What Are Repeatability and Reproducibility?

Part 1: A D02 Viewpoint for Laboratories[1]

BY ALEX LAU

Q: Many ASTM standard test methods contain repeatability and reproducibility statements and values. What is the conceptual difference between repeatability and reproducibility, and how can this information be applied?

A. A standard test method prescribes a fixed set of equipment requirements, measurement methodologies and procedural protocols on how to produce a measurement for a specific property of interest. In a perfect world, if the same material is repeatedly tested using the same test method, whether by the same operator in the same laboratory using the same apparatus or by different laboratories and operators using different apparatus meeting the specified requirements, all measurements generated will be numerically identical.

However, we do not live in a perfect world. Repeated execution of the same test method on the same material, whether by the same operator in the same laboratory using the same apparatus or by different operators in different laboratories using apparatus of similar design, will not always yield numerically identical results. Minor differences will exist between test results due to variations in factors that are inherent in the test method.

For instance, different operators may prepare the test material in slightly different manners. Different apparatus, while designed to the same specification, may differ slightly in performance. The quality of reagents may differ slightly between laboratories. The variation in environmental conditions, for example, humidity, temperature, pressure, within or between laboratories, may affect instrumentation responses. Variation attributable to factors considered inherent in the test method is generally referred as common-cause variation and can be thought of conceptually as the inherent "imperfections" associated with the test method. In statistical nomenclature, this inherent "imperfection" is known as *precision* although some might prefer to think of it as *im*precision.

Intuitively, the precision associated with multiple test results obtained by the same operator on the same test material using the same apparatus and test method within a very short time should be less than the variation associated with an equivalent number of test results, where each test result is produced by a different laboratory on the same test material using the same test method. This intuitive expectation comes from recognizing that, in the former scenario, the number of factors that can affect the precision of the test results is less than the latter. Therefore, precision must be associated with specific conditions.

Repeatability and reproducibility are standardized terms[2] adopted by ASTM and other standardization organizations, and they are associated with the precision of measurements generated on the same material using the same test method under specific conditions. Table 1, Conditions for Precision, outlines three different sets of specific conditions associated with precision as defined in ASTM E177, Practice for Use of the Terms Precision and Bias in ASTM Test Methods, and E456, Terminology Relating to Quality and Statistics.

Precision associated with repeatability and reproducibility conditions is quantified by the repeatability and reproducibility limits published in the precision and bias section of an ASTM test method. A nonstatistical interpretation of these values is that these are the maximum difference between two results obtained under specified conditions that can be attributed to the test method precision. Hence, the published repeatability and reproducibility limits (r, R) can be used as decision limits to support or challenge the validity of the assumption that both test results have been produced on the same material in a correct manner under the associated specific conditions. It should be noted that *r* and *R* may not necessarily remain constant throughout the measurement range or material types included in the scope of the test method. Repeatability and reproducibility limits are derived statistically from specially designed studies. Various standard practices prescribing specific requirements associated with these studies have been developed and adopted by various industries to address their specific needs. ASTM standards C802, D2777, E180, E691, D6300, and ISO 4259 and ISO 5725[3] are examples of these standard practices. While the statistical approaches and philosophies may appear different to the practitioner, these practices are all based on similar statistical principles, and the interpretation and application of the derived end results are essentially the same.

The application of *r* and *R* can best be illustrated with the following two examples.

EXAMPLE 1 — DIESEL FUEL AND D7039

Two test results for total sulfur obtained successively by the same operator and instrument on aliquots prepared from the same test sample of diesel fuel using ASTM D7039, Test Method for Sulfur in Gasoline and Diesel Fuel by Monochromatic Wavelength Dispersive X-ray Fluorescence Spectrometry, an X-ray-based test method where precision is known to vary with sulfur level, were returned at 17 and 14 weight parts per million. A decision needs to be made on whether the difference between the two results can be attributable solely to test method precision.

The repeatability and reproducibility limits as published in D7039 are:

$$\text{Repeatability } (r) = 0.23*(\bar{X})^{0.5} = 0.23*(15.5)^{0.5} = 0.91$$

$$\text{Reproducibility } (R) = 0.5*(\bar{X})^{0.5}$$

where $\bar{X}$ = average of two results (Note: Limits are dependent on the sulfur level).

Table 1 — Conditions for Precision

	Repeatability Condition	Intermediate Precision Condition	Reproducibility Condition
Laboratory	Same	Same	Different
Operator	Same	Different	Different
Apparatus	Same	Same[a]	Different
Time between Tests	Short[b]	Multiple Days	Not Specified

[a]This situation can be different instruments meeting the same design requirement.
[b]Standard test method dependent, typically does not exceed one day

Because both results are obtained under repeatability conditions, the published repeatability limit is evaluated as 0.91 wppm at the average value of 15.5 wppm. Because the absolute difference between the two results of 3 wppm exceeds 0.91, the difference cannot be attributed solely to test method precision.

EXAMPLE 2 — GASOLINE AND D2700

A supplier released a batch of gasoline with a single lab inspection of 82.2 motor octane number using ASTM D2700, Test Method for Motor Octane Number of Spark-Ignition Engine Fuel. The customer tested the gasoline at their receiving facilities using the same test method and obtained a single test result of 81.5 MON. A decision needs to be made using the two results as to whether there is evidence of octane degradation due to transit.

The repeatability and reproducibility values (between 80.0 and 90.0) as published in ASTM D2700 are:

Repeatability = 0.2 MON
Reproducibility = 0.9 MON

Because both results are obtained under reproducibility conditions, and the absolute difference of 0.7 is less than the published reproducibility of 0.9, the difference can be attributable solely to test method precision. Hence there is no compelling evidence to suggest that there is octane degradation in transit.

REFERENCES

1. *ASTM Committee D02 on Petroleum Products and Lubricants*
2. For exact ASTM definitions of precision, repeatability and reproducibility, readers are referred to ASTM E177, Practice for Use of the Terms Precision and Bias in ASTM Test Methods, and ASTM E456, Terminology Relating to Quality and Statistics.
3. The complete titles of the ASTM standards are:
C802, Practice for Conducting an Interlaboratory Test Program to Determine the Precision of Test Methods for Construction Materials;
D2777, Practice for Determination of Precision and Bias of Applicable Test Methods of Committee D19 on Water;
E180, Practice for Determining the Precision of ASTM Methods for Analysis and Testing of Industrial and Specialty Chemicals;
E691, Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method;
D6300, Practice for Determination of Precision and Bias Data for Use in Test Methods for Petroleum Products and Lubricants;
ISO 4259, Petroleum Products – Determination and Application of Precision Data in Relation to Methods of Test; and
ISO 5725, Rubber and Rubber Products – Determination of Precision for Test Method Standards.

Next issue: Part 2 — The practical application of statistics standards for repeatability and reproducibility in interlaboratory studies.

ALEX T. LAU, *Engineering Services Canada, is chair of Subcommittee D02.94 on Quality Assurance and Statistics, part of ASTM Committee D02 on Petroleum Products and Lubricants, and a contributing member of Committee E11 on Quality and Statistics.*
DEAN NEUBAUER *is the DataPoints column coordinator and E11.90.03 publications chair.*

Statistics play an important role in the ASTM International standards you write, such as the development of precision and bias statements for test methods, running interlaboratory studies, knowing how to round numbers properly and determining sample size. A panel of experts is ready to answer your questions about how to use statistical principles in ASTM standards. Please send your questions to SN *Editor in Chief Maryann Gorman at mgorman@astm.org or ASTM International, 100 Barr Harbor Drive, P.O. Box C700, W. Conshohocken, PA 19428-2959.*