C-89-05

Designation: F 465 – 76 (Reapproved 1981)

Standard Practice for DEVELOPING PRECISION AND ACCURACY DATA ON ASTM METHODS FOR THE ANALYSIS OF MEAT AND MEAT PRODUCTS[1]

This standard is issued under the fixed designation F 465; 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 establishes uniform guidelines for expressing the precision and accuracy of methods for the analysis of meat and meat products. It includes a procedure for developing this information. There is no intent to restrict qualified groups in their use of other techniques.

1.2 Statements of precision are restricted to those variables specifically mentioned. Task groups are referred to **(1, 2, 3).**[2]

2. Applicable Documents

2.1 *ASTM Standards:*

E 178 Recommended Practice for Dealing with Outlying Observations[3]

E 180 Recommended Practice for Developing Precision Data on ASTM Methods for Analysis and Testing of Industrial Chemicals[4]

F 463 Test for Fat in Meat and Meat Products by Ether Extraction[5]

3. Definitions

3.1 *error*—in a statistical sense, any deviations of an observed value from the true value. When expressed as a fraction or percentage of the value measured, it is called a relative error. All statements of precision or accuracy should indicate clearly whether they are expressed in absolute or relative sense.

3.2 *random error*—the chance variation encountered in all experimental work despite the closest possible control of variables. It is characterized by the random occurrence of both positive and negative deviations from the mean value for the method, the algebraic average of which will approach zero in a long series of measurements.

3.3 *bias*—a constant or systematic error as opposed to a random error. It manifests itself as a persistent positive or negative deviation of the method average from the accepted reference value.

3.4 *precision*—the degree of agreement of repeated measurements of the same property. Precision statements in ASTM methods for analysis of meat and meat products will be derived from the estimated standard deviation of a series of measurements and will be expressed in terms of the repeatability and reproducibility of the method.

3.5 *accuracy*—the agreement between an experimentally determined value and the accepted reference value.

3.6 *variance*—a measure of the dispersion of a series of results around their average. It is the sum of the squares of the individual deviations from the average of the results, divided by the number of results minus one.

3.7 *standard deviation*—a measure of the dispersion of a series of results around their average, expressed as the square root of the quantity obtained by summing the squares of the deviations from the average of the results and dividing by the number of observations

[1] This practice is under the jurisdiction of ASTM Committee F-10 on Meat and Meat Products.

Current edition approved Aug. 27, 1976. Published December 1976.

[2] The **boldface** numbers in parentheses refer to the list of references at the end of this recommended practice.

[3] *Annual Book of ASTM Standards*, Part 41.

[4] *Annual Book of ASTM Standards*, Parts 29 and 30.

[5] *Annual Book of ASTM Standards*, Part 46.

minus one. It is also the square root of the variance and is calculated as follows:

$$s = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n - 1}}$$

where:

s = estimated standard deviation of the series of results,
X_i = each individual value,
$\bar{X}$ = average (arithmetic mean) of all values, and
n = number of values.

3.7.1 The following forms of this equation are more convenient for computation, especially when using a calculator:

$$s = \sqrt{\frac{\Sigma X^2 - (\Sigma X)^2/n}{n - 1}} \quad \text{or}$$

$$s = \sqrt{\frac{n\Sigma X^2 - (\Sigma X)^2}{n(n - 1)}}$$

where:

s = estimated standard deviation,
ΣX^2 = sum of the squares of all of the individual values,
$(\Sigma X)^2$ = square of the total of the individual values, and
n = number of values.

NOTE 1—Care must be taken in using either of these equations to be sure that a sufficient number of decimal places is carried in the sum of the values and in the sum of their squares so that serious rounding errors do not occur. For best results, all rounding should be postponed until after a value has been obtained for s. In this recommended practice, the standard deviation is obtained from an analysis of variance of the results of an interlaboratory test program (see Section 7).

3.8 *coefficient of variation*—a measure of relative precision calculated as the standard deviation of a series of values divided by their average. It is usually multiplied by 100 and expressed as a percentage.

3.9 *range*—the absolute value of the algebraic difference between the highest and the lowest values in a set of data.

3.10 *duplicates*—paired determinations performed by one analyst at essentially the same time. This concept also applies to other such multiple determinations.

3.11 *95 % confidence interval or confidence limits*—that interval or range of values around an observed value which will, in 95 % of the cases, include the expected value. The expected value is the average of an infinite series of such determinations.

3.12 *95 % confidence level*—a term commonly used for establishing the probability of precision statements and means that there are 95 in 100 chances of being correct, and 5 in 100 chances of being wrong, when predicting that the expected precision (or expected value) will fall within the specified limits or range.

3.13 *repeatability*—the precision of a method expressed as the agreement attainable between independent determinations performed by a single analyst using the same apparatus and techniques (see 5.2.6, 7.2, and 7.2.12).

3.14 *reproducibility*—the precision of a method expressed as the agreement attainable between determinations that are performed in different laboratories.

4. Preliminary Studies

4.1 *General*—This section covers the preliminary work that should be carried out in a few laboratories before undertaking a full interlaboratory evaluation of a method.

4.2 When a task group is asked to provide a specific analytical procedure, there may be one or more methods available from the literature or from laboratories already performing such analyses. In such cases, these methods have usually been the subject of considerable research, therefore any additional study of variables, at this stage, would be a waste of time. It is recommended that such methods be rewritten in ASTM format, with full descriptions of the equipment and procedure, and be evaluated in a pilot run by a few laboratories on selected materials. Three laboratories and at least three such materials, using one or two analysts performing duplicate determinations on each of two days, by each method, constitutes a practical plan that can be analyzed by the procedures described in Sections 6, 7, and 8. Such a pilot study will confirm the adequacy of the methods and supply qualitative indications of relative precision and accuracy.

4.3 When the method to be evaluated is new, or represents an extensive modification of an available method, it is recommended that a study on variables be carried out by at least one laboratory to establish the parameters and conditions to be used in the descrip-

tion of the method. This should be followed by a three-laboratory pilot study before undertaking a full interlaboratory evaluation.

4.4 Detailed procedures for executing such preliminary studies are not described in this recommended practice but are available in the general statistical literature **(4)**.

5. Planning the Interlaboratory Study

5.1 *General*—This section covers the recommendations for the planning of interlaboratory studies.

5.2 *Variables*—The major variables to be considered are methods, materials or levels, laboratories, apparatus, analysts, days, and runs as follows:

5.2.1 *Methods*—The preliminary studies of Section 4 should lead to an agreement on a single method, which can then be evaluated in a full interlaboratory study. If it is necessary to evaluate two or more methods, the complete program must be carried out on each method. In either case, it will be assumed that the variables for each method have been explored and that a well-standardized, fully detailed procedure has been prepared. The time and expense required for an extensive precision study cannot be justified if the preparation is incomplete.

5.2.2 *Materials or Levels*—The number of samples distributed should be held to the minimum needed to evaluate the method adequately. (Increasing the number of samples will not significantly increase the degrees of freedom available for predicting the reproducibility of the method. This can be achieved only by increasing the number of laboratories.) Some interlaboratory studies can be limited to a single sample, as in the case of preparing a specific standard solution. Methods applicable to a single product of high purity can usually be evaluated with one or two samples. When different concentrations of a constituent or values of a chemical property are involved, the samples should represent the approximate lower, middle, and top levels of the expected range. If these vary over a wide range, the number of levels should be increased and spaced to cover the range.

5.2.3 *Laboratories*—To obtain a reliable precision estimate, it is recommended that the interlaboratory study include approximately ten qualified laboratories. When this number of independent laboratories cannot be recruited, advantage can be taken of a liberalized definition of collaborating laboratories, quoted as follows from p. 9 of STP 335 **(3)**:

> Here the term "collaborating laboratory" has a more specific meaning than in common usage. For example, a testing process often consists of an integrated sequence of operations using apparatus, reagents, and measuring instruments; and several more or less independent installations may be set up in the same area or "laboratory." Each such participating installation should be considered as a collaborating laboratory so far as this procedure is concerned. Similarly, sets of test results obtained with different participants or under different conditions of calibration would in general constitute results from different collaborating laboratories even though they were obtained on the same sets of equipments.

This concept makes it possible to increase the available "laboratories" by using two analysts (but not more than two) in as many laboratories as needed to bring the total to the recommended minimum of ten. In such cases, the two analysts must evaluate the method independently in the fullest sense of the word, interpret as using different samples, different reagents, different apparatus where possible, and perform the work on different calendar days. (In the design in 5.8, laboratories using two analysts are designated as A-1, A-2, B-1, B-2, etc.). The most desirable laboratories and analysts are those having previous experience with the proposed method or with similar methods. It is essential that enough experience be acquired to establish confidence in the performance of a laboratory before starting the interlaboratory test series. Such preliminary work must be done with samples other than those to be used in the formal interlaboratory test program.

5.2.4 *Apparatus*—The effect of duplicate setups is not often a critical variable in chemical analysis. In instrumental methods, however, apparatus can become an important factor because the various laboratories may be using different makes or types of equipment, for example, the various colorimeters and spectrophotometers used in photometric methods. In such cases, the effect of apparatus becomes confounded with between-laboratory variability, and special care must be used to avoid misinterpreting the results. Of course, if enough laboratories have instruments of each type, the apparatus can be made a planned variable in the study.

5.2.5 *Analysts*—The use of a single analyst in each laboratory (as described in 5.2.3) is adequate to provide the information needed for calculating the repeatability and reproducibility of the method as covered in this recommended practice. It is essential that all analysts complete the entire interlaboratory test program. With regard to analyst qualifications, an analyst who is proficient in the method should be selected.

5.2.6 *Days*—The repeatability of the method (see 3.13) shall be evaluated in terms of independent determinations by the same analyst. To achieve this, all scheduled determinations must be performed on each of two days (see 5.8 and 7.2).

NOTE 2—As used in this recommended practice, the term "days" represents replication of a set of determinations performed on any day other than that on which the first set was run. It may become a systematic variable to the extent that it is desirable that a given laboratory run the entire set on another. Although this may introduce a bias for that laboratory, there appears to be little chance that such a bias would be common to all laboratories. When prelminary studies suggest that instability may result in an overall systematic "days" effect, special planning will be required to take care of this problem.

5.2.7 *Runs*—The multiple determinations performed at the same time or within a very short time interval, on each day should be taken into consideration.

5.3 *Number of Determinations*—Each analyst is required to perform duplicate determinations on each sample on each of two days. If one determination of a paired set is accidentally ruined, another pair must be run. An odd or unusual value does not constitute a ruined determination. In such cases, an additional set of duplicate determinations should be run and all values reported, with an assignable cause, if at all possible.

5.4 *Samples:*

5.4.1 One person should be made responsible for accumulating, subdividing, and distributing the materials to be used in the test program. Extra samples should be held in reserve to permit the necessary replacement of any that may be lost or damaged in transit. Proper techniques in packaging and sampling should be followed. If a collaborating laboratory should receive a sample that shows evidence of leakage, or that is suspect for any other reason, the recipient should not use it and should immediately request a replacement.

5.4.2 The most important requirement is that the subsamples be representative and as homogeneous as possible. This is to be emphasized in preparing samples involving the comparison of a test with a reference method. If collaborating laboratories are involved in evaluating a single method, a high degree of homogeneity should be assured. In such cases, it is recommended that the person responsible for preparing and distributing samples conduct analyses on representative lots to determine that the sampling error is acceptably low.

5.4.3 Instability of any type may impose other restrictions on the execution of a planned program. It is the responsibility of the task group chairman to include in the plans for the interlaboratory study specific instructions on selecting, preparing, storing, and handling of the standard samples.

5.4.4 The samples distributed for the formal interlaboratory test program should not be used for practice runs. Where dryruns are performed to develop proficiency in an inexperienced analyst or laboratory, conduct the test on samples other than those distributed.

5.5 *Scheduling and Timing*—Interlaboratory studies fail occasionally because a timetable had not been established to cover the program, particularly in cases where the materials have changes in storage, after opening the container, etc. The instructions to the collaborators should cover such points as the time between receipt of samples and their testing, the time elapsing between start and finish of the program, and the order of performing the tests, etc., with particular attention to randomizing as a means of avoiding systematic errors.

NOTE 3—A discussion of randomizing is beyond the scope of this recommended practice. Refer to standard textbooks on statistics and specifically to Refs **(5, 6)**.

5.6 *Instructions and Preliminary Questionnaire*—Having decided on the variables and levels for each, the task group chairman should distribute to all participants a complete description of the planned collaborative study, emphasizing any special conditions or precautions to be observed. A detailed procedure and description of equipment, prepared

in ASTM format, must be included. A questionnaire similar to the one in Table 1 will aid materially in the successful execution of the interlaboratory study.

5.7 *Report Form*—A form for reporting the essential data should be prepared and distributed (in duplicate) to all collaborators, who should be instructed on the number of decimal places to be used. It is recommended that interlaboratory studies be reported to one decimal place beyond that called for in the "Report" instructions of the method under study. Any subsequent rounding-off should be done by the task group chairman or the data analyst.

5.8 *Design for an Interlaboratory Test Program*—The plan given in Table 2 should cover most cases where laboratories and levels (or materials) are the principle variables. This plan calls for each analyst to perform two determinations in parallel on each of two days, at each level. Where additional variables must be included, the proposed program should be referred to a statistician or to the subcommittee on precision and accuracy for a specific recommendation.

6. Determination of Outlying Observations

6.1 *General*—This section covers some elementary recommendations for dealing with outlying observations and rejection of data. Lacking a universally accepted practice for the rigid application of available statistical tests, considerable technical and common-sense judgments must be exercised in using them. Accordingly, the following procedures are offered only as guides for the data analyst and all decisions to exclude or to include any suspect data shall be subject to the approval of the task group concerned.

NOTE 4—The test for outlying observations should be applied only once to a set of interlaboratory test data. Although two or more values can be rejected simultaneously, in no case should the remaining data again be tested for outliers.

6.2 *Principle of Method:*

6.2.1 The tests for outliers among the "multiple runs" and "different days" data are based on control chart limits for the range, as described in ASTM STP 15 C **(7)**.

6.2.2 The test for outlying observations among laboratory averages is described in Recommended Practice E 178. For supplemental tests which are optional, task forces are referred to the methods of graphical diagnosis **(1)** and of laboratory ranking using averages **(2)**.

6.2.3 The choice of probability levels for each of the three tests is based on practical experience gained from a number of interlaboratory studies involving chemical or physical properties.

NOTE 5—In choosing probability levels, there are two alternatives: *(1)* using a high probability level, accepting the divergent data, inflating variances, and perhaps failing to find significant differences; or *(2)* using a lower probability level, rejecting the divergent data, deflating variances, and perhaps finding significance where none exists. In the case of multiple runs in an interlaboratory test program, the choice of the 99.9 % level is based on the premise that only a high degree of divergence should justify rejection of data from a laboratory for this reason. The 99.0 % level for days also reflects this premise. The 95 % level for laboratories is frequently used and is chosen here because an outlying laboratory average, even at this probability level, may have a pronounced effect on the claimed reproducibility of the method (see also 6.7.2).

6.2.4 The procedures are illustrated by data developed in an interlaboratory study on the determination of fat content (see Table 3).

6.3 *Outliers Between Runs:*

6.3.1 Using the data of Table 3, tabulate the results of the duplicate runs on each of two days, in each of the twelve laboratories. Calculate the individual ranges and the average range as shown in Table 4.

6.3.2 Multiply the average range by the factor 3.488 to obtain the critical range at a 99.9 % probability level (0.1 % significance level). For the seven samples in question, these values are:

Sample	Average Range	Critical Range
Beef-1	0.27	0.94
Beef-2	0.50	1.74
Beef-3	0.67	2.33
Pork-1	0.21	0.73
Pork-2	0.90	3.13
Frankfurter	0.35	1.22
Bologna	0.18	0.62

NOTE 6—The factor 3.488 is the D_4 value used to calculate the upper control limit for the range and is derived by the following equation:

$$D_4 = 1 + td_3/d_2$$

where d_3 and d_2 apply to the range of two values (see Table B2, p. 115 of ASTM STP 15 C) and t is the two-tailed value of the "t" distribution for $p = 0.001$ and DF = ∞.

The following are the D_4 factors at other probability levels, for values of $n = 2, 3$, and 4 replicates:

Probability Level, %	$n = 2$	$n = 3$	$n = 4$
99.9	3.488	2.728	2.405
99.73 (3s)	3.267	2.575	2.282
99.0	2.947	2.352	2.100
95.0	2.482	2.029	1.837

6.3.3 Scan the individual ranges of Table 4 for values exceeding the critical range. For this example, the following values occur:

Sample	Critical Range	Observed Range	Suspect Laboratory
Beef-1	0.94	2.80, 1.10, 0.45	5
Beef-2	1.74	4.80, 3.75, 0.46	5
Beef-3	2.33	9.95, 3.30, 0.50	5
Pork-1	0.73	2.25, 1.15, 0.35	5
Pork-2	3.13	18.30, 0.70	5
Frankfurter	1.22	3.85, 0.78	5
Bologna	0.62	(0.55, max)	none

The indicated laboratories are suspect as rejectable at a 99.9 % probability level.

6.4 *Outliers Between Days:*

6.4.1 Calculate the averages (to 0.1 unit) of the duplicate runs performed each day. Tabulate and determine the individual ranges and the average range as shown in Table 5.

6.4.2 Multiply the average range by the factor 2.947 to obtain the critical range at a 99.0 % probability level (1.0 % significance level). Scan the individual ranges of Table 5 for values exceeding the critical range. For this example, the values are as follows:

Sample	Average Range	Critical Range	Observed Range	Suspect Laboratory
Beef-1	0.70	2.06	5.90, 0.55	5
Beef-2	0.37	1.09	2.13, 0.43	5
Beef-3	1.01	2.97	9.52, 0.83	5
Pork-1	0.23	0.67	1.90, 0.20	5
Pork-2	0.32	0.94	1.10, 0.65	9
Frankfurter	0.54	1.59	4.57, 0.41	5
Bologna	0.29	0.85	1.17, 0.50	5

The indicated laboratories are suspect as rejectable at a 99.0 % probability level.

6.5 *Outliers Among Laboratory Averages:*

6.5.1 Calculate the laboratory averages (to one place beyond normal reporting level) and tabulate (see Table 6).

6.5.2 *Optional*—Rank the average of each laboratory's four determinations on each sample from 1 for the highest average, ranging to 12 for the lowest, as shown in Table 7, and calculate the collaborator's scores as shown. Suspect laboratories are any whose score is beyond limits as tabulated by Youden **(1)** in Table 8.

6.5.3 *Optional*—Prepare two-sample plots of laboratory averages (means) from Table 6, as previously described **(1)** and as illustrated in Figs. 1 and 2, and examine the plots to identify the laboratories whose data are erratic or exhibit marked systematic errors relative to the consensus.

6.5.4 Determine the standard deviation among laboratory averages for each material using the calculating form of the equation given in 3.7 (see Table 6).

6.5.5 Referring to Recommended Practice E 178, calculate the test criteria as follows:

$$T_n = (X_n - \bar{X})/s$$
$$T_1 = (\bar{X} - X_1)/s$$

6.5.6 Using Table 9, obtain the critical value of T at the 5 % significance level (95 % probability) for $n = 12$. Comparing the observed with the critical values, the data show:

Sample	Critical T	Observed T_n or T_1	Suspect Laboratory
Beef-1	2.285	(2.10 max)	none
Beef-2	2.285	3.02	5
Beef-3	2.285	3.06	5
Pork-1	2.285	3.02	5
Pork-2	2.285	3.17	5
Frankfurter	2.285	(2.25 max)	none
Bologna	2.285	2.83	5

The indicated laboratories are suspect as rejectable at a 95.0 % confidence level.

6.5.7 Recommended Practice E 178 also indicates that an alternative system based entirely on ratios of simple differences among the determinations is given in the literature **(8, 9)**. That procedure may be used if it is felt highly desirable to avoid the calculation of s.

6.6 *Summary*—The data of 6.3, 6.4, and 6.5 can be summarized as follows:

Sample	Laboratories Suspect as Rejectionable		
	Runs at 99.9 %	Days at 99.0 %	Laboratory Averages at 95.0 %
Beef-1	5	5	none
Beef-2	5	5	5
Beef-3	5	5	5
Pork-1	5	5	5
Pork-2	5	9	5
Frankfurter	5	5	none
Bologna	none	5	5

6.7 *Discussion:*

6.7.1 When the operations in Section 6 show any set of data from a laboratory to be suspect, every effort should be made to find an assignable cause that will justify rejection.

6.7.2 As this recommended practice does not provide procedures for the analysis of data in which values are missing, rejection in any one of the three categories (runs, day, or laboratories) makes it necessary to exclude from the analysis of variance all of the data from that laboratory which is pertinent to the material or sample in question.

6.7.3 Although rejected data are usually excluded before performing the analysis of variance, it is advisable to perform the analysis using the entire set, as well as after the elimination of the suspect data. Calculation of the variances with and without the suspect data yields two sets of results which may be compared and the comparison may be useful in appraising the results of the entire program, as well as in deciding whether or not the rejection is justified.

7. Determination of Components of Precision

7.1 *General:*

7.1.1 This section demonstrates the statistical analysis of typical data obtained with the design of 5.8.

7.1.2 An abridged analysis of variance (see 7.2) gives the basic information needed for calculating repeatability and reproducibility as defined in this recommended practice. It determines the between-laboratories and within-laboratories variances for each level and combines them to give the two pertinent standard deviations or coefficients of variation.

7.1.3 This procedure disregards interactions. An optional procedure is outlined for use in determining interactions (see 7.4). Task groups are referred to the literature **(1, 2)** and specifically to ASTM STP 335 **(3)**. Laboratory-sample interaction of the collaborative data is calculated by referring to the published procedure **(2)** for analysis of variance using data on all samples at one time rather than one at a time and is outlined in 7.4.

7.2 *Analysis of Variance*—The abridged analysis of variance is illustrated in the following sections by an example representing a collaborative study of a single method involving five meat and two meat product samples and an adequate number of laboratories, with one qualified analyst in each laboratory carrying out two determinations (paired duplicates) on each of 2 days. Although by some definitions, the repeatability estimate can be based on the variation between paired duplicates, experience in chemical testing shows that such estimates are usually more optimistic and imply a superior level of precision than when they are derived from independent determinations performed on different days. This recommended practice uses the duplicate results only for calculating acceptable checking limits for such duplicates (see 7.2.10 and 7.2.12.1) and elects to base estimates of repeatability on the averages of the duplicate determinations obtained on each of 2 days. Accordingly, the analysis of variance determines the within-laboratory, between-days variance and the between-laboratory variance for each sample and provides for combining the data for all samples to give overall standard deviations (or coefficients of variation) that are used to calculate the repeatability and reproducibility of the method.

7.2.1 *Specific Example*—Seven meat and meat product samples were analyzed for fat content by a single analyst, in each of twelve laboratories. The entire set of data is in accordance with Section 6 and Table 3. Only the results for sample Beef-1 are used in the following sections to demonstrate the analysis of variance.

7.2.2 *Homogeneity of Data and Testing of Outliers*—On applying the tests for outliers (see 6.6), the results of Laboratory 5 were excluded because of divergent values between runs and days and among the laboratory averages. Table 10 shows the remaining data as the averages of the duplicate determinations.

7.2.3 *Coded Data*—To avoid handling large numbers in the analysis of variance, data can be coded by arbitrarily selecting a constant for each sample and then subtracting the selected value from all determinations of a sample without affecting the results. In performing this abridged analysis of variance, it was elected to use the collaborative fat determinations without coding.

7.2.4 *Analysis of Variance*—Perform the following operations either directly on the values determined or on the coded data.

7.2.4.1 Square the individual values and add them, as follows (use day averages, Table 5, Table 10, or coded data):

$$10.70^2 + 10.85^2 + 10.75^2 + \ldots + 10.10^2 + 10.98^2 + 10.80^2 = 2556.3800 \quad (1)$$

7.2.4.2 Square the column totals, add them, and divide by the number of values in each column, as follows:

$$(21.93^2 + 21.77^2 + \ldots + 22.51^2 + 21.58^2)/2 = 2555.8933 \quad (2)$$

7.2.4.3 Add the individual values, square this total, and divide by the number of values, as follows:

$$(10.70 + 10.85 + \ldots + 10.98 + 10.80)^2/22 = 2553.5672 \quad (3)$$

7.2.4.4 From the analysis of variance, shown in Table 11, calculate the components of variance as follows:

$$s_a^2 = 0.0442; s_a = \sqrt{0.0442} = 0.2103$$

$$s_b^2 = (0.2326 - s_a^2)/2 = (0.2326 - 0.0442)/2 = (0.1884)/2 = 0.0942$$

$$s_{a+b}^2 = s_a^2 + s_b^2 = (0.0442 + 0.0942) = 0.1384;$$

$$s_{a+b} = \sqrt{0.1384} = 0.3721$$

where:

s_a = estimated standard deviation of a single result (average of duplicates) within a laboratory, based on 11 degrees of freedom, and

s_{a+b} = estimated standard deviation of a single result (average of duplicates) in any laboratory, based on approximately 10 degrees of freedom (see 7.2.11).

7.2.5 *Other Materials*—Perform analyses of variance on the data for the other six samples, using the example in 7.2.4 as a model. These analyses are not illustrated, but the results are shown in 7.2.6.

7.2.6 *Pooling of Data*—Summarize the data for the seven materials in the format shown in Table 12.

7.2.6.1 The tabulated values should exhibit one of the following three patterns: (*1*) the s_a or the s_{a+b} values are in good agreement for the seven samples, in which case, proceed with pooling as shown in 7.3; (*2*) the coefficients of variation are in agreement for the seven samples; or (*3*) neither show the desired uniformity.

7.2.6.2 In Table 12, the range of s_a is 0.0745 − 0.2478 and represents a three-fold variation. The range of the corresponding coefficient of variation is 0.4066 − 2.1326 and represents a five-fold variation. The lower variation of s_a is the first indication that it would be preferable to pool the values of s_a corresponding to the 7 samples rather than those of the coefficient of variation. As a second indication, inspection of the graphical treatment of s_a versus fat content and coefficient of variation versus fat content, shows that except for the lowest fat content (*P*-1), s_a does not vary in proportion to fat content, but the coefficient of variation does decrease smoothly for increasing fat content.

7.2.6.3 Confirmation of which values of s_a can be pooled is obtained by testing the sample variances (s_a^2) obtained in 7.2.4.4 for homogeneity by Bartlett's chi-square test **(10)** as described in 7.2.7.

7.2.7 *Homogeneity of Variance by Bartlett's Test (Optional)*—Perform Bartlett's chi-square test on any group of variances (s^2 duplicates, s_a^2, s_b^2, or s_{a+b}^2) to be able to select which individual sample variances of the group can be pooled to obtain an overall variance that is representative of the pooled group of samples, where:

$$\text{chi-square} = 2.3026 \left\{ \left[\sum_1^a (n_i - 1) \right] \log \bar{s}^2 - \sum_1^a (n_i - 1) \log s_i^2 \right\}$$

7.2.7.1 *Step 1*—Tabulate the sample variances (from 1 to *a*) as shown in Table 13 for s_a^2, code the variances with a multiplicative code to avoid calculating negative logarithms, and obtain the common logarithms of the coded values. This coding has no effect upon the results of the test.

7.2.7.2 *Step 2*—Sum the degrees of freedom of the individual variances as follows: 21 + 21 + 21 + 21 + 19 + 21 + 21 = 145.

7.2.7.3 *Step 3*—Compute a weighted average coded variance ($\bar{s}^2$) as follows:

$$[(21 \times 44.245) + (21 \times 33.591) + (21 \times 61.391) + (21 \times 5.545) + (19 \times 39.64) + (21 \times 22.336) + (21 \times 36.573)]/145 = 34.6928$$

7.2.7.4 *Step 4*—Obtain the logarithm of coded $\bar{s}^2$ as follows:

$$\log 34.6928 = 1.5402$$

7.2.7.5 *Step 5*—Compute the sum of the products of the log-coded variances and multiply by the degrees of freedom:

$$(21 \times 1.6459) + (21 \times 1.5262) + (21 \times 1.7881) + (21 \times 0.7439) + (19 \times 1.5981) + (21 \times 1.3490) + (21 \times 1.5632) = 211.3059$$

7.2.7.6 *Step 6*—Compute the chi-square value as a natural logarithmic value as follows:

$$\chi^2 = 2.3026\,[(\text{quantity 2} \times \text{quantity 3}) - \text{quantity 4}]$$
$$= 2.3026\,[(145 \times 1.5402) - 211.3059]$$
$$= 27.6977$$

7.2.7.7 *Step 7*—Compute a correction factor to adjust the chi-square value for the degrees of freedom involved. The correction factor is especially useful if the chi-square value lies slightly above the borderline of significance.

$$\text{Correction factor, } C = 1 + \frac{1}{3(a-1)}\left[\sum_1^a \frac{1}{(n_i - 1)} - \frac{1}{\sum_1^a (n_i - 1)}\right]$$
$$= 1 + \frac{1}{3(7-1)}\left[\frac{1}{21} + \frac{1}{21} + \frac{1}{21} + \frac{1}{21} + \frac{1}{19} + \frac{1}{21} + \frac{1}{21} - \frac{1}{145}\right]$$
$$= 1 + (0.0555)\,(0.3314)$$
$$= 1.0184$$

7.2.7.8 *Step 8*—Compute the adjusted chi-square value using the correction factor:

$$\text{chi-square (adjusted)} = 27.6977/1.0184 = 27.19; \text{ where } a - 1 = 6 \text{ DF}$$

7.2.7.9 *Step 9*—Compare the calculated chi-square (adjusted) value with the tabular value (see Table 14) for a selected probability level and the appropriate degrees of freedom. For the present example, the calculated value exceeds the tabular value of 12.59 ($P = 0.05$, 6 DF) indicating that the seven variances are significantly heterogeneous.

7.2.7.10 *Step 10*—When heterogeneity is indicated, compute a new chi-square value excluding the highest or lowest variance of the group. By inspection of the variances listed in Table 13, there is a greater arithmetic difference between the two lowest variances than between the two highest. Excluding the lowest variance (sample P-1), a chi-square (adjusted) value of 5.6 (5 DF) is obtained. This does not exceed the tabular value of 11.1 ($P = 0.05$, 5 DF), and indicates homogeneity so that the six variances can be pooled as shown in 7.3.2 with its respective (k) (m) $(n - 1)$ DF.

7.2.8 When the values of the coefficient of variation of repeatability show good agreement, the mathematical procedure for pooling them is analogous to pooling variances. Using the values for within-laboratories for six of the samples shown in Table 12, for example, a pooled coefficient can be calculated as follows using the same respective (k) (m) $(n - 1)$ degrees of freedom for weighting:

$$\frac{s_a \times 100}{\bar{X}} = \{[(11)\,(1.952)^2 + (11)\,(0.895)^2 + (11)\,(0.972)^2 + (10)\,(0.407)^2 + (11)\,(0.550)^2 + (11)\,(0.846)^2]/(11 + 11 + 11 + 10 + 11 + 11)\}^{1/2}$$
$$= 1.07\ \%$$

7.2.9 The values of standard deviation (s_{a+b}) and the coefficient of variation of reproducibility from Table 12 are also examined for agreement or pattern of uniformity. The s_{a+b} values, except for the lowest fat content (sample P-1), generally increase smoothly in proportion to increasing fat content. The values of the coefficient of variation of reproducibility, except for the two lowest fat contents (samples P-1 and B-1), generally decrease in proportion to increasing fat content and do not exhibit more uniformity than standard deviation. To confirm which values of s_{a+b} can be pooled, the sample variances (s_{a+b}^2) in 7.2.4.4 are tested for homogeneity by Bartletts's chi-square test as described in 7.2.7. The homogeneous groups of sample variances of reproducibility can be pooled as shown in 7.3.2; or the sample values of coefficient of variation, grouped according to agreement, can be pooled as shown for repeatability in 7.2.8, using the respective $(m - 1)$ degrees of freedom as follows:

$$s_{a+b} \times 100/\bar{X} = \{[(10 \times 3.453^2) + (10 \times 1.905^2) + (10 \times 1.756^2) + (9 \times 1.309^2) + (10 \times 1.891^2) + (10 \times 2.126^2)]/(10 + 10 + 10 + 9 + 10 + 10)\}^{1/2} = 2.19\ \%$$

7.2.10 *Checking Limits for Duplicates*—A useful precision estimate can be obtained from the values for the duplicate determinations in the form of the permissible range for such paired determinations. The standard de-

viation for duplicates can be calculated from the original data for paired determinations as illustrated for Beef-1 in Table 15.

s (from duplicates)

$$= \sqrt{\frac{\text{sum of the squares of all differences}}{2 \times (\text{number of sets})}}$$

$$= \sqrt{\frac{0.6601}{2 \times 22}}$$

= 0.12, based on 22 degrees of freedom

7.2.10.1 The data for the other six samples are analyzed similarly, after eliminating outliers (6.3.3). These operations are not illustrated but the results are summarized in Table 16. As was the case in 7.2.6, the sample variances are tested for homogeneity by Bartlett's test. The chi-square (adjusted) value for the 7-sample variances is 14.2. This exceeds the tabular value of 12.6 (P = 0.05, 6 DF) and indicates significant inhomogeneity. Excluding the lowest variance (sample P-1), a chi-square (adjusted) value of 6.1 is obtained, which does not exceed the tabular value of 11.1 (P = 0.05, 5 DF). This indicates that the sample variances can be pooled for six samples (excluding sample P-1) using the respective (k) (m) (n) $(r - 1)$ degrees of freedom for weighting as shown in the example in 7.3.2. Similarly, a pooled value of the coefficient of variation between duplicates of the same six samples can be calculated as follows:

Pooled coefficient of variation

$$= \{[(22 \times 1.1509^2) + (22 \times 0.8231^2) + (22 \times 0.5200^2) + (23 \times 0.3331^2) + (23 \times 0.7251^2) + (24 \times 0.7290^2)] / (22 + 22 + 22 + 23 + 23 + 24)\}^{1/2}$$

$$= \sqrt{0.5691} = 0.75\ \%$$

7.2.11 *Degrees of Freedom*—Calculation of the exact number of degrees of freedom applicable to the pooled coefficient of variation (or to the pooled standard deviation) is a complex procedure which is beyond the scope of this recommended practice. To permit making predictions concerning the reproducibility in a universe of laboratories based on a study among m laboratories, a conservative estimate of $(m - 1)$ degrees of freedom is used. For an estimate of the repeatability of the method, the available degrees of freedom can be approximated from the following equation:

DF = k materials or levels × m laboratories × $(n - 1)$ days

7.2.11.1 In view of the fact that tests for outlying observations may reject some data and result in different values of m for each level of material, it is more correct to calculate the total degrees of freedom by adding the DF values for the pertinent levels. For the example cited, the between-days, within-laboratories DF values of Table 12 are used. With regard to checking limits for duplicates, the available DF can be approximated as follows:

DF = k materials or levels × m laboratories × n days × $(r - 1)$ multiples

For the reasons given in 7.2.11, it is also more correct to total the DF values for the applicable levels, as shown in Table 16.

7.2.12 *Calculation of Precision Estimates*—The following precision estimates should be calculated from the pertinent coefficients of variation, previously described, as illustrated below:

7.2.12.1 *Checking Limits for Duplicates (95 % Confidence Level)*—Multiply the coefficient of variation for duplicate runs by the factor for the applicable degrees of freedom obtained from Table 17 **(11, 12, 13)**. For the range of two results, these factors can be calculated as follows:

$$\text{Factor} = \sqrt{2} \times t_{0.05}$$

For the example cited in 7.2.10.1, where $(s \times 100)/\bar{X} = 0.75$ % and DF = (22 + 22 + 22 + 23 + 23 + 24) = 136; 0.75 × 2.79 = 2.11 % relative, for the fat content range of 10.8 to 47.8 %, which is the maximum range for duplicate values acceptable at a 95 % confidence level.

7.2.12.2 *Repeatability (95 % Confidence Level)*—Similarly, multiply the overall coefficient of variation for the between-days, within-laboratories data by the indicated factor. For the example in 7.2.8, where $(s_a \times 100)/\bar{X} = 1.07$ % and DF = (11 + 11 + 11 + 10 + 11 + 11 + 11) = 65; 1.07 × 2.82 = 3.02 % relative, for the fat content range of 10.8 to 49 %, the maximum range between two values (each the average of duplicates obtained by the same analyst on different days) acceptable at a 95 % confidence level.

7.2.12.3 *Reproducibility*—These values