~~Designation:D6300–07~~ Designation: D 6300 – 08

An American National Standard

Standard Practice for Determination of Precision and Bias Data for Use in Test Methods for Petroleum Products and Lubricants[1]

This standard is issued under the fixed designation D 6300; 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.

INTRODUCTION

Both Research Report RR:D02–1007,[2] *Manual on Determining Precision Data for ASTM Methods on Petroleum Products and Lubricants*[2] and the ISO 4259, benefitted greatly from more than 50 years of collaboration between ASTM and the Institute of Petroleum (IP) in the UK. The more recent work was documented by the IP and has become ISO 4259.

ISO 4259 encompasses both the determination of precision and the application of such precision data. In effect, it combines the type of information in RR:D02–1007[2] regarding the determination of the precision estimates and the type of information in Practice D 3244 for the utilization of test data. The following practice, intended to replace RR:D02–1007,[2] ~~differs slightly from related portions of the ISO standard. This new practice is consistent with the computer software, ADJD6300 D2PP, Version 4.43, Determination of Precision and Bias Data for Use in Test Methods for Petroleum Products.~~ differs slightly from related portions of the ISO standard.

1. Scope

1.1 This practice covers the necessary preparations and planning for the conduct of interlaboratory programs for the development of estimates of precision (determinability, repeatability, and reproducibility) and of bias (absolute and relative), and further presents the standard phraseology for incorporating such information into standard test methods.

1.2 This practice is generally limited to homogeneous products with which serious sampling problems do not normally arise.

1.3 This practice may not be suitable for solid or semisolid products such as petroleum coke, industrial pitches, paraffin waxes, greases, or solid lubricants when the heterogeneous properties of the substances create sampling problems. In such instances, use Practice E 691 or consult a trained statistician.

~~1.4A software program (ADJD6300) performs the necessary computations prescribed by this practice.~~

2. Referenced Documents

2.1 *ASTM Standards:*[3]

D 123 Terminology Relating to Textiles

D 3244 Practice for Utilization of Test Data to Determine Conformance with Specifications

E 29 Practice for Using Significant Digits in Test Data to Determine Conformance with Specifications

E 456 Terminology Relating to Quality and Statistics

E 691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method

2.2 *ISO Standards:*

ISO 4259 Petroleum Products-Determination and Application of Precision Data in Relation to Methods of Test

~~2.3ASTM Adjuncts:~~

~~ADJD6300D2PP, Version 4.43, Determination of Precision and Bias Data for Use in Test Methods for Petroleum Products[3, 4]~~

---

[1] This practice is under the jurisdiction of ASTM Committee D02 on Petroleum Products and Lubricants and is the direct responsibility of Subcommittee D02.94 on Coordinating Subcommittee on Quality Assurance and Statistics.

~~Current edition approved July 1, 2007. Published September 2007. Originally approved in 1998. Last previous edition approved in 2006 as D6300–06.~~

Current edition approved Dec. 15, 2008. Published February 2009. Originally approved in 1998. Last previous edition approved in 2007 as D 6300–07.

[2] Supporting data have been filed at ASTM International Headquarters and may be obtained by requesting Research Report RR:D02–1007.

~~[3] Available from ASTM International Headquarters. Order Adjunct No. ADJD6300.~~

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

~~[4] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.~~

3. Terminology

3.1 *Definitions:*

3.1.1 *analysis of variance (ANOVA)*, *n*—a procedure for dividing the total variation of a set of data into two or more parts, one of which estimates the error due to selecting and testing specimens and the other part(s) possible sources of added variation. **D 123**

3.1.2 *bias*, *n*—the difference between the population mean of the test results and an accepted reference value. **E 456**

3.1.3 *bias, relative*, *n*—the difference between the population mean of the test results and an accepted reference value, which is the agreed upon value obtained using an accepted reference method for measuring the same property.

3.1.4 *degrees of freedom*, *n*—the divisor used in the calculation of variance.

3.1.4.1 *Discussion*—This definition applies strictly only in the simplest cases. Complete definitions are beyond the scope of this practice. **ISO 4259**

3.1.5 *determinability*, *n*—a quantitative measure of the variability associated with the same operator in a given laboratory obtaining successive determined values using the same apparatus for a series of operations leading to a single result; it is defined as that difference between two such single determined values as would be exceeded in the long run in only one case in 20 in the normal and correct operation of the test method.

3.1.5.1 *Discussion*—This definition implies that two determined values, obtained under determinability conditions, which differ by more than the determinability value should be considered suspect. If an operator obtains more than two determinations, then it would usually be satisfactory to check the most discordant determination against the mean of the remainder, using determinability as the critical difference **(1)**.[5]

3.1.6 *mean square*, *n— in analysis of variance*, a contraction of the expression "mean of the squared deviations from the appropriate average(s)" where the divisor of each sum of squares is the appropriate degrees of freedom. **D 123**

3.1.7 *normal distribution*, *n*—the distribution that has the probability function:

$$f(x) = (1/\sigma)(2\pi)^{-1/2}\exp\,[-(x-\mu)^2/2\sigma^2] \quad (1)$$

where:

x = a random variate,
μ = the mean distribution, and
σ = the standard deviation of the distribution.

(*Syn.* Gaussian distribution, law of error) **D 123**

3.1.8 *outlier*, *n*—a result far enough in magnitude from other results to be considered not a part of the set. **RR:D02–1007[2]**

3.1.9 *precision*, *n*—the degree of agreement between two or more results on the same property of identical test material. In this practice, precision statements are framed in terms of *repeatability* and *reproducibility* of the test method.

3.1.9.1 *Discussion*—The testing conditions represented by repeatability and reproducibility should reflect the normal extremes of variability under which the test is commonly used. Repeatability conditions are those showing the least variation; reproducibility, the usual maximum degree of variability. Refer to the definitions of each of these terms for greater detail. **RR:D02–1007[2]**

3.1.10 *random error*, *n*—the chance variation encountered in all test work despite the closest control of variables. **RR:D02–1007[2]**

3.1.11 *repeatability*, *n*—the quantitative expression of the random error associated with the same operator in a given laboratory obtaining repetitive results by applying the same test method with the same apparatus under constant operating conditions on identical test material within short intervals of time. It is defined as the difference between two such results at the 95 % confidence level. **RR:D02–1007[2]**

3.1.11.1 *Discussion*—Interpret as the value equal to or below which the absolute difference between two single test results obtained in the above conditions may expect to lie with a probability of 95 %. **ISO 4259**

3.1.11.2 *Discussion*—The difference is related to the repeatability standard deviation but it is not the standard deviation or its estimate. **RR:D02–1007[2]**

3.1.12 *reproducibility*, *n*—a quantitative expression of the random error associated with different operators from different laboratories using different apparatus, each obtaining a single result by applying the same test method on an identical test sample. It is defined as the 95 % confidence limit for the difference between two such single and independent results.

3.1.12.1 *Discussion*—Interpret as the value equal to or below which the absolute difference between two single test results on identical material obtained by operators in different laboratories, using the standardized test, may be expected to lie with a probability of 95 %. **ISO 4259**

3.1.12.2 *Discussion*—The difference is related to the reproducibility standard deviation but is not the standard deviation or its estimate. **RR:D02–1007[2]**

[4] Available from International Organization for Standardization, 1 rue de Varembé, Case postale 56, CH-1211 Geneva 20, Switzerland.
[5] Available from International Organization for Standardization, 1 rue de Varembé, Case postale 56, CH-1211 Geneva 20, Switzerland.
[5] The bold numbers in parentheses refers to the list of references at the end of this standard.

3.1.12.3 *Discussion*—In those cases where the normal use of the test method does not involve sending a sample to a testing laboratory, either because it is an in-line test method or because of serious sample instabilities or similar reasons, the precision test for obtaining reproducibility may allow for the use of apparatus from the participating laboratories at a common site (several common sites, if feasible). The statistical analysis is not affected thereby. However, the interpretation of the reproducibility value will be affected, and therefore, the precision statement shall, in this case, state the conditions to which the reproducibility value applies.

3.1.13 *standard deviation*, *n*—the most usual measure of the dispersion of observed values or results expressed as the positive square root of the variance. **E 456**

3.1.14 *sum of squares*, *n*— *in analysis of variance*, a contraction of the expression "sum of the squared deviations from the appropriate average(s)" where the average(s) of interest may be the average(s) of specific subset(s) of data or of the entire set of data. **D 123**

3.1.15 *variance*, *n*—a measure of the dispersion of a series of accepted results about their average. It is equal to the sum of the squares of the deviation of each result from the average, divided by the number of degrees of freedom. **RR:D02–1007[2]**

3.1.16 *variance, between-laboratory*, *n*—that component of the overall variance due to the difference in the mean values obtained by different laboratories. **ISO 4259**

3.1.16.1 *Discussion*—When results obtained by more than one laboratory are compared, the scatter is usually wider than when the same number of tests are carried out by a single laboratory, and there is some variation between means obtained by different laboratories. Differences in operator technique, instrumentation, environment, and sample "as received" are among the factors that can affect the between laboratory variance. There is a corresponding definition for between-operator variance.

3.1.16.2 *Discussion*—The term "between-laboratory" is often shortened to "laboratory" when used to qualify representative parameters of the dispersion of the population of results, for example as "laboratory variance."

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *determination*, *n*—the process of carrying out a series of operations specified in the test method whereby a single value is obtained.

3.2.2 *operator*, *n*—a person who carries out a particular test.

3.2.3 *probability density function*, *n*—function which yields the probability that the random variable takes on any one of its admissible values; here, we are interested only in the normal probability.

3.2.4 *result*, *n*—the final value obtained by following the complete set of instructions in the test method.

3.2.4.1 *Discussion*—It may be obtained from a single determination or from several determinations, depending on the instructions in the method. When rounding off results, the procedures described in Practice E 29 shall be used.

4. Summary of Practice

4.1 A draft of the test method is prepared and a pilot program can be conducted to verify details of the procedure and to estimate roughly the precision of the test method.

4.2 A plan is developed for the interlaboratory study using the number of participating laboratories to determine the number of samples needed to provide the necessary degrees of freedom. Samples are acquired and distributed. The interlaboratory study is then conducted on an agreed draft of the test method.

4.3 The data are summarized and analyzed. Any dependence of precision on the level of test result is removed by transformation. The resulting data are inspected for uniformity and for outliers. Any missing and rejected data are estimated. The transformation is confirmed. Finally, an analysis of variance is performed, followed by calculation of repeatability, reproducibility, and bias. When it forms a necessary part of the test procedure, the determinability is also calculated.

5. Significance and Use

5.1 ASTM test methods are frequently intended for use in the manufacture, selling, and buying of materials in accordance with specifications and therefore should provide such precision that when the test is properly performed by a competent operator, the results will be found satisfactory for judging the compliance of the material with the specification. Statements addressing precision and bias are required in ASTM test methods. These then give the user an idea of the precision of the resulting data and its relationship to an accepted reference material or source (if available). Statements addressing determinability are sometimes required as part of the test method procedure in order to provide early warning of a significant degradation of testing quality while processing any series of samples.

5.2 Repeatability and reproducibility are defined in the precision section of every Committee D02 test method. Determinability is defined above in Section 3. The relationship among the three measures of precision can be tabulated in terms of their different sources of variation (see Table 1).

5.2.1 When used, determinability is a mandatory part of the Procedure section. It will allow operators to check their technique for the sequence of operations specified. It also ensures that a result based on the set of determined values is not subject to excessive variability from that source.

5.3 A bias statement furnishes guidelines on the relationship between a set of test results and a related set of accepted reference values. When the bias of a test method is known, a compensating adjustment can be incorporated in the test method.

5.4 This practice is intended for use by D02 subcommittees in determining precision estimates and bias statements to be used

TABLE 1 Sources of Variation

	Method	Apparatus	Operator	Laboratory	Time
Reproducibility	Complete (Result)	Different	Different	Different	Specified
Repeatability	Complete (Result)	Same	Same	Same	Almost same
Determinability	Incomplete (Part result)	Same	Same	Same	Almost same

in D02 test methods. Its procedures correspond with ISO 4259 and are the basis for the Committee D02 computer software, *Calculation if Precision Data: Petroleum Test Methods*. The use of this practice replaces that of Research Report RR:D02–1007.[2]

5.5 Standard practices for the calculation of precision have been written by many committees with emphasis on their particular product area. One developed by Committee E11 on Statistics is Practice E 691. Practice E 691 and this practice differ as outlined in Table 2.

6. Stages in Planning of an Interlaboratory Test Program for the Determination of the Precision of a Test Method

6.1 The stages in planning an interlaboratory test program are: preparing a draft method of test (see 6.2), planning and executing a pilot program with at least two laboratories (optional but recommended for new test methods) (see 6.3), planning the interlaboratory program (see 6.4), and executing the interlaboratory program (see 6.5). The four stages are described in turn.

6.2 *Preparing a Draft Method of Test*—This shall contain all the necessary details for carrying out the test and reporting the results. Any condition which could alter the results shall be specified. The section on precision will be included at this stage only as a heading.

TABLE 2 Differences in Calculation of Precision in Practices D 6300 and E 691

Element	This Practice	Practice E 691
Applicability	Limited in general to homogeneous samples for which serious sampling problems do not normally arise.	Permits heterogeneous samples.
Number of duplicates	Two	Any number
Precision is written for	Test method	Each sample
Outlier tests:	Sequential	Simultaneous
Within laboratories	Cochran test	*k*-value
Between laboratories	Hawkins test	*h*-value
Outliers	Rejected, subject to subcommittee approval.	Rejected if many laboratories or for cause such as blunder or not following method.
	Retesting not generally permitted.	Laboratory may retest sample having rejected data.
Rejection limit	20 %	5 %
Analysis of variance	Two-way, applied globally to all the remaining data at once.	One-way, applied to each sample separately.
Precision multiplier	$t\sqrt{2}$, where t is the two-tailed Student's t for 95 % probability.	$2.8=1.96\sqrt{2}$
	Increases with decreasing laboratories × samples particularly below 12.	Constant.
Variation of precision with level	Minimized by data transformation. Equations for repeatability and reproducibility are generated in the retransformation process.	User may assess from individual sample precisions.

6.3 *Planning and Executing a Pilot Program with at Least Two Laboratories*:

6.3.1 A pilot program is recommended to be used with new test methods for the following reasons: (*1*) to verify the details in the operation of the test; (*2*) to find out how well operators can follow the instructions of the test method; (*3*) to check the precautions regarding sample handling and storage; and (*4*) to estimate roughly the precision of the test.

6.3.2 At least two samples are required, covering the range of results to which the test is intended to apply; however, include at least 12 laboratory-sample combinations. Test each sample twice by each laboratory under repeatability conditions. If any omissions or inaccuracies in the draft method are revealed, they shall now be corrected. Analyze the results for precision, bias, and determinability (if applicable) using this practice. If any are considered to be too large for the technical application, then consider alterations to the test method.

6.4 *Planning the Interlaboratory Program*:

6.4.1 There shall be at least five participating laboratories, but it is preferable to exceed this number in order to reduce the number of samples required and to make the precision statement as representative as possible of the qualified user population. In the absence of pilot test program information to permit use of Fig. 1 (see 6.4.3) to determine the number of laboratories, the minimum number of laboratories shall be six.

6.4.2 The number of samples shall be sufficient to cover the range of the property measured, and to give reliability to the precision estimates. If any variation of precision with level was observed in the results of the pilot program, then at least five samples shall be used in the interlaboratory program. In any case, it is necessary to obtain at least 30 degrees of freedom in both repeatability and reproducibility. For repeatability, this means obtaining a total of at least 30 pairs of results in the program. In the absence of pilot test program information to permit use of Fig. 1 (see 6.4.3) to determine the number of samples, the number of samples shall be greater than five, and chosen such that the number of laboratories times the number of samples is greater than or equal to 42.

6.4.3 For reproducibility, Fig. 1 gives the minimum number of samples required in terms of L, P , and Q, where L is the number of participating laboratories, and P and Q are the ratios of variance component estimates (see 8.3.1) obtained from the pilot program. Specifically, P is the ratio of the interaction component to the repeats component, and Q is the ratio of the laboratories component to the repeats component.

NOTE 1—Appendix X1 gives the derivation of the equation used. If Q is much larger than P, then 30 degrees of freedom cannot be achieved; the blank entries in Fig. 1 correspond to this situation or the approach of it (that is, when more than 20 samples are required). For these cases, there is likely to be a significant bias between laboratories. The program organizer shall be informed; further standardization of the test method may be necessary.

6.5 *Executing the Interlaboratory Program*:

6.5.1 One person shall oversee the entire program, from the distribution of the texts and samples to the final appraisal of the results. He or she shall be familiar with the test method, but should not personally take part in the actual running of the tests.

6.5.2 The text of the test method shall be distributed to all the laboratories in time to raise any queries before the tests begin. If any laboratory wants to practice the test method in advance, this shall be done with samples other than those used in the program.

6.5.3 The samples shall be accumulated, subdivided, and distributed by the organizer, who shall also keep a reserve of each sample for emergencies. It is most important that the individual laboratory portions be homogeneous. Instructions to each laboratory shall include the following:

6.5.3.1 The agreed draft method of test;

6.5.3.2 Material Safety Data Sheets, where applicable, and the handling and storage requirements for the samples;

6.5.3.3 The order in which the samples are to be tested (a different random order for each laboratory);

6.5.3.4 The statement that two test results are to be obtained in the shortest practical period of time on each sample by the same operator with the same apparatus. For statistical reasons it is imperative that the two results are obtained independently of each other, that is, that the second result is not biased by knowledge of the first. If this is regarded as impossible to achieve with the operator concerned, then the pairs of results shall be obtained in a blind fashion, but ensuring that they are carried out in a short period of time (preferably the same day). The term *blind fashion* means that the operator does not know that the sample is a duplicate of any previous run.

6.5.3.5 The period of time during which repeated results are to be obtained and the period of time during which all the samples are to be tested;

6.5.3.6 A blank form for reporting the results. For each sample, there shall be space for the date of testing, the two results, and any unusual occurrences. The unit of accuracy for reporting the results shall be specified. This should be, if possible, more digits reported than will be used in the final test method, in order to avoid having rounding unduly affect the estimated precision values.

6.5.3.7 When it is required to estimate the determinability, the report form must include space for each of the determined values as well as the test results.

6.5.3.8 A statement that the test shall be carried out under normal conditions, using operators with good experience but not exceptional knowledge; and that the duration of the test shall be the same as normal.

6.5.4 The pilot program operators may take part in the interlaboratory program. If their extra experience in testing a few more samples produces a noticeable effect, it will serve as a warning that the test method is not satisfactory. They shall be identified in the report of the results so that any such effect may be noted.

6.5.5 It can not be overemphasized that the statement of precision in the test method is to apply to test results obtained by

L = number of participating laboratories

P = interaction variance component/repeats variance component

Q = laboratories variance component/repeats variance component

L=5

P:	Q: 0	1	2	3	4	5	6	7	8	9
0	4									
1	5									
2	6	11								
3	6	9								
4	7	8	16							
5	7	8	12							
6	7	8	11	19						
7	7	8	10	15						
8	7	8	9	13						
9	7	8	9	11	17					

L=6

P:	Q: 0	1	2	3	4	5	6	7	8	9
0	3									
1	4	11								
2	5	7								
3	5	7	14							
4	5	6	10							
5	6	6	8	15						
6	6	6	8	11						
7	6	6	7	10	15					
8	6	6	7	9	12					
9	6	6	7	8	10	15				

L=7

P:	Q: 0	1	2	3	4	5	6	7	8	9
0	4									
1	5									
2	6	11								
3	6	9								
4	7	8	16							
5	7	8	12							
6	7	8	11	19						
7	7	8	10	15						
8	7	8	9	13						
9	7	8	9	11	17					

L=8

P:	Q: 0	1	2	3	4	5	6	7	8	9
0	3									
1	3	5								
2	4	5	9							
3	4	5	7	14						
4	4	4	6	9	20					
5	4	4	5	7	11					
6	4	4	5	6	8	13				
7	4	4	5	6	7	10	16			
8	4	5	5	6	6	8	11	18		
9	4	5	5	5	6	7	9	13		

L=9

P:	Q: 0	1	2	3	4	5	6	7	8	9
0	2									
1	3	4								
2	3	4	7							
3	3	4	5	9						
4	4	4	5	6	11					
5	4	4	5	6	7	12				
6	4	4	4	5	6	9	14			
7	4	4	4	5	6	7	10	15		
8	4	4	4	5	5	6	8	10	16	
9	4	4	4	5	5	6	7	8	11	18

L=10

P:	Q: 0	1	2	3	4	5	6	7	8	9
1	2	8								
1	3	4	11							
2	3	4	5	12						
3	3	3	4	6	13					
4	3	4	4	5	7	14				
5	3	4	4	5	6	8	14			
6	3	4	4	4	5	6	9	14		
7	3	4	4	4	5	6	7	9	14	
8	3	4	4	4	5	5	6	7	10	14
9	4	4	4	4	4	5	6	6	8	10

L=11

P:	Q: 0	1	2	3	4	5	6	7	8	9
0	2	4								
1	2	3	5							
2	3	3	3	7						
3	3	3	4	5	8					
4	3	3	4	4	6	8	18			
5	3	3	4	4	5	6	9	15		
6	3	3	3	4	4	5	6	9	14	
7	3	3	3	4	4	5	5	7	9	13
8	3	3	3	4	4	4	5	6	7	9
9	3	3	3	4	4	4	5	5	6	7

L=12

P:	Q: 0	1	2	3	4	5	6	7	8	9
0	2	4								
1	2	3	5							
2	2	3	4	6	14					
3	3	3	3	4	6	11				
4	3	3	3	4	5	6	9			
5	3	3	3	4	4	5	6	9	16	
6	3	3	3	3	4	4	5	6	9	13
7	3	3	3	3	3	4	4	5	5	6	8
8	3	3	3	3	3	4	4	4	5	5	6
9	3	3	3	3	3	3	4	4	4	5	6

L=13

P:	Q: 0	1	2	3	4	5	6	7	8	9
0	2	3								
1	2	3	4	12						
2	2	3	3	4	8					
3	2	3	3	4	5	7	14			
4	3	3	3	3	4	5	7	10		
5	3	3	3	3	4	4	5	6	9	15
6	3	3	3	3	4	4	4	5	6	8
7	3	3	3	3	3	4	4	4	5	6
8	3	3	3	3	3	3	4	4	4	5
9	3	3	3	3	3	3	4	4	4	5

L=14

P:	Q: 0	1	2	3	4	5	6	7	8	9
0	2	3								
1	2	2	3	7						
2	2	2	3	4	6	12				
3	2	2	3	3	4	5	8	18		
4	2	3	3	3	3	4	5	7	11	
5	2	3	3	3	3	4	4	5	6	8
6	3	3	3	3	3	3	4	4	5	6
7	3	3	3	3	3	3	3	4	4	5
8	3	3	3	3	3	3	3	4	4	4
9	3	3	3	3	3	3	3	3	4	4

L=15

P:	Q: 0	1	2	3	4	5	6	7	8	9
0	2	2	13							
1	2	2	3	5	19					
2	2	2	3	3	4	7				
3	2	2	3	3	3	4	6	9		
4	2	2	3	3	4	4	5	7	10	
5	2	2	3	3	3	3	4	4	5	6
6	2	2	3	3	3	3	3	4	4	5
7	2	2	3	3	3	3	3	3	4	4
8	2	2	3	3	3	3	3	3	3	4
9	2	2	3	3	3	3	3	3	3	3

L=16

P:	Q: 0	1	2	3	4	5	6	7	8	9
0	2	5								
1	2	2	3	4	8					
2	2	2	2	3	4	5	9			
3	2	2	2	3	3	4	4	6	9	
4	2	2	2	3	3	3	4	4	5	6
5	2	2	2	3	3	3	3	4	4	5
6	2	2	2	2	3	3	3	3	4	4
7	2	2	2	2	3	3	3	3	3	4
8	2	2	2	2	3	3	3	3	3	3
9	2	2	2	2	3	3	3	3	3	3

FIG. 1 Determination of Number of Samples Required (see 6.4.3)

running the agreed procedure exactly as written. Therefore, the test method must not be significantly altered after its precision statement is written.

7. Inspection of Interlaboratory Results for Uniformity and for Outliers

7.1 *Introduction*:

7.1.1 This section specifies procedures for examining the results reported in a statistically designed interlaboratory program (see Section 6) to establish:

7.1.1.1 The independence or dependence of precision and the level of results;

7.1.1.2 The uniformity of precision from laboratory to laboratory, and to detect the presence of outliers.

NOTE 2—The procedures are described in mathematical terms based on the notation of Annex A1 and illustrated with reference to the example data (calculation of bromine number) set out in Annex A2. Throughout this section (and Section 8), the procedures to be used are first specified and then illustrated by a worked example using data given in Annex A2.

NOTE 3—It is assumed throughout this section that all the deviations are either from a single normal distribution or capable of being transformed into such a distribution (see 7.2~~). In ADJD6300 (D2PP),[3] this can be visually examined on a sample-by-sample basis using the "probability" plot in the VIEW DATA screen.Other cases (which are rare) would require different treatment that is beyond the scope of this practice. Also, see~~ **). Other cases (which are rare) would require different treatment that is beyond the scope of this practice. Also, see (2) for a statistical test of normality. Note4—Although the procedures shown here are in a form suitable for hand calculation, it is strongly advised that an electronic computer be used to store and analyze interlaboratory test results, based on the procedures of this practice. ADJD6300 D2PP, Version 4.43, Determination of Precision and Bias Data for Use in Test Methods for Petroleum Products,[3] has been designed for this purpose.**

7.2 *Transformation of Data*:

7.2.1 In many test methods the precision depends on the level of the test result, and thus the variability of the reported results is different from sample to sample. The method of analysis outlined in this practice requires that this shall not be so and the position is rectified, if necessary, by a transformation.

7.2.2 The laboratories' standard deviations D_j, and the repeats standard deviations d_j (see Annex A1) are calculated and plotted separately against the sample means m_j. If the points so plotted may be considered as lying about a pair of lines parallel to the m-axis, then no transformation is necessary. If, however, the plotted points describe non-horizontal straight lines or curves of the form $D = f_1(m)$ and $d = f_2(m)$, then a transformation will be necessary.

7.2.3 The relationships $D = f_1(m)$ and $d = f_2(m)$ will not in general be identical. It is frequently the case, however, that the ratios $u_j = d_j / D_j$ are approximately the same for all m_j, in which case f_1 is approximately proportional to f_2 and a single transformation will be adequate for both repeatability and reproducibility. The statistical procedures of this practice are greatly facilitated when a single transformation can be used. For this reason, unless the u_j clearly vary with property level, the two relationships are combined into a single dependency relationship $D = f(m)$ (where D now includes d) by including a dummy variable T. This will take account of the difference between the relationships, if one exists, and will provide a means of testing for this difference (see A4.1).

7.2.4 In the event that the rations u_j do vary with level (mean, m_j), as confirmed with a regression of u_j on m_j, or $\log(u_j)$ on $\log(m_j)$, follow the instructions in Annex A5. Otherwise, continue with 7.2.5.

7.2.5 The single relationship $D = f(m)$ is best estimated by weighted linear regression analysis. Strictly speaking, an iteratively weighted regression should be used, but in most cases even an unweighted regression will give a satisfactory approximation. The derivation of weights is described in A4.2, and the computational procedure for the regression analysis is described in A4.3. Typical forms of dependence $D = f(m)$ are given in A3.1. These are all expressed in terms of at most two (2) transformation parameters, B and B_0.

7.2.6 The typical forms of dependence, the transformations they give rise to, and the regressions to be performed in order to estimate the transformation parameters B, are all summarized in A3.2. This includes statistical tests for the significance of the regression (that is, is the relationship $D = f(m)$ parallel to the m-axis), and for the difference between the repeatability and reproducibility relationships, based at the 5 % significance level. If such a difference is found to exist, follow the procedures in Annex A5.

7.2.7 If it has been shown at the 5 % significance level that there is a significant regression of the form $D = f(m)$, then the appropriate transformation $y = F(x)$, where x is the reported result, is given by the equation

$$F(x) = K\int \frac{dx}{f(x)} \tag{2}$$

where K = a constant. In that event, all results shall be transformed accordingly and the remainder of the analysis carried out in terms of the transformed results. Typical transformations are given in A3.1.

7.2.8 The choice of transformation is difficult to make the subject of formalized rules. Qualified statistical assistance may be required in particular cases. The presence of outliers may affect judgement as to the type of transformation required, if any (see 7.7).

7.2.9 *Worked Example*:

7.2.9.1 Table 3 lists the values of m, D, and d for the eight samples in the example given in Annex A2, correct to three significant digits. Corresponding degrees of freedom are in parentheses. Inspection of the values in Table 3 shows that both D and d increase with m, the rate of increase diminishing as m increases. A plot of these figures on log-log paper (that is, a graph of log D and log d against log m) shows that the points may reasonably be considered as lying about two straight lines (see Fig. A4.1 in Annex A4).

TABLE 3 Computed from Bromine Example Showing Dependence of Precision on Level

Sample Number	3	8	1	4	5	6	2	7
m	0.756	1.22	2.15	3.64	10.9	48.2	65.4	114
D	0.0669 (14)	0.159 (9)	0.729 (8)	0.211 (11)	0.291 (9)	1.50 (9)	2.22 (9)	2.93 (9)
d	0.0500 (9)	0.0572 (9)	0.127 (9)	0.116 (9)	0.0943 (9)	0.527 (9)	0.818 (9)	0.935 (9)

From the example calculations given in A4.4, the gradients of these lines are shown to be the same, with an estimated value of 0.638. Bearing in mind the errors in this estimated value, the gradient may for convenience be taken as 2/3.

$$\int x^{-\frac{2}{3}}\,dx = 3x^{\frac{1}{3}} \tag{3}$$

7.2.9.2 Hence, the same transformation is appropriate both for repeatability and reproducibility, and is given by the equation. Since the constant multiplier may be ignored, the transformation thus reduces to that of taking the cube roots of the reported bromine numbers. This yields the transformed data shown in Table A1.3, in which the cube roots are quoted correct to three decimal places.

7.3 *Tests for Outliers*:

7.3.1 The reported data or, if it has been decided that a transformation is necessary, the transformed results shall be inspected for outliers. These are the values which are so different from the remainder that it can only be concluded that they have arisen from some fault in the application of the test method or from testing a wrong sample. Many possible tests may be used and the associated significance levels varied, but those that are specified in the following subsections have been found to be appropriate in this practice. These outlier tests all assume a normal distribution of errors.

7.3.2 *Uniformity of Repeatability* —The first outlier test is concerned with detecting a discordant result in a pair of repeat results. This test **(3)** involves calculating the $e_{ij}{}^2$ over all the laboratory/sample combinations. Cochran's criterion at the 1 % significance level is then used to test the ratio of the largest of these values over their sum (see A1.5). If its value exceeds the value given in Table A2.2, corresponding to one degree of freedom, n being the number of pairs available for comparison, then the member of the pair farthest from the sample mean shall be rejected and the process repeated, reducing n by 1, until no more rejections are called for. In certain cases, specifically when the number of digits used in reporting results leads to a large number of repeat ties, this test can lead to an unacceptably large proportion of rejections, for example, more than 10 %. If this is so, this rejection test shall be abandoned and some or all of the rejected results shall be retained. A decision based on judgement will be necessary in this case.

7.3.3 *Worked Example*— In the case of the example given in Annex A2, the absolute differences (ranges) between transformed repeat results, that is, of the pairs of numbers in Table A1.3, in units of the third decimal place, are shown in Table 4. The largest range is 0.078 for Laboratory G on Sample 3. The sum of squares of all the ranges is

$$0.042^2 + 0.021^2 + \ldots + 0.026^2 + 0^2 = 0.0439.$$

Thus, the ratio to be compared with Cochran's criterion is

$$\frac{0.078^2}{0.0439} = 0.138 \tag{4}$$

where 0.138 is the result obtained by electronic calculation of unrounded factors in the expression. There are 72 ranges and as, from Table A2.2, the criterion for 80 ranges is 0.1709, this ratio is not significant.

7.3.4 *Uniformity of Reproducibility* :

7.3.4.1 The following outlier tests are concerned with establishing uniformity in the reproducibility estimate, and are designed to detect either a discordant pair of results from a laboratory on a particular sample or a discordant set of results from a laboratory on all samples. For both purposes, the Hawkins' test **(4)** is appropriate.

7.3.4.2 This involves forming for each sample, and finally for the overall laboratory averages (see 7.6), the ratio of the largest absolute deviation of laboratory mean from sample (or overall) mean to the square root of certain sums of squares (A1.6).

7.3.4.3 The ratio corresponding to the largest absolute deviation shall be compared with the critical 1 % values given in Table A1.5, where n is the number of laboratory/sample cells in the sample (or the number of overall laboratory means) concerned and where v is the degrees of freedom for the sum of squares which is additional to that corresponding to the sample in question. In the test for laboratory/sample cells v will refer to other samples, but will be zero in the test for overall laboratory averages.

7.3.4.4 If a significant value is encountered for individual samples the corresponding extreme values shall be omitted and the process repeated. If any extreme values are found in the laboratory totals, then all the results from that laboratory shall be rejected.

7.3.4.5 If the test leads to an unacceptably large proportion of rejections, for example, more than 10 %, then this rejection test

TABLE 4 Absolute Differences Between Transformed Repeat Results: Bromine Example

Laboratory	Sample							
	1	2	3	4	5	6	7	8
A	42	21	7	13	7	10	8	0
B	23	12	12	0	7	9	3	0
C	0	6	0	0	7	8	4	0
D	14	6	0	13	0	8	9	32
E	65	4	0	0	14	5	7	28
F	23	20	34	29	20	30	43	0
G	62	4	78	0	0	16	18	56
H	44	20	29	44	0	27	4	32
J	0	59	0	40	0	30	26	0

shall be abandoned and some or all of the rejected results shall be retained. A decision based on judgement will be necessary in this case.

7.3.5 *Worked Example*:

7.3.5.1 The application of Hawkins' test to cell means within samples is shown below.

7.3.5.2 The first step is to calculate the deviations of cell means from respective sample means over the whole array. These are shown in Table 5, in units of the third decimal place. The sum of squares of the deviations are then calculated for each sample. These are also shown in Table 5 in units of the third decimal place.

7.3.5.3 The cell to be tested is the one with the most extreme deviation. This was obtained by Laboratory D from Sample 1. The appropriate Hawkins' test ratio is therefore:

$$B^* = \frac{0.314}{\sqrt{0.117 + 0.015 + \ldots + 0.017}} = 0.7281 \quad (5)$$

7.3.5.4 The critical value, corresponding to $n = 9$ cells in sample 1 and $v = 56$ extra degrees of freedom from the other samples is interpolated from Table A1.5 as 0.3729. The test value is greater than the critical value, and so the results from Laboratory D on Sample 1 are rejected.

7.3.5.5 As there has been a rejection, the mean value, deviations, and sum of squares are recalculated for Sample 1, and the procedure is repeated. The next cell to be tested will be that obtained by Laboratory F from Sample 2. The Hawkins' test ratio for this cell is:

$$B^* = \frac{0.097}{\sqrt{0.006 + 0.015 + \ldots + 0.017}} = 0.3542 \quad (6)$$

7.3.5.6 The critical value corresponding to $n = 9$ cells in Sample 2 and $v = 55$ extra degrees of freedom is interpolated from Table A1.5 as 0.3756. As the test ratio is less than the critical value there will be no further rejections.

7.4 *Rejection of Complete Data from a Sample*:

7.4.1 The laboratories standard deviation and repeats standard deviation shall be examined for any outlying samples. If a transformation has been carried out or any rejection made, new standard deviations shall be calculated.

7.4.2 If the standard deviation for any sample is excessively large, it shall be examined with a view to rejecting the results from that sample.

7.4.3 Cochran's criterion at the 1 % level can be used when the standard deviations are based on the same number of degrees of freedom. This involves calculating the ratio of the largest of the corresponding sums of squares (laboratories or repeats, as appropriate) to their total (see A1.5). If the ratio exceeds the critical value given in Table A2.2, with n as the number of samples and v the degrees of freedom, then all the results from the sample in question shall be rejected. In such an event, care should be taken that the extreme standard deviation is not due to the application of an inappropriate transformation (see 7.1), or undetected outliers.

7.4.4 There is no optimal test when standard deviations are based on different degrees of freedom. However, the ratio of the largest variance to that pooled from the remaining samples follows an F-distribution with v_1 and v_2 degrees of freedom (see A1.7). Here v_1 is the degrees of freedom of the variance in question and v_2 is the degrees of freedom from the remaining samples. If the ratio is greater than the critical value given in A2.6, corresponding to a significance level of $0.01/S$ where S is the number of samples, then results from the sample in question shall be rejected.

7.4.5 *Worked Example*:

7.4.5.1 The standard deviations of the transformed results, after the rejection of the pair of results by Laboratory D on Sample 1, are given in Table 6 in ascending order of sample mean, correct to three significant digits. Corresponding degrees of freedom are in parentheses.

7.4.5.2 Inspection shows that there is no outlying sample among these. It will be noted that the standard deviations are now independent of the sample means, which was the purpose of transforming the results.

TABLE 5 Deviations of Cell Means from Respective Sample Means: Transformed Bromine Example

Laboratory	Sample							
	1	2	3	4	5	6	7	8
A	20	8	14	15	10	48	6	3
B	75	7	20	9	10	47	6	3
C	64	35	3	20	30	4	22	25
D	314	33	18	42	7	39	80	50
E	32	32	30	9	7	18	18	39
F	75	97	31	20	30	8	74	53
G	10	34	32	20	20	61	9	62
H	42	13	4	42	13	21	8	50
J	1	28	22	29	14	8	10	53
Sum of Squares	117	15	4	6	3	11	13	17

TABLE 6 Standard Deviations of Transformed Results: Bromine Example

Sample number	3	8	1	4	5	6	2	7
m	0.9100	1.066	1.240	1.538	2.217	3.639	4.028	4.851
D	0.0278	0.0473	0.0354	0.0297	0.0197	0.0378	0.0450	0.0416
	(14)	(9)	(13)	(11)	(9)	(9)	(9)	(9)
d	0.0214	0.0182	0.028	0.0164	0.0063	0.0132	0.0166	0.0130
	(9)	(9)	(8)	(9)	(9)	(9)	(9)	(9)

7.4.5.3 The values in Table 7, taken from a test program on bromine numbers over 100, will illustrate the case of a sample rejection.

7.4.5.4 It is clear, by inspection, that the laboratories standard deviation of Sample 93 at 15.76 is far greater than the others. It is noted that the repeats standard deviation in this sample is correspondingly large.

7.4.5.5 Since laboratory degrees of freedom are not the same over all samples, the variance ratio test is used. The variance pooled from all samples, excluding Sample 93, is the sum of the sums of squares divided by the total degrees of freedom, that is

$$\frac{(8 \times 5.10^2 + 9 \times 4.20^2 + ... + 8 \times 3.85^2)}{(8 + 9 + ... + 8)} = 19.96 \tag{7}$$

7.4.5.6 The variance ratio is then calculated as

$$\frac{15.26^2}{19.96} = 11.66 \tag{8}$$

where 11.66 is the result obtained by electronic calculation without rounding the factors in the expression.

7.4.5.7 From Table A1.8 the critical value corresponding to a significance level of 0.01/8 = 0.00125, on 8 and 63 degrees of freedom, is approximately 4. The test ratio greatly exceeds this and results from Sample 93 shall therefore be rejected.

7.4.5.8 Turning to repeats standard deviations, it is noted that degrees of freedom are identical for each sample and that Cochran's test can therefore be applied. Cochran's criterion will be the ratio of the largest sum of squares (Sample 93) to the sum of all the sums of squares, that is

$$2.97^2/(1.13^2 + 0.99^2 + ... + 1.36^2) = 0.510 \tag{9}$$

This is greater than the critical value of 0.352 corresponding to $n = 8$ and $v = 8$ (see Table A2.2), and confirms that results from Sample 93 shall be rejected.

7.5 *Estimating Missing or Rejected Values*:

7.5.1 *One of the Two Repeat Values Missing or Rejected*—If one of a pair of repeats (Y_{ij1} or Y_{ij2}) is missing or rejected, this shall be considered to have the same value as the other repeat in accordance with the least squares method.

7.5.2 *Both Repeat Values Missing or Rejected*:

7.5.2.1 If both the repeat values are missing, estimates of a_{ij} (= $Y_{ij1} + Y_{ij2}$) shall be made by forming the laboratories × samples interaction sum of squares (see Eq 17), including the missing values of the totals of the laboratories/samples pairs of results as unknown variables. Any laboratory or sample from which all the results were rejected shall be ignored and new values of *L* and *S* used. The estimates of the missing or rejected values shall be those that minimize the interaction sum of squares.

7.5.2.2 If the value of single pair sum a_{ij} has to be estimated, the estimate is given by the equation:

$$a_{ij} = \frac{1}{(L-1)\,(S'-1)}\,(LL_1 + S'S_1 - T_1) \tag{10}$$

where:

L_1 = total of remaining pairs in the *i*th laboratory,
S_1 = total of remaining pairs in the *j*th sample,
S' = S – number of samples rejected in 7.4, and
T_1 = total of all pairs except a_{ij}.

7.5.2.3 If more estimates are to be made, the technique of successive approximation can be used. In this, each pair sum is estimated in turn from Eq 10, using L_1, S_1, and T_1, values, which contain the latest estimates of the other missing pairs. Initial values for estimates can be based on the appropriate sample mean, and the process usually converges to the required level of accuracy within three complete iterations **(5)**.

TABLE 7 Example Statistics Indicating Need to Reject an Entire Sample

Sample number	90	89	93	92	91	94	95	96
m	96.1	99.8	119.3	125.4	126.0	139.9	139.4	159.5
D	5.10	4.20	15.26	4.40	4.09	4.87	4.74	3.85
	(8)	(9)	(8)	(11)	(10)	(8)	(9)	(8)
d	1.13	0.99	2.97	0.91	0.73	1.32	1.12	1.36
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)

7.5.3 *Worked Example*:

7.5.3.1 The two results from Laboratory D on Sample 1 were rejected (see 7.3.4) and thus a_{41} has to be estimated.

Total of remaining results in Laboratory 4 = 36.354
Total of remaining results in Sample 1 = 19.845
Total of all the results except a_{41}= 348.358
Also S' = 8 and L = 9.

Hence, the estimate of a_{41} is given by

$$a_{ij} = \frac{1}{(9-1)\,(8-1)}\,[(9 \times 36.354) + (8 \times 19.845) - 348.358] \tag{11}$$

Therefore,

$$a_{ij} = \frac{137.588}{56} = 2.457 \tag{12}$$

7.6 *Rejection Test for Outlying Laboratories*:

7.6.1 At this stage, one further rejection test remains to be carried out. This determines whether it is necessary to reject the complete set of results from any particular laboratory. It could not be carried out at an earlier stage, except in the case where no individual results or pairs are missing or rejected. The procedure again consists of Hawkins' test (see 7.3.4), applied to the laboratory averages over all samples, with any estimated results included. If any laboratories are rejected on all samples, new estimates shall be calculated for any remaining missing values (see 7.5).

7.6.2 *Worked Example*:

7.6.2.1 The procedure on the laboratory averages shown in Table 8 follows exactly that specified in 7.3.4. The deviations of laboratory averages from the overall mean are given in Table 9 in units of the third decimal place, together with the sum of squares. Hawkins' test ratio is therefore:

$$B^* = 0.026/\sqrt{0.00222} = 0.5518 \tag{13}$$

Comparison with the value tabulated in Table A1.5, for n = 9 and v = 0, shows that this ratio is not significant and therefore no complete laboratory rejections are necessary.

7.7 *Confirmation of Selected Transformation* :

7.7.1 At this stage it is necessary to check that the rejections carried out have not invalidated the transformation used. If necessary, the procedure from 7.2 shall be repeated with the outliers replaced, and if a new transformation is selected, outlier tests shall be reapplied with the replacement values reestimated, based on the new transformation.

7.7.2 *Worked Example*:

7.7.2.1 It was not considered necessary in this case to repeat the calculations from 7.2 with the outlying pair deleted.

8. Analysis of Variance and Calculation of Precision Estimates

8.1 After the data have been inspected for uniformity, a transformation has been performed, if necessary, and any outliers have been rejected (see Section 7), an analysis of variance shall be carried out. First an analysis of variance table shall be constructed, and finally the precision estimates derived.

8.2 *Analysis of Variance*:

8.2.1 *Forming the Sums of Squares for the Laboratories × Samples Interaction Sum of Squares*— The estimated values, if any, shall be put in the array and an approximate analysis of variance performed.

$$\text{M} = \text{mean correction} = T^2/2L'S' \tag{14}$$

where:

L' = L – number of laboratories rejected in 7.6 – number of laboratories with no remaining results after rejections in 7.3.4,
S' = total of remaining pairs in the jth sample, and
T = the total of all duplicate test results.

$$\text{Samples sum of squares} = [\sum_{j=1}^{S'} (g_j^2/2L')\,] - M \tag{15}$$

where g_j is the sum of sample j test results.

$$\text{Laboratories sum of squares} = [\sum_{i=1}^{L'} (h_i^2/2S')] - M \tag{16}$$

TABLE 8 Averages of All Transformed Results from Each Laboratory

Laboratory	A	B	C	D	E	F	G	H	J	Grand Average
Average	2.437	2.439	2.424	2.426[A]	2.444	2.458	2.410	2.428	2.462	2.436

[A] Including estimated value.

TABLE 9 Absolute Deviations of Laboratory Averages from Grand Average × 1000

Laboratory	A	B	C	D	E	F	G	H	J	Sum of Squares
Deviation	1	3	12	10	8	22	26	8	26	2.22

where h_i is the sum of laboratory i test results.

$$\text{Pairs sum of squares} = (1/2)\left[\sum_{i=1}^{L'}\sum_{j=1}^{S'} a_{ij}^2\right] - M \tag{17}$$

I = Laboratories × samples interaction sum of squares
= (pairs sum of squares) – (laboratories sum of squares) – (sample sum of squares)

Ignoring any pairs in which there are estimated values, repeats sum of squares,

$$E = (1/2)\sum_{i=1}^{L'}\sum_{j=1}^{S'} e_{ij}^2 \tag{18}$$

The purpose of performing this approximate analysis of variance is to obtain the minimized laboratories × samples interaction sum of squares, I. This is then used as indicated in 8.2.2, to obtain the laboratories sum of squares. If there were no estimated values, the above analysis of variance is exact and paragraph 8.2.2 shall be disregarded.

8.2.1.1 *Worked Example*:

$$\text{Mean correction} = \frac{350.815^2}{144} = 854.6605 \tag{19}$$

where 854.6605 is the result obtained by electronic calculation without rounding the factors in the expression.

$$\text{Samples sum of squares} = \frac{22.302^2 + 72.512^2 + \ldots + 19.192^2}{18} - 854.6605 = 293.5409 \tag{20}$$

$$\text{Laboratories sum of squares} = \frac{38.992^2 + 39.020^2 + \ldots + 39.387^2}{16} - 854.6605 = 0.0356 \tag{21}$$

$$\text{Pairs sum of squares} = (1/2)\,(2.520^2 + 8.041^2 + \ldots + 2.238^2) - 854.6605 = 293.6908 \tag{22}$$

$$\text{Repeats sum of squares} = (1/2)\,(0.042^2 + 0.021^2 + \ldots + 0^2) = 0.0219 \tag{23}$$

Table 10 can then be derived.

8.2.2 *Forming the Sum of Squares for the Exact Analysis of Variance*:

8.2.2.1 In this subsection, all the estimated pairs are disregarded and new values of g_j are calculated. The following sums of squares for the exact analysis of variance **(6)** are formed.

$$\text{Uncorrected sample sum of squares} = \sum_{j=1}^{S'} \frac{g_j^2}{S_j} \tag{24}$$

TABLE 10 Sums of Squares: Bromine Example

Sources of Variation	Sum of Squares
Samples	293.5409
Laboratories	0.0356
Laboratories × samples interaction	0.1143
Pairs	293.6908
Repeats	0.0219