INTERNATIONAL

Designation: D3244 – 21a

An American National Standard

Standard Practice for Utilization of Test Data to Determine Conformance with Specifications[1]

This standard is issued under the fixed designation D3244; 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

The properties of commercial petroleum products are measured by standardized laboratory test methods to assess their conformance to specifications. Two or more measurement results obtained by performing the same test method for the same property of a specific sample usually will not be numerically identical. Therefore, the test methods generally include a paragraph on the precision of results. This precision (or, a more appropriate term is imprecision) is an expression of the degree of agreement that can be expected between the aforementioned measurements.

Many difficulties that arise in assessing conformance to specifications are due to test imprecision. Because of this, a true value of a property can never be determined exactly; and it is necessary to infer from measured values the range within which the "true value" is likely to lie. The main purpose of this practice is to indicate how test imprecision should be interpreted relative to specification limit values.

1. Scope*

1.1 This practice covers guidelines and statistical methodologies with which two parties (see Note 1) can compare and combine independently obtained test results to obtain an Assigned Test Value (*ATV*) for the purpose of resolving a dispute over product property conformance with specification.

NOTE 1—Application of this practice is usually, but not limited to, between supplier and receiver of a product.

1.2 This practice defines a technique for establishing an Acceptance Limit (*AL*) and Assigned Test Value (*ATV*) to resolve the dispute over a property conformance with specification by comparing the *ATV* to the *AL*.

1.3 This practice applies only to those test methods which specifically state that the repeatability and reproducibility values conform to the definitions herein.

1.4 The statistical principles and methodology outlined in this practice can also be used to obtain an *ATV* for specification conformance decision when multiple results are obtained for the same batch of product within a single laboratory. For this application, site precision (R') as defined in Practice D6299 shall be used in lieu of test method published reproducibility (R).

1.5 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

D1319 Test Method for Hydrocarbon Types in Liquid Petroleum Products by Fluorescent Indicator Adsorption
D4057 Practice for Manual Sampling of Petroleum and Petroleum Products
D4177 Practice for Automatic Sampling of Petroleum and Petroleum Products
D6299 Practice for Applying Statistical Quality Assurance and Control Charting Techniques to Evaluate Analytical Measurement System Performance
D6300 Practice for Determination of Precision and Bias Data for Use in Test Methods for Petroleum Products, Liquid Fuels, and Lubricants
D6792 Practice for Quality Management Systems in Petroleum Products, Liquid Fuels, and Lubricants Testing Laboratories

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

Current edition approved Oct. 1, 2021. Published October 2021. Originally published as an appendix to the *1968 Annual Book of ASTM Standards*, Part 18. Originally approved as a standard in 1974. Last previous edition approved in 2021 as D3244 – 21. DOI: 10.1520/D3244-21A.

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

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

D7372 Guide for Analysis and Interpretation of Proficiency Test Program Results

E29 Practice for Using Significant Digits in Test Data to Determine Conformance with Specifications

2.2 *ISO Standard:*[3]

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

3. Terminology

3.1 *Definitions:*

3.1.1 *acceptance limit (AL), n*—a numerical value that defines the point between making the property conformance or non-conformance to a specification decision.

3.1.1.1 *Discussion*—The *AL* is not necessarily the specification limit. It is a value that takes into account the specification limit, the test method precision, and the desired probability of making the conformance to specification decision if the true value (see 3.1.17) of the property is at the specification limit.

3.1.2 *assigned test value (ATV), n*—the average of all results obtained in the several laboratories which are considered acceptable based on the reproducibility of the test method.

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

3.1.4 *dispute, n*—when there is a question as to product property conformance to specification because a test value obtained falls outside the specification limit(s).

3.1.5 *dispute adjudication sample, n*—a mutually agreed sample between the parties in dispute to be used for the purpose of arriving at the ATV for the property that is in dispute with regards to its specification conformance status.

3.1.6 *operator, n*—a person who normally and regularly carries out a particular test.

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

3.1.8 *receiver, n*—any individual or organization who receives or accepts the product delivered by the supplier.

3.1.9 *receiver's risk, n*—the probability of accepting a product that fails to meet the specification.

3.1.10 *repeatability (a.k.a. Repeatability Limit) (r), n*—the quantitative expression for the random error associated with the difference between two independent results obtained under repeatability conditions that would be exceeded with an approximate probability of 5 % (one case in 20 in the long run) in the normal and correct operation of the test method. **D6300**

3.1.11 *repeatability conditions, n*—conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time. **D6300**

3.1.12 *reproducibility (a.k.a. Reproducibility Limit) (R), n*—a quantitative expression for the random error associated with the difference between two independent results obtained under reproducibility conditions that would be exceeded with an approximate probability of 5 % (one case in 20 in the long run) in the normal and correct operation of the test method. **D6300**

3.1.13 *reproducibility conditions, n*—conditions where independent test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment. **D6300**

3.1.14 *result, n*—the value obtained by following the complete set of instructions of a test method. It may be obtained from a single determination or several determinations, depending on the instruction of the test method.

3.1.15 *supplier, n*—any individual or organization responsible for the quality of a product just before it is taken over by the receiver.

3.1.16 *supplier's risk, n*—the probability of rejecting a product that meets the specification.

3.1.17 *true value (μ), n*—for practical purposes, the value towards which the average of single results obtained by *N* laboratories using the same standard test method tends, when *N* becomes very large. Consequently, this definition of true value is associated with the particular test method employed.

4. Significance and Use

4.1 This practice provides a means whereby the parties can resolve disputes over specification conformance for those product properties which can be tested and expressed numerically.

4.1.1 This practice can be used to ensure that such properties are correctly stated on labels or in other descriptions of the product.

4.1.2 This practice can be implemented in those cases where a supplier uses an in-house or a commercial testing laboratory to sample and test a product prior to releasing the product to a shipper (intermediate receiver) and the ultimate receiver also uses an in-house or commercial testing laboratory to sample and test the product upon arrival at the destination. The *ATV* would still be determined according to 8.3.

4.2 This practice can be applied in the determination of tolerances from specification limits based on a mutually agreed probability between parties for making the conformance to specification decision if the true value of a property is sufficiently close to the specification limit. Such tolerances are bounded by an *acceptance limit (AL)*. If the *ATV* value determined by applying this practice falls on the *AL* or on the acceptable side of the *AL*, the product property can be considered to have met the specification; otherwise it shall be considered to have failed to meet the specification.

4.3 Application of this practice requires the *AL* be determined prior to actual commencement of testing. Therefore, the degree of criticality of the specification, as determined by the Probability of Acceptance (P value) that is required to calculate

[3] Available from American National Standards Institute (ANSI), 25 W. 43rd St., 4th Floor, New York, NY 10036, http://www.ansi.org.

the *AL*, shall have been mutually agreed upon between both parties prior to execution of actual product testing.

4.3.1 This agreement should include a decision as to whether the *ATV* is to be determined by the absolute or rounding-off method of Practice E29, as therein defined.

4.3.1.1 If the rounding-off method is to be used, the number of significant digits to be retained must also be agreed upon.

4.3.1.2 These decisions must also be made in the case where only one party is involved, as in the case of a label.

4.3.1.3 In the absence of such an agreement, this practice recommends the *ATV* be rounded in accordance with the rounding-off method in Practice E29 to the number of significant digits that are specified in the governing specification.

4.4 This practice is designed to be suitable for reference in contracts governing the transfer of petroleum products and lubricants from a supplier to a receiver.

4.5 As a prerequisite for acceptance for lab test results to be used in this practice, the following conditions shall be satisfied:

4.5.1 Site precision (R') as defined in Practice D6299 for the appropriate test method(s) from each lab, as substantiated by control charts meeting the requirement of D6299 from in-house quality control programs, for property typical of the product in dispute, should have a TPI > 1.2 for methods with Precision Ratio <4 and TPI > 2.4 for methods with Precision Ratio ≥4 (see Practice D6792 for TPI explanation).

4.5.2 Each lab shall be able to demonstrate, by way of results from interlaboratory exchange programs, a lack of a systemic bias relative to exchange averages for the appropriate test method(s) as per methodology outlined in Guide D7372.

4.5.3 In the event that the site precision of laboratories from two parties are statistically different as confirmed by the F-test (see Annex A4), then, for the purpose of establishing the *ATV*, each laboratory's test result shall be inversely weighted in accordance with laboratory's demonstrated variance.

4.6 It is recommended that this practice be conducted under the guidance of a qualified statistician.

5. Sampling

5.1 The disputing parties shall agree on the sampling procedure to obtain the dispute adjudication sample. Obtain enough sample to allow for all required determinations to be made by at least two, and a possible third party.

6. Applying Test Method Precision Data to Accept or Reject Test Results

6.1 This section describes procedures in which the precision limits of test methods can be used as a decision criterion to accept or reject test results.

6.2 *Significance of Repeatability (r):*

6.2.1 *Acceptance of Results*—When only two results are obtained under conditions of repeatability and the difference is equal to or less than the repeatability of the method, the operator may report the average of the two results as being applicable to the sample tested.

6.2.2 *Rejection of Results*—When two results are obtained that differ by more than the repeatability of the method, both should be rejected. Obtain two additional results immediately under conditions of repeatability. If the difference between these two results is equal to or less than the repeatability of the method, the operator should report the average of the two as being applicable to the sample tested. If, however, the difference so obtained again exceeds the repeatability, reject the results and investigate the application of the method.

6.3 *Significance of Reproducibility (R):*

6.3.1 *Acceptance of Results*—When two results are obtained and comprise one result from each laboratory (Note 2), if the difference is equal to or less than the reproducibility of the method, then both results should be considered acceptable.

NOTE 2—When a comparison for reproducibility is made between results from two laboratories, it is a common practice that single results from each will be compared. If each of the laboratories has produced more than a single result, see 6.4.

6.3.2 *Rejection of Results*—When the results from two laboratories differ by more than the reproducibility of the method, reject both results and each laboratory should repeat the test on the retained sample. If the difference is now equal to or less than the reproducibility, both results should be considered acceptable. If, however, the difference between these results is still greater than the reproducibility, reject the results and investigate the application of the method at each laboratory, sampling, sample preparation and storage and all other factors which can contribute to the variance.

6.4 *Significance of Reduced Reproducibility (R_reduced) from Multiple Testing*—If the number of results obtained in either one or both laboratories is more than one, then the allowable difference between the averages from the two laboratories is given as follows:

$$\text{Difference, } R\_reduced = \sqrt{R^2 - r^2\left(1 - \frac{1}{2n_1} - \frac{1}{2n_2}\right)} \qquad (1)$$

where:

R = reproducibility of the method,
r = repeatability of the method,
n_1 = number of results of the first laboratory, and
n_2 = number of results of the second laboratory.

6.5 *Referee Laboratory*—In the event a third or referee laboratory is invited to perform the test using a portion of one of the samples described in 6.3.2, multiply the reproducibility, *R*, by 1.2 (to convert a range for two to a range for three) and compare this value with the difference between the two extreme results for acceptance. If acceptance is indicated, the *ATV* for the sample should be the average of the three results.

7. Determination of Acceptance Limits by Applying Test Method Precision Data and Specification Criticality Considerations to Specification Limits

7.1 *Specifications*—A specification fixes a limit to the *true value* of a given property. In practice, however, this *true value* can never be established exactly. The property is measured in the laboratory by applying a standard test method, the results of which may show some random scattering within tolerances as defined by the test method repeatability and reproducibility limits. Therefore, there is always some uncertainty as to the *true value* of the tested property.

7.2 Although the *true value* is never known exactly, the probability of obtaining any specific test result, relative to a hypothesized true value, can be calculated if the probability distribution function for the test method is known (for example, the Normal or Gaussian distribution).

7.2.1 Some specifications, because of the product characteristic or the end use of the product, or both, require that the receiver have a high degree of assurance that the true value of the product property actually meets or exceeds the quality level indicated by the specification limit value. For the purpose of this practice, such specifications are called *critical* specifications.

7.2.2 Specifications that require assurance only that the product property is not substantially poorer than is indicated by the specification limit are called *noncritical* specifications for the purposes of this practice.

7.3 *Specification Conformance Decision Guidelines:*

7.3.1 Whenever a product is tested for conformity to a specification for a specific property, a decision must ultimately be made as to whether the property conforms to specification.

7.3.2 The numerical value that delineates the regions of property conformance or nonconformance to specification is the *AL*. The *AL* may or may not coincide with the specification limit value (*S*) used to define the requirements for the product.

7.3.3 The *AL* value, calculated as described in this practice, shall be agreed upon between the disputing parties prior to commencement of testing.

7.3.4 The probability (*P*) of making the decision that the property conforms to specification when the true value of the property exactly equals the specification limit value is shown in Fig. 1 and Fig. 2 as a function of $D = (AL - S)/(0.255R)$, where *D* is a direct measure of the difference between *AL* and *S*. This relationship is based (*1*) on the assumption of normally (Gaussian) distributed testing errors, which is adequate for most test procedures, and (*2*) on using an *ATV* for making the specification conformance decision that is the average of precision-acceptable results from two laboratories.

7.3.4.1 For values of *P* greater than 0.5 (Noncritical Spec Region in Fig. 1), the *AL* decision is primarily driven by supplier's risk considerations (that is, probability of rejecting a product which actually meets the specification).

7.3.4.2 For values of *P* less than 0.5 (Critical Spec Region in Fig. 1), the *AL* decision is primarily driven by receiver's risk considerations (that is, probability of accepting a product which fails to meet the specification).

7.3.4.3 When *P* = 0.5, the *AL* coincides with the specification limit; the conformance to specification decision is based on equal sharing of test method imprecision related risks between the disputing parties for making the incorrect decision.

7.3.5 The *AL* associated with probability *P* of accepting the product when the true value exactly equals the specification limit value *S* is then given by:

$$AL = S + 0.255 \cdot R \cdot D \quad (2)$$

7.3.5.1 The factor 0.255 in Eq 2 is for *N* (no. of labs) = 2. For *N* greater than 2, the 0.255 factor should be multiplied by $\sqrt{2/N}$.

		$D = (AL - S)/0.255\,R$	
	Probability (*P*) of Acceptance	Maximum Specification Limit	Minimum Specification Limit
Critical Spec Region	0.001	−3.090	3.090
	0.005	−2.576	2.576
	0.010	−2.326	2.326
	0.025	−1.960	1.960
	0.050	−1.645	1.645
	0.100	−1.282	1.282
	0.150	−1.036	1.036
	0.200	−0.842	0.842
	0.300	−0.524	0.524
Noncritical Spec Region	0.500	0.000	0.000
	0.700	0.524	−0.524
	0.800	0.842	−0.842
	0.850	1.036	−1.036
	0.900	1.282	−1.282
	0.950	1.645	−1.645
	0.975	1.960	−1.960
	0.990	2.326	−2.326
	0.995	2.576	−2.576
	0.999	3.090	−3.090

NOTE 1—Based on *N* = 2 = number of different laboratories' results used to obtain *ATV*. See text for use of this table.

FIG. 1 Deviation of *AL* from Specification for Product Acceptance at a Given Probability

7.3.6 For specifications having both minimum and maximum limits, the procedure in 7.3.5 must be applied twice to give both upper and lower *AL*s. There must be some allowable region remaining between the lower and upper *AL*s.

7.3.7 When only a single test result is or will be available, the relationships given should be used with *N* = 1 (7.3.5.1). Obviously, no check on reproducibility precision can be made with a single test result, and the single value becomes the *ATV* for the sample.

7.3.8 The relationships between the *AL*s for critical and noncritical specifications are shown in Fig. 3 for a minimum specification.

7.3.9 For the risk of making the decision that the property conforms to specification when the true value fails to meet specification by various off-spec amounts expressed in units of R, see Table 1 and example in Annex A5.

8. Obtaining the *ATV*

8.1 The following procedure will produce an *ATV* with precision control based on the reproducibility of the test method.

8.2 Obtain two independent test results, one from each party, herein labelled as X_R and X_S, respectively, for the dispute adjudication sample using a mutually agreed test method.

8.3 *ATV Procedure:*

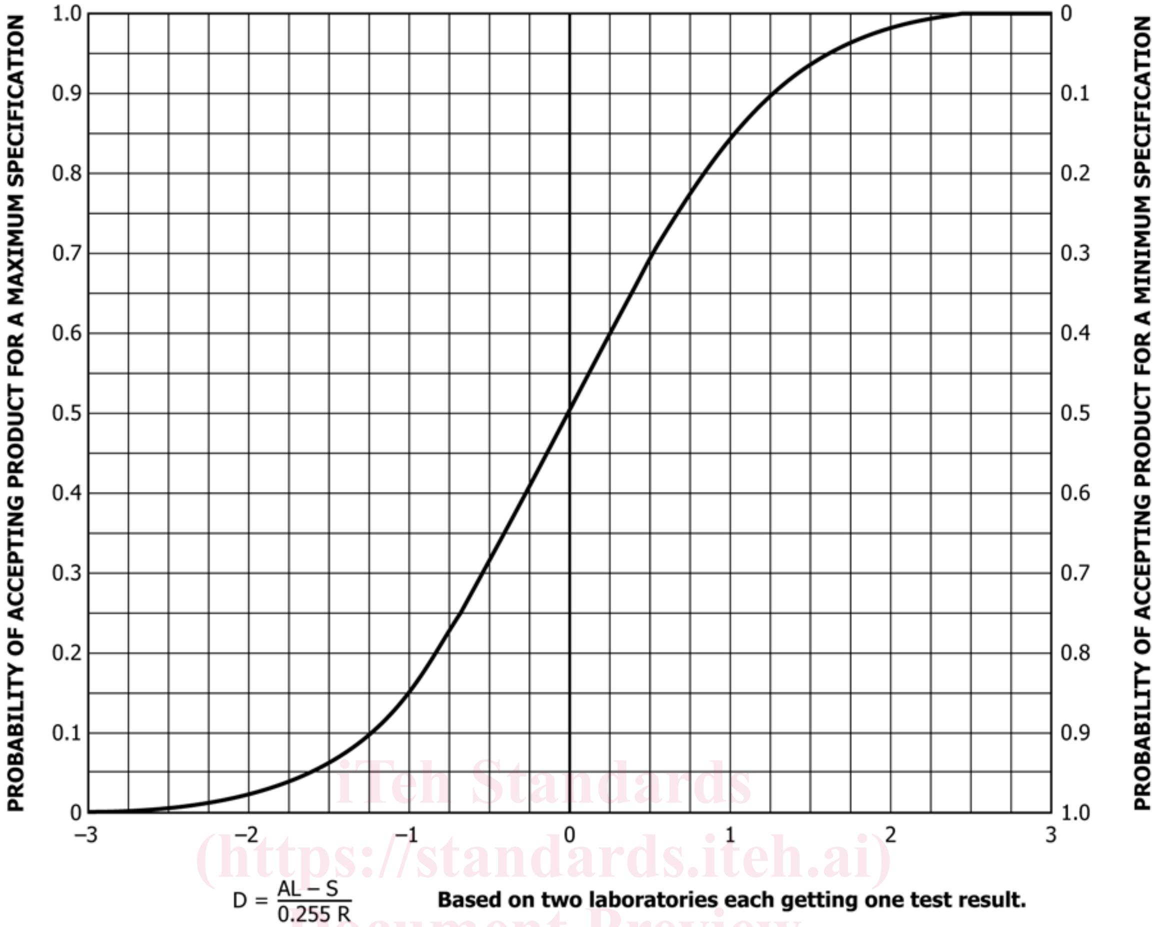
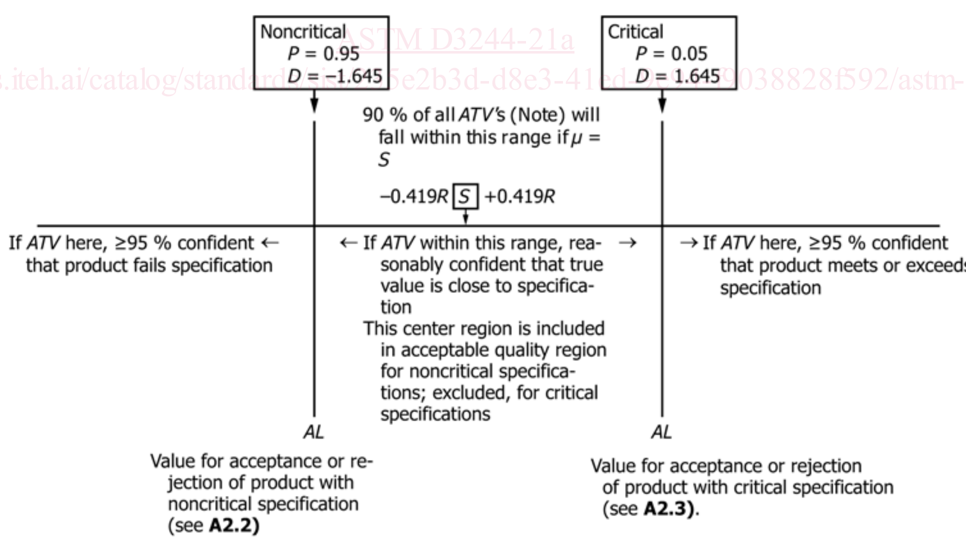


FIG. 2 Probability of Acceptance vs Deviation of *AL* from True Value = S



NOTE 1—This applies when *ATV* is established by the average of two results, one each from two different laboratories.

FIG. 3 Relationships Between *AL*s for Critical and Noncritical Specifications

8.3.1 If the absolute value of $\Delta = X_R - X_S$ is less than or equal to *R*, the reproducibility of the test method, average the two results to obtain the following in accordance with 6.3.1:

$$ATV = \left(X_R + X_S\right)/2 \quad (3)$$

8.3.2 If the absolute value of Δ exceeds R, reject both results and retest on portions of the retain sample to obtain X_R', X_S'.

8.3.3 If the absolute value of $\Delta' = X_R' - X_S'$ is less than or equal to R, average the two results to obtain the following in accordance with 6.3.2:

$$ATV = (X_R' + X_S')/2 \quad (4)$$

8.3.4 If the absolute value of Δ' exceeds R, obtain a new test value X_{RL} from a referee laboratory (6.5).

8.3.5 If $\Delta_3 = X_{max} - X_{min}$ is less than or equal to 1.2 R, obtain the following:

$$ATV = (X_R' + X_S' + X_{RL})/3 \quad (5)$$

8.3.6 If Δ_3 exceeds 1.2 R, obtain ATV as the average of the closer pair.

NOTE 3—This last step for obtaining an *ATV* does not comply rigidly to statistical concepts. It is done in this manner because in most cases the test sample (see Section 5) is depleted.

8.4 The above procedure will always yield an *ATV*. If the supplier's and receiver's laboratories have little or no bias relative to each other, then the procedure will end at 8.3.1 about 95 % of the time, and some 95 % of the remaining 5 %, at 8.3.3.

8.5 If any particular supplier and receiver pair find they frequently must go as far as calling for a reference laboratory test, they should carefully check their running of the test, as well as examine their calibration practice versus other laboratories that have demonstrated proficieny in the conduct of the particular test method.

8.6 This procedure for obtaining an *ATV* is designed for the test of samples obtained according to Section 5.

8.6.1 If more extensive testing is needed for special situations, comparable procedures can be developed. A statistician or quality control expert should be consulted to do this.

9. Product Property Conformance to Specification Consideration

9.1 If the *ATV* is equal to or better than the *AL*, the product property is to be considered as having met specification.

9.2 If the *ATV* fails the *AL* value, the product property is to be considered as failing to meet specification.

9.3 These concepts are presented graphically in Fig. 4.

9.3.1 The plotted lines are boundary conditions separating acceptable results from those indicating other alternative actions.

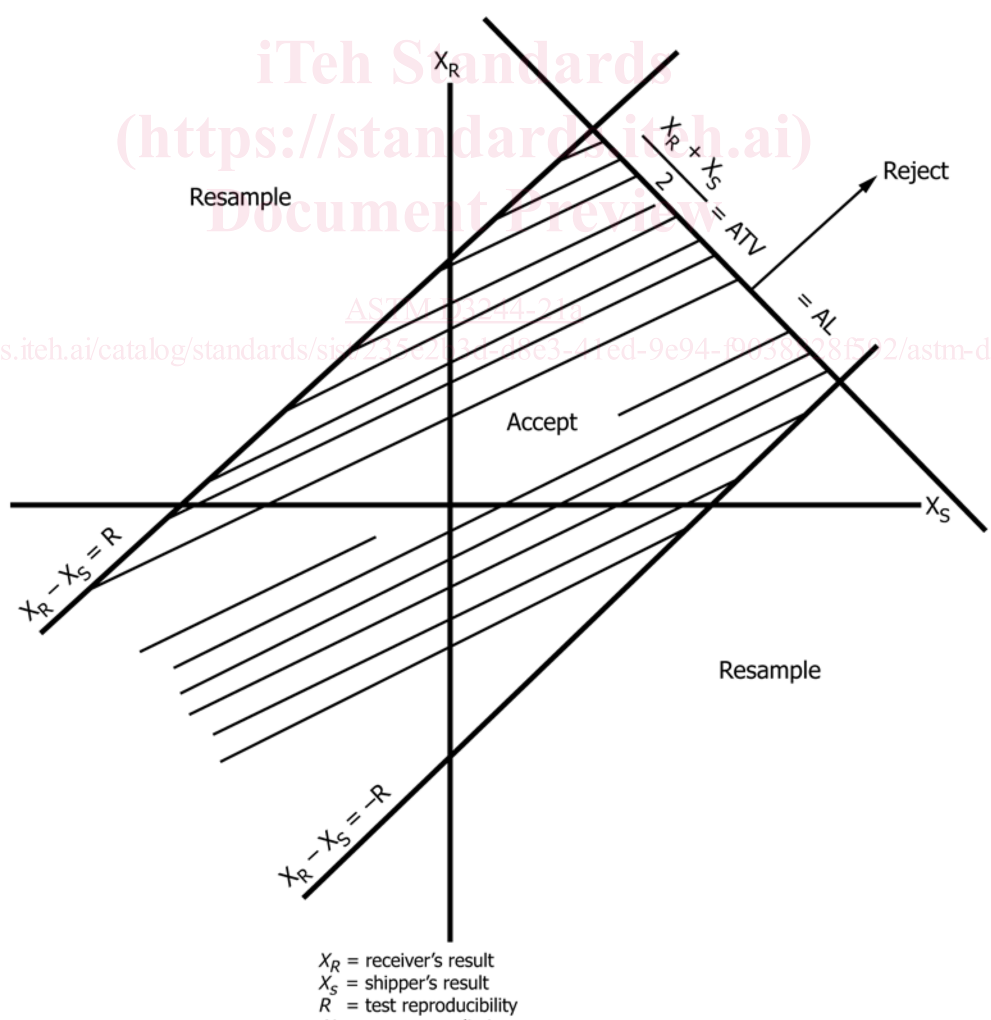


FIG. 4 Diagram Showing Regions of Acceptance, Rejection, and Resampling

9.3.1.1 The property is considered as having met specification if the two results fall to the left of the line, $(X_R + X_S)/2 = ATV = AL$ and if they are also within the lines $X_R - X_S = \pm R$.

9.3.2 The property is considered as failing to meet specification if the results lie to the right of the line $(X_R + X_S)/2 = ATV = AL$.

9.3.3 Initial results falling in the region labeled *resample* call for a retest.

9.3.3.1 If results for a second sample also fall in the resample region, a referee laboratory should be included in the new testing program.

9.4 The actual consequences of rejecting a product for failure to meet specification are subject to prior agreement or negotiation between the parties concerned.

10. Keywords

10.1 acceptance; acceptance limits; agreement; conformance; dispute; precision; rejection; specifications

TABLE 1 Receiver's Risk of Acceptance of Product that Fails to Meet Specification

		$D = (AL - S)/0.255\ R$		Supplier's Risk	Receiver's Risk			
	Probability (*P*) of Acceptance if true value (μ) is exactly at *S*	Maximum Specification Limit = *S*	Minimum Specification Limit = *S*	Probability (*P*) of Rejection if true value (μ) is exactly at *S*	Probability (*P*) of Acceptance if true value (μ) is not exactly at *S*, but is off spec by 0.25*R*	Probability (*P*) of Acceptance if true value (μ) is not exactly at *S*, but is off spec by 0.5*R*	Probability (*P*) of Acceptance if true value (μ) is not exactly at *S*, but is off spec by 0.75*R*	Probability (*P*) of Acceptance if true value (μ) is not exactly at *S*, but is off spec by 1*R*
Critical Spec Region	0.001	–3.090	3.090	0.999	<0.001	<0.001	<0.001	<0.001
	0.005	–2.576	2.576	0.995	<0.001	<0.001	<0.001	<0.001
	0.010	–2.326	2.326	0.990	<0.001	<0.001	<0.001	<0.001
	0.025	–1.960	1.960	0.975	0.002	<0.001	<0.001	<0.001
	0.050	–1.645	1.645	0.950	0.004	<0.001	<0.001	<0.001
	0.100	–1.282	1.282	0.900	0.012	<0.001	<0.001	<0.001
	0.150	–1.036	1.036	0.850	0.022	0.001	<0.001	<0.001
	0.200	–0.842	0.842	0.800	0.034	0.003	<0.001	<0.001
	0.300	–0.524	0.524	0.700	0.066	0.006	<0.001	<0.001
Noncritical Spec Region	0.500	0.000	0.000	0.500	0.163	0.025	0.002	<0.001
	0.700	0.524	–0.524	0.300	0.324	0.075	0.008	<0.001
	0.800	0.842	–0.842	0.200	0.445	0.132	0.018	0.001
	0.850	1.036	–1.036	0.150	0.522	0.178	0.028	0.002
	0.900	1.282	–1.282	0.100	0.619	0.249	0.049	0.004
	0.950	1.645	–1.645	0.050	0.747	0.376	0.097	0.011
	0.975	1.960	–1.960	0.025	0.836	0.500	0.163	0.025
	0.990	2.326	–2.326	0.010	0.911	0.643	0.269	0.055
	0.995	2.576	–2.576	0.005	0.945	0.731	0.357	0.089
	0.999	3.090	–3.090	0.001	0.983	0.871	0.559	0.203