test—Calculate the difference of the two proportions in sample standard deviation units as follows:

$$z = [(p_U - p_L)/(s_d^2)]^{1/2} \tag{A6.4}$$

If  $-1.96 \leq z \leq 1.96$ , then conclude that the effect of the factor level change is not significant. The value of  $z$  and its associated probability (two sided) of 5 % is found in a table of areas under the normal distribution curve.

A6.1.6 Critical Difference—For those data whose distribution cannot be approximated by a normal distribution, prepare a table of critical differences between two levels, using an exact test of significance for 2-by-2 contingency tables containing small frequencies (see section 9.1.1 and section 9.1.2 of Practice D 2906), using published tables (5), or using an algorithm for use with a computer.

A6.2 Example:

A6.2.1 Following is a ruggedness test of three factors in fabric flammability testing (Test Method D 3659). The data are notations of passing or failing based on an arbitrary specification. For this reason, it is assumed that these data may be modeled by a binomial distribution. Three factors which

should logically affect flammability behavior are: finish on fabric, conditioning prior to ignition, and ignition time (the standard three seconds or forced). Since the fabric tested was polyester batiste, and Test Method D 3659 is a semi-restraint method, it was possible that ignition could not be forced due to the fabric curling away from the flame.

A6.2.2 The following levels were assigned to each of the three factors:

- A—Finish
  - 1 (upper level)—Flame Retardant Treated
  - 0 (lower level)—Scoured; no finish applied
- B—Conditioning
  - 1 (upper level)—Conditioned as in Practice D 1776
  - 0 (lower level)—Oven dried and desiccated
- C—Ignition Time
  - 1 (upper level)—Three seconds
  - 0 (lower level)—Forced ignition

A6.2.3 The design for this test was developed as directed in Annex A3, and is shown in Table A6.1.

A6.2.4 Table A6.2 shows the results of the tests for each replicate of each treatment combination. Table A6.3 shows the data organized by levels within each of the three factors. The data are summarized on the last line of Table A6.2, the results being expressed as the fraction passing by level within each factor.

A6.2.5 Following are the sample standard deviations for each factor-level combination, calculated as directed in A6.1:

| Factor | Level | $p$  | $s$  |
|--------|-------|------|------|
| A      | 1     | 0.83 | 0.15 |
|        | 0     | 0.50 | 0.20 |
| B      | 1     | 0.67 | 0.19 |
|        | 0     | 0.67 | 0.19 |
| C      | 1     | 0.83 | 0.15 |
|        | 0     | 0.50 | 0.20 |

A6.2.6 Since in each case the interval  $p \pm 3s$  includes either zero or one, the approximation to the normal distribution cannot be used (4).

A6.2.7 Calculate the critical differences as directed in A6.1.6. Put the results in a table as shown for this example in Table A6.4. This shows that no factor had a significant effect at the 95 % probability level.

A6.2.8 This technique can also be used to determine the minimum number of replicates from which conclusions can be drawn. The results of such calculations show that, if fewer than four replicates are run for each level of each factor, then there are not enough data. Even with four replicates per level of a factor, one level must produce all successes and the other all failures in order to be able to say that a factor has a significant effect (see Table A6.5).

TABLE A6.1 Flammability Ruggedness Test Design

| Factor          | Treatment Combination |   |   |   |
|-----------------|-----------------------|---|---|---|
|                 | 1                     | 2 | 3 | 4 |
| A—Finish        | 1                     | 0 | 0 | 1 |
| B—Conditioning  | 1                     | 1 | 0 | 0 |
| C—Ignition Time | 1                     | 0 | 1 | 0 |





**TABLE A6.2 Flammability Test Results**

| Replicate             | Treatment Combination |      |      |      |
|-----------------------|-----------------------|------|------|------|
|                       | 1                     | 2    | 3    | 4    |
| 1                     | pass                  | pass | fail | fail |
| 2                     | pass                  | fail | pass | pass |
| 3                     | pass                  | fail | pass | pass |
| Fraction Passing, $p$ | 1.00                  | 0.33 | 0.67 | 0.67 |

**TABLE A6.3 Flammability Results by Levels Within Factors**

| A—Finish   |      | B—Conditioning |      | C—Ignition Time |      |
|------------|------|----------------|------|-----------------|------|
| Level      |      | Level          |      | Level           |      |
| 1          | 0    | 1              | 0    | 1               | 0    |
| pass       | pass | pass           | fail | pass            | pass |
| pass       | fail | pass           | pass | pass            | fail |
| pass       | fail | pass           | pass | pass            | fail |
| fail       | fail | pass           | fail | fail            | fail |
| pass       | pass | fail           | pass | pass            | pass |
| pass       | pass | fail           | pass | pass            | pass |
| $p^A$ 0.83 | 0.50 | 0.67           | 0.67 | 0.83            | 0.50 |

<sup>A</sup>  $p$  is the fraction passing.

**TABLE A6.4 Significantly Different Numbers or Proportions of Successes (Failures) in Two Sets of Six Specimens—5 % Probability Level<sup>A</sup>**

| Successes in One Set of Six Specimens |            | Successes in Another Set of Six Specimens |                 |
|---------------------------------------|------------|---|-----------------|
| Number                                | Proportion | Number                                    | Proportion      |
| 0                                     | 0.00       | 5 or more                                 | 0.83 or greater |
| 1                                     | 0.17       | 6   | 1.00            |
| 2                                     | 0.33       | —   | —               |
| 3                                     | 0.50       | —   | —               |
| 4                                     | 0.67       | —   | —               |
| 5                                     | 0.83       | 0   | 0.00            |
| 6                                     | 1.00       | 1 or fewer                                | 0.17 or less    |

<sup>A</sup> For two-sided tests. Successes in one set of specimens are compared with successes in the other. Failures are compared with failures. See A6.1.6.

**TABLE A6.5 Significantly Different Numbers or Proportions of Successes (Failures) in Two Sets of Four Specimens—5 % Probability Level<sup>A</sup>**

| Successes in One Set of Four Specimens |            | Successes in Another Set of Four Specimens |            |
|--|------------|--|------------|
| Number                                 | Proportion | Number                                     | Proportion |
| 0                                      | 0.00       | 4  | 1.00       |
| 1                                      | 0.25       | —  | —          |
| 2                                      | 0.50       | —  | —          |
| 3                                      | 0.75       | —  | —          |
| 4                                      | 1.00       | 0  | 0.00       |

<sup>A</sup> For two-sided tests. Successes in one set of specimens are compared with successes in the other. Failures are compared with failures. See A6.1.6.

**A7. POISSON DISTRIBUTION**

**A7.1 Procedure:**

**A7.1.1 Calculation of Significance**—According to McClave and Dietrich (4), if the average  $c$  of a Poisson distribution (see A7.1.2) is equal to or greater than nine, then that particular

distribution can be approximated by a normal distribution. If this requirement is met for each factor level, then perform a  $z$ -test as directed in A7.1.4. Otherwise calculate critical differences as directed in A7.1.5.

A7.1.2 *Calculation of  $\bar{c}$* —For each factor level calculate the average number of times the event of interest occurs in the particular units observed:  $\bar{c} = x/n$  where  $x$  is the total number of times the event occurs in the observation of  $n$  units in each factor level.

A7.1.3 *Calculation of Difference Variance*—Calculate the variance of the difference of the two  $\bar{c}$ 's for the two levels of each factor as follows:

$$s_d^2 = \bar{c}_U + \bar{c}_L \quad (\text{A7.1})$$

where:

$s_d^2$  = the sample variance of the difference of the  $\bar{c}$ 's,

$\bar{c}_U$  = average number of occurrences per unit at the upper level, and

$\bar{c}_L$  = average number of occurrences per unit at the lower level.

A7.1.4 *z-test*—Calculate the difference between the two average number of occurrences per unit at each of the two levels for each factor in standard deviation units as follows:

$$z = (\bar{c}_U - \bar{c}_L) / (s_d^2)^{1/2} \quad (\text{A7.2})$$

If  $-1.96 \leq z \leq 1.96$ , conclude that the effect of the factor level change is not significant. The value of  $z$  and its associated probability (two-sided) of 5 % is found in a table of areas under the normal curve.

A7.1.5 *Critical Difference*—For those data whose distribution cannot be approximated by a normal distribution, prepare a table of critical differences of counts between the two levels of each factor, using existing tables (5), the methods specified in 10.1.1, and 10.1.2 of Practice D 2906, or an algorithm for use with a computer.<sup>6</sup>

## A7.2 Example:

A7.2.1 Following is a ruggedness test of spinnability of two cotton blends. The data evaluated are ends-down counts from one shift from one of each type of spinning frame by each operator. The three factors included in the evaluation are type frame, operators, and blends of different average staple lengths.

A7.2.2 The following levels were assigned to each of the three factors:

- A—Type Spinning Frame
  - 1 (upper level)—Ring
  - 0 (lower level)—Open-End
- B—Operator
  - 1 (upper level)—Susie Smith
  - 0 (lower level)—Betty Jones
- C—Blend
  - 1 (upper level)—1 in.
  - 0 (lower level)—<sup>30</sup>/<sub>32</sub> in.

A7.2.3 The design for this test was done as directed in Annex A2.

A7.2.4 Table A7.2 shows the results of the tests for each replicate of each treatment combination. Table A7.1 shows the data organized by levels within each of the three factors. The data are summarized in Table A7.3.

A7.2.5 Since the data are a count of the number of things per unit, it is assumed that these data may be modeled by a Poisson distribution. Since each of the factor levels has an average of nine or larger (see A7.1.1), the distribution from which they come may be approximated by a normal distribution. For these reasons, the  $z$ -test is used to evaluate the significance of the differences of the factor levels.

A7.2.6 Column 5 of Table A7.3, obtained by using Eq A7.1 and A7.2, shows the values of  $|z|$  for the differences between the averages of the levels of the three factors. For factor A,  $s_d^2 = 12.38/8 + 15.00/8 = 3.42$ , and  $z = 2.62/1.85 = 1.42$ .

A7.2.7 Each of the absolute values of  $z$  is less than 1.96. Thus it is concluded that none of the three factors has a significant effect on the test result.

**TABLE A7.1 Number of Ends Down per Shift by Factor Levels**

| Factor A |     | Factor B |     | Factor C |    |
|----------|-----|----------|-----|----------|----|
| Level    |     | Level    |     | Level    |    |
| 1        | 0   | 1        | 0   | 1        | 0  |
| 10       | 10  | 10       | 21  | 10       | 10 |
| 11       | 12  | 11       | 14  | 11       | 12 |
| 18       | 15  | 18       | 24  | 18       | 15 |
| 11       | 11  | 11       | 13  | 11       | 11 |
| 15       | 21  | 10       | 15  | 21       | 15 |
| 18       | 14  | 12       | 18  | 14       | 18 |
| 11       | 24  | 15       | 11  | 24       | 11 |
| 5        | 13  | 11       | 5   | 13       | 5  |
| Totals:  |     |          |     |          |    |
| 99       | 120 | 98       | 121 | 122      | 97 |

**TABLE A7.2 Number of Ends Down per Shift**

| Replicate | Treatment Combination |    |    |    |
|-----------|-----------------------|----|----|----|
|           | 1                     | 2  | 3  | 4  |
| 1         | 10                    | 10 | 21 | 15 |
| 2         | 11                    | 12 | 14 | 18 |
| 3         | 18                    | 15 | 24 | 11 |
| 4         | 11                    | 11 | 13 | 5  |

**TABLE A7.3 Summary of Ends Down per Shift**

| Factor | Average Level 1 | Average Level 0 | Diff. | z    |
|--------|-----------------|-----------------|-------|------|
| A      | 12.38           | 15.00           | 2.62  | 1.42 |
| B      | 12.25           | 15.12           | 2.87  | 1.55 |
| C      | 15.25           | 12.12           | 3.13  | 1.69 |

## A8. NORMAL DISTRIBUTION

### A8.1 Procedure:

A8.1.1 *Calculation of Sample Variance*—Calculate the sample variance of the replicates within each treatment combination (Note A8.1):

$$s_R^2 = [\sum x^2 - (\sum x)^2/n]/(n - 1) \quad (\text{A8.1})$$

where:

- $s_R^2$  = variance of replicates made in a treatment combination,
- $x$  = test result for each replicate in a treatment combination, and
- $n$  = number of replicates made in a treatment combination.

NOTE A8.1—Eq A8.1-A8.4 apply only to the design as described in Annex A3.

A8.1.2 Calculate the pooled error variance of the test as follows:

$$s_p^2 = [\sum (n_i - 1)s_i^2]/(r - R) \quad (\text{A8.2})$$

where:

- $s_p^2$  = pooled error variance of the test,
- $\sum$  = summation is from 1 to  $R$  with respect to  $i$ ,
- $n_i$  = the number of replications in the  $i$ th replicate,
- $s_i^2$  = variance of the replications of the  $i$ th replicate,
- $r$  = total number of replications in the test that is, the sum of the replications for all treatment combinations, and
- $R$  = number of runs.

A8.1.3 *Calculation of Critical Differences*—To evaluate the difference between levels of a factor, calculate the critical value of the difference of the averages for the two levels of a factor as follows:

$$s_d^2 = s_p^2(1/n_U + 1/n_L) \quad (\text{A8.3})$$

$$CD = 1.414 t \sqrt{v_d} \quad (\text{A8.4})$$

where:

- $s_d^2$  = the variance of the difference of the averages for the two levels of a factor,
- $n_U$  = the number of replications included in the average for the upper level,
- $n_L$  = the number of replications included in the average for the lower level,
- $CD$  = the critical value of the difference of the two levels of the factor, and
- 1.414 = square root of 2, a constant that converts the standard error of an average to the standard error of the difference between two such averages.
- $t$  = value of Student's  $t$  for two-sided limits, for  $n_U + n_L - 2$  degrees of freedom and a selected probability level.

A8.1.4 *Draw Conclusion*—The critical difference is the maximum expected value for a difference between two observed averages at a specified probability level when the two samples are from the same distribution. Therefore, if the observed difference between the averages of the upper and lower levels of a factor exceeds  $CD$ , then conclude that the test method is sensitive to that factor.

### A8.2 Example:

A8.2.1 The following is an example of a ruggedness test which produced numbers (Test Method D 1907) which may be modeled by a normal distribution.

A8.2.2 Four factors which can influence the outcome of yarn number determinations are: conditioning time, reeling, yarn tensioning device, and skeining.

A8.2.3 For this ruggedness test a 7/1 ring spun cotton yarn was chosen. The following levels were assigned to each of the four factors:

- A—Conditioning Time
  - 1 (upper level)—Four hours
  - 0 (lower level)—One hour
- B—Reel
  - 1 (upper level)—Motor driven
  - 0 (lower level)—Hand driven
- C—Yarn Tensioning Device
  - 1 (upper level)—Post and disks

0 (lower level)—Ball  
 D—Skeining  
 1 (upper level)—Before preconditioning  
 0 (lower level)—After conditioning

A8.2.4 The design of this test was done as directed in Annex A3, and is shown in Table A8.1. Note that in this instance the number of observations at the upper and lower levels of each factor are not equal.

TABLE A8.1 Yarn Number Ruggedness Test Design

| Factor                   | Treatment Combination |   |   |   |   |
|--------------------------|-----------------------|---|---|---|---|
|                          | 1                     | 2 | 3 | 4 | 5 |
| A—Conditioning Time      | 1                     | 1 | 1 | 0 | 0 |
| B—Reel                   | 1                     | 0 | 1 | 1 | 0 |
| C—Yarn Tensioning Device | 1                     | 0 | 0 | 1 | 1 |
| D—Skeining               | 1                     | 1 | 0 | 0 | 1 |

TABLE A8.2 Yarn Number Test Results—Test Method D 1907

| Replicate | Treatment Combination |      |      |      |      |
|-----------|-----------------------|------|------|------|------|
|           | 1                     | 2    | 3    | 4    | 5    |
| 1         | 7.1                   | 7.1  | 6.9  | 6.9  | 7.5  |
| 2         | 7.3                   | 7.0  | 6.9  | 6.4  | 7.4  |
| 3         | 7.0                   | 6.6  | 6.8  | 6.7  | 7.4  |
| Average   | 7.1                   | 6.9  | 6.9  | 6.8  | 7.4  |
| $s_R^2$   | 0.02                  | 0.07 | 0.00 | 0.07 | 0.00 |

A8.2.5 Table A8.2 shows the results of the tests for each replicate of each treatment combination. Table A8.3 shows the data organized by level within each of the four factors. A summary of results is shown in Table A8.4.

A8.2.6 Calculate the sample variance of the replicates within each of the five treatment combinations using Eq A8.1. The results are shown in the last row of Table A8.2. The pooled error variance,  $s_p^2 = 0.0338$  from Eq A8.2.

A8.2.7 Using Eq A8.3 and A8.4, and  $t = 2.228$ , calculate the variance of the difference of the averages for the two levels

TABLE A8.3 Yarn Number Results by Levels Within Factors

| A—Conditioning |     | B—Reel |     | C—Yarn Tension |     | D—Skeining |     |
|----------------|-----|--------|-----|----------------|-----|------------|-----|
| Level          |     | Level  |     | Level          |     | Level      |     |
| 1              | 0   | 1      | 0   | 1              | 0   | 1          | 0   |
| 7.1            | 6.9 | 7.1    | 7.1 | 7.1            | 7.1 | 7.1        | 6.9 |
| 7.3            | 6.4 | 7.3    | 7.0 | 7.3            | 7.0 | 7.3        | 6.9 |
| 7.0            | 6.7 | 7.0    | 6.6 | 7.0            | 6.6 | 7.0        | 6.8 |
| 7.1            | 7.5 | 6.9    | 7.5 | 6.9            | 6.9 | 7.1        | 6.9 |
| 7.0            | 7.4 | 6.9    | 7.4 | 6.4            | 6.9 | 7.0        | 6.4 |
| 6.6            | 7.4 | 6.8    | 7.4 | 6.7            | 6.8 | 6.6        | 6.7 |
| 6.9            |     | 6.9    |     | 7.5            |     | 7.5        |     |
| 6.9            |     | 6.4    |     | 7.4            |     | 7.4        |     |
| 6.8            |     | 6.7    |     | 7.4            |     | 7.4        |     |
| Avg. 7.0       | 7.0 | 6.9    | 7.2 | 7.1            | 6.9 | 7.2        | 6.8 |

TABLE A8.4 Summary of Yarn Number Determinations

| Factor                   | Average for: |             | Difference Between Averages |
|--------------------------|--------------|-------------|-----------------------------|
|                          | Upper Level  | Lower Level |                             |
| A—Conditioning Time      | 7.0          | 7.0         | 0.0                         |
| B—Reel                   | 6.9          | 7.2         | -0.3                        |
| C—Yarn Tensioning Device | 7.1          | 6.9         | 0.2                         |
| D—Skeining               | 7.2          | 6.8         | 0.4                         |

of each factor and the  $CD$  of the two levels of each factor. For this example,  $CD = 0.22$  yarn number. Compare the calculated value of  $CD$  with the absolute value of the observed differences between levels of each factor shown in Table A8.4.

A8.2.8 Conclude that the test procedure is not sensitive to either the change in conditioning time or the use of the two different yarn tensioning devices. This means that the shorter conditioning time may be used, and that either of the two tensioning devices may be used. Motor driven and hand driven reeling do give different results as do reeling before preconditioning and after conditioning. Thus either motor or hand driven reeling should be specified, and the order of reeling and conditioning should be specified.

A9. DESIGN OF A RANDOMIZED BLOCK EXPERIMENT

A9.1 Factor Level and Block Selection—Select the factor and the several levels,  $k$ , to be investigated, such as five operators, three shifts, or six bags (used to contain fiber when being scoured). Choose a blocking factor. Such things as days, machines, or operators may be used as blocks. The number of blocks is  $n$ ; provide at least two blocks.

A9.1.1 The general format for recording data is shown in Table A9.1.

A9.2 Randomization and Performance—Randomize the assignment of levels to blocks and the order of running the levels of the factor in each block. Always keep the same number of levels in each block. Record the results in a table having the format shown in Table A9.1.

A9.3 Degrees of Freedom—Table A14.1 shows how to calculate the number of degrees of freedom with which the mean squares for levels, blocks, and error are estimated.

TABLE A9.1 Format for Recording Data from Randomized Block Experiments

| Block | Factor Level |     |     |     |     | Sum       |
|-------|--------------|-----|-----|-----|-----|-----------|
|       | A            | B   | C   | ... | n   |           |
| 1     | x            | x   | x   | ... | x   | Sum       |
| 2     | x            | x   | x   | ... | x   | Sum       |
| 3     | x            | x   | x   | ... | x   | Sum       |
| .     | .            | .   | .   | ... | .   | .         |
| .     | .            | .   | .   | ... | .   | .         |
| m     | x            | x   | x   | ... | x   | Sum       |
|       | Sum          | Sum | Sum | ... | Sum | Grand Sum |

Considering costs and time, design the test so that there are at least ten degrees of freedom in the estimate of the mean square for error.

A9.3.1 Only when the analysis of the data is made using the  $F$ -test will the degrees of freedom for error variance be used.

Even if an *F*-test is not to be run, the rule of thumb for at least ten degrees of freedom given above is the proper one to use in designing the experiment.

**A9.4 Interaction**—When using this experimental design it is impossible to test for the significance of interaction between

factor levels and blocks. If replications are made of one or more treatment combinations within blocks, then it will be possible to separate factor level-by-block interaction and experimental error variances (see Practice D 2904 and Practice D 4467).

## RANDOMIZED BLOCK EXPERIMENT ANALYSIS

### A10. UNKNOWN OR UNDEFINED DISTRIBUTION—SMALL SAMPLE

#### A10.1 Procedure:

**A10.1.1** For data whose distribution is unknown or undefined, whose variates may be discrete or continuous, use the Friedman Rank Sum Test, described in Hollander and Wolf (3). Use the small sample test, if the number of blocks and factor levels is within the range of Table A10.1. Arrange the data in a table whose rows are blocks and whose columns are factor levels. Assign ranks to each observation within a block. In the event of ties, assign the average rank of the tied observations. For an example, see Table A10.2.

**A10.1.2** To determine if the levels of the factors are significantly different, calculate Friedman’s *S* statistic, using Eq A10.1, and compare it with the entry in Table A10.1.

$$S = (12n\sum T^2 / nk(k + 1) - 3n(k + 1)) \quad (A10.1)$$

where:

- S* = Friedman’s *S* statistic,
- T* = sum of the ranks for each factor level,
- n* = number of blocks, and
- k* = number of factor levels.

**TABLE A10.1 Critical Values of Friedman’s *S* Statistic—5% Probability Level (2)**

NOTE 1—*n* = number of blocks, and *k* = number of factor levels.

| <i>n</i> | <i>k</i> |     |     |
|----------|----------|-----|-----|
|          | 3        | 4   | 5   |
| 2        | —        | 6.0 | —   |
| 3        | 6.0      | 7.4 | 8.5 |
| 4        | 6.5      | 7.8 | 8.8 |
| 5        | 6.4      | 7.8 | 8.9 |
| 6        | 7.0      | 7.6 | —   |
| 7        | 7.1      | 7.8 | —   |
| 8        | 6.2      | 7.6 | —   |
| 9        | 6.2      | —   | —   |
| 10       | 6.2      | —   | —   |
| 11       | 6.5      | —   | —   |
| 12       | 6.5      | —   | —   |
| 13       | 6.6      | —   | —   |

**TABLE A10.2 Pilling Test Results and Their Rankings Within Blocks (Fabric Type)**

| Fabric       | Operator |    |     |   |     |   |     |   |
|--------------|----------|----|-----|---|-----|---|-----|---|
|              | 1        |    | 2   |   | 3   |   | 4   |   |
|              | Ob.      | R  | Ob. | R | Ob. | R | Ob. | R |
| Polyester    | 4.0      | 4  | 2.7 | 2 | 3.0 | 3 | 2.5 | 1 |
| Satin        | 3.0      | 4  | 2.0 | 2 | 1.0 | 1 | 2.5 | 3 |
| Wool/Acrylic | 4.3      | 4  | 2.0 | 2 | 1.0 | 1 | 2.5 | 3 |
| Sum          | 11.3     | 12 | 6.7 | 6 | 5.0 | 5 | 7.5 | 7 |

**A10.1.2.1** Hollander and Wolf (3) give instructions for adjusting *S* in the event that there are ties in some ranks. Adjusting *S* for ties changes *S* by no more than about 5%. For this reason, it is usually not necessary to make this adjustment unless it is desired to calculate the probability that an error of the first kind will be made in drawing a conclusion, or unless *S* is within about 5% of its critical value.

#### A10.2 Example:

**A10.2.1** An example is given to illustrate this procedure. A components of variance analysis was made on a method for pilling resistance determination (Test Method D 3512) to examine the influence of operators on the results of the test. The experiment included four operators (factor levels) and three fabrics (blocks). The three fabrics were a double knit polyester, a satin, and a knit wool/acrylic blend. The operator-fabric pairs were run all in the same day, but in a random order.

**A10.2.2** Table A10.2 shows the resulting observations and their rankings within blocks. Using Eq A10.1, *S* = 5.8. From Table A10.1, for *n* = 3 and *k* = 4, the critical value of *S* is 7.4. Since *S* is less than its critical value, this leads to the tentative conclusion that there is no significant difference among operators. The conclusion is tentative, because the size of the experiment does not give ten or more degrees of freedom when calculated as directed in Table A14.1.

### A11. UNKNOWN OR UNDEFINED DISTRIBUTION—LARGE SAMPLE

A11.1 *Procedure*—If the number of factor levels and blocks are such that the critical value of Friedman’s  $S$  is beyond the range of Table A8.1, then proceed as directed in

A10.1, but use the large sample approximation given on pp. 68–69 in Hollander and Wolf (3).

### A12. BINOMIAL DISTRIBUTION

A12.1 *Procedure*—For a randomized block experiment, limited to that described in Annex A9, and producing data that can be modeled by a binomial distribution, transform the data, using the appropriate transformations found in (6) or in (7).

Using these transformed data, proceed as directed in Annex A14. Or alternatively, proceed as directed in Annex A10 or Annex A11 whichever is appropriate, depending on the sample size.

### A13. POISSON DISTRIBUTION

A13.1 *Procedure*—For a randomized block experiment, limited to that described in Annex A9, and producing data that can be modeled by a Poisson distribution, transform the data, using the appropriate transformation found in (6) or in (7).

Using these transformed data, proceed as directed in Annex A14. Alternatively, proceed as directed in Annex A10 or Annex A11, whichever is appropriate, depending on the sample size.

### A14. NORMAL DISTRIBUTION

#### A14.1 *Procedure*:

A14.1.1 For a randomized block experiment limited to that described in Annex A9, with  $n$  blocks and  $k$  factor levels, make the following calculations:

- (1) = sum of all the observations divided by the number of observations,  $kn$ .
- (2) = sum of the squares of each observation.
- (3) = sum of the squares of each factor level total divided by the number of blocks,  $n$ .
- (4) = sum of the squares of each block total divided by the number of factor levels,  $k$ .

A14.1.2 Write an analysis of variance table, using the format shown in Table A14.1.

NOTE A14.1—The residual sum of squares is the total of the experimental error sum of squares and the sum of squares due to interaction of factor levels and blocks. If there is no interaction, then the residual sum of squares is equal to the experimental error sum of squares.

A14.1.3 Determine the critical values of the  $F$ 's by finding the values at the degrees of freedom for the two mean squares and a probability level of 95 % in a table of the percentage points of the  $F$ -distribution (5). If the calculated  $F$  for factor levels exceeds the critical value, then conclude that there is a significant difference between factor levels. Significance

**TABLE A14.1 ANOVA Format**

| Source of Variation | Sum of Squares <sup>A</sup> | Degrees of Freedom | Mean Squares <sup>B</sup> | $F$ |
|---------------------|-----------------------------|--------------------|---------------------------|-----|
| Factor              | (3)–(1)                     | $n - 1$            | $L$                       | L/R |
| Blocks              | (4)–(1)                     | $m - 1$            | $B$                       |     |
| Residual            | (2)–(3)<br>–(4)+(1)         | $(m - 1)(n - 1)$   | $R$                       |     |

<sup>A</sup> The legend for the sum of squares is given in A14.1.1.

<sup>B</sup> Each mean square  $L$ ,  $B$ , and  $R$  is calculated by dividing the sum of squares by the corresponding degrees of freedom.

means that some of the test variability is explained by changes in levels of the factor. Test variability may be reduced by reducing the difference in factor levels.

#### A14.2 *Example*:

A14.2.1 A plant had three tensile testing machines of the same make and type. It was desired to determine whether or not different results were being obtained from the three machines. In order to do this, the following randomized block experiment was run. The design of this experiment is as directed in Annex A9. From the same cone of 20's rayon yarn, one specimen was tested for breaking strength by the same operator on each machine on each of six days. The tests were made using Test Method D 2256 with option 3A. Table A14.2 gives the results of testing in the experiment.

**TABLE A14.2 Breaking Strength in Pounds-Force**

| Day   | Machine |     |     | Total |
|-------|---------|-----|-----|-------|
|       | A       | B   | C   |       |
| 1     | 1.7     | 1.3 | 1.5 | 4.5   |
| 2     | 1.6     | 1.4 | 1.4 | 4.4   |
| 3     | 1.8     | 1.5 | 1.7 | 5.0   |
| 4     | 1.3     | 1.3 | 1.6 | 4.2   |
| 5     | 1.5     | 1.9 | 1.7 | 5.1   |
| 6     | 1.7     | 1.5 | 1.5 | 4.7   |
| Total | 9.6     | 8.9 | 9.4 | 27.9  |

A14.2.2 The preliminary calculations for determining the sums of squares in an ANOVA were done as directed in A14.1.1 with the following results:

- (1) = 43.2450;
- (2) = 43.7700;
- (3) = 43.2883;
- (4) = 43.4500.

A14.2.3 The calculations for the ANOVA table were done as directed in A14.1.2. The resulting table follows:

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Squares | <i>F</i> |
|---------------------|----------------|--------------------|--------------|----------|
| Machines            | 0.0433         | 2                  | 0.0216       | 0.78     |
| Days                | 0.2050         | 5                  | 0.0410       |          |
| Residual            | 0.2767         | 10                 | 0.0277       |          |
| Total               | 0.5250         | 17                 |              |          |

A14.2.4 From a table in (5), at the 5 % probability level, the critical value of *F* for machines is 4.10 at two and ten degrees

of freedom. Since the sample value of  $F = 0.78$  is less than the critical value, conclude that the variation due to machines is not significantly different from the experimental error (see Note A14.1). Therefore, conclude that neither different machines, nor change of conditions from day to day affect the precision of the test.

## AVERAGING

### A15. NO COMPOSITING

A15.1 *Variance Components*—For variables data (Note A15.1), determine the variance components of the three stages (lot, laboratory, and specimen) of sampling as directed in Section 5 of Guide D 4854.

NOTE A15.1—Sampling plans that produce attribute data usually do not take specimens in stages, but require that specimens be taken at random from all of the individual items in the lot.

A15.2 *Test Results Variability*—Calculate the test results variability,  $s^2$  (Note A15.2), for as many different sampling plans as desired, using Eq A15.1:

$$s_2 = L/m + T/mn + E/mnk \quad (\text{A15.1})$$

where:

- $s^2$  = variance of test results,
- $L$  = lot sample component of variance,
- $m$  = number of lot sampling units,
- $T$  = laboratory sample component of variance,
- $n$  = number of laboratory sampling units per lot sampling unit,
- $E$  = specimen component of variance, and
- $k$  = number of specimens per laboratory sampling unit.

NOTE A15.2—(Eq A15.1 and A16.1) are correct regardless of the distribution from which the data come.

A15.3 *Sampling Plan Selection*—Select the desired sampling plan and put it into the test method instead of any sampling plan that may have been in the test method.

### A16. COMPOSITING

A16.1 *Test Results Variability*—Calculate the test results variability,  $s^2$  (Note A15.2), for as many different sampling and compositing plans as desired, using Eq A16.1:

$$s_2 = L/m + T/mn + E/a \quad (\text{A16.1})$$

where:

- $E$  = specimen testing component of variance,
- $a$  = total number of tests on all the composite samples, and the other symbols are as defined for (Eq A15.1).

A16.2 *Example*—A test method specifies as follows: (1) from a lot of staple fiber take four bales as lot sampling units, (2) from each bale take three laboratory sampling units of 10.0 g each, (3) blend the three laboratory sampling units from a bale, and (4) test two specimens from each of the four blended laboratory sampling units. In calculating the variance of test results, using Eq A16.1, the values of the three denominators will be:  $m = 4$ ;  $mn = (4)(3) = 12$ ; and  $a = (4)(2) = 8$ .

**REFERENCES**

- (1) *ASTM Manual on Presentation of Data and Control Chart Analysis*, ASTM STP 15D, ASTM, 1976.
- (2) Shapiro, S. S., *How to Test Normality and Other Distribution Assumptions*, American Society for Quality Control, Milwaukee, Wisconsin, 1980, Vol 3 of the series *The ASQC Basic References in Quality Control: Statistical Techniques*, Edward J. Dudewitz, Ed.
- (3) Hollander, M., and Wolf, D. A., *Nonparametric Statistical Methods*, John Wiley & Sons, New York, NY, 1973.
- (4) McClave, J. T., and Dietrich, F. H., II, *Statistics*, Dellen Publishing Company, San Francisco, 1970, pp. 129, 142, and 260.
- (5) Pearson, E. S., and Hartley, H. O., *Biometrika Tables for Statisticians*, Vol 1, Cambridge University Press, Cambridge, England, 1954.
- (6) Brownlee, K. A., *Industrial Experimentation*, Chemical Publishing Company, Brooklyn, NY, 1953.
- (7) Hald, A., *Statistical Theory with Engineering Applications*, John Wiley & Sons, Inc. London, 1952.

*ASTM International takes no position respecting the validity of any patent rights asserted in connection with any item mentioned in this standard. Users of this standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, are entirely their own responsibility.*

*This standard is subject to revision at any time by the responsible technical committee and must be reviewed every five years and if not revised, either reapproved or withdrawn. Your comments are invited either for revision of this standard or for additional standards and should be addressed to ASTM International Headquarters. Your comments will receive careful consideration at a meeting of the responsible technical committee, which you may attend. If you feel that your comments have not received a fair hearing you should make your views known to the ASTM Committee on Standards, at the address shown below.*

*This standard is copyrighted by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, United States. Individual reprints (single or multiple copies) of this standard may be obtained by contacting ASTM at the above address or at 610-832-9585 (phone), 610-832-9555 (fax), or [service@astm.org](mailto:service@astm.org) (e-mail); or through the ASTM website ([www.astm.org](http://www.astm.org)).*