1. **SCOPE**

1.1. **Description of Test**

An interlaboratory testing program is used to evaluate the consistency of test results obtained from a well-defined test procedure by different laboratories testing the same controlled material. An interlaboratory testing program removes the material and sampling variability and provides information about the equipment and test procedure variability that can be expected between laboratories or within a laboratory. The test results obtained from an interlaboratory testing program on a specific test procedure can be used to develop a precision statement for that test procedure.

This standard test procedure defines a process that can be used to design, coordinate and analyze an interlaboratory testing program on a specific test procedure.

This standard test procedure is used to determine precision statements for the testing procedures involved in the interlaboratory testing program.

1.2. **Application of Test**

This standard test procedure is used to design, coordinate and analyze an interlaboratory testing program. It can be used to determine precision statements for the test procedures used in the interlaboratory testing program when six or more laboratories are involved.

1.3. **Definitions**

1.3.1. *Repeatability and Reproducibility*

These terms deal with the variability of test results obtained under specified laboratory conditions.

*Repeatability* deals with the variability between independent test results obtained within a single laboratory in the shortest practical time period by a single operator with a specific set of testing equipment using random samples from a homogenous material prepared for the interlaboratory study. “Shortest practical time period” means that the test results, at least for one material, are obtained in a time not less than in normal testing and not so long as to permit significant changes in test material, equipment or environment.

*Reproducibility* deals with the variability between single test results obtained from different laboratories, each of which has applied the test procedures to random samples from a homogenous material prepared for the interlaboratory study.
1.3.2. **Repeatability Conditions**

Repeatability conditions are the within laboratory conditions specified for repeatability. The single operator, single set of test equipment means that for a particular step in the measurement process the same combination of operator and apparatus is used for every test result on every material. One operator may prepare the test specimens, a second measure the dimensions and a third measure the breaking force.

1.3.3. **Precision, Bias and Accuracy of a Test Method**

The *precision of a test procedure* is a generic concept related to the closeness of the agreement between test results obtained by the test procedure being evaluated under prescribed like conditions. The measurement process must be in a state of statistical control or the precision of the process has no meaning. The greater the scatter of the test results the poorer the precision. Precision is usually expressed as the standard deviation or some multiple of the standard deviation.

The *bias of a test procedure* is a generic concept related to a consistent or systematic difference between a set of test results obtained from the test procedure and an accepted value of the property being measured. The measuring process must be in a state of statistical control or the bias of the process has no meaning. In determining the bias, the effect of imprecision is averaged out by using the average of a very large set of test results. This average minus the accepted reference value is an estimate of the test procedure’s bias. If an accepted reference value is not available, bias can not be determined.

*Accuracy* is a generic concept of exactness related to the closeness of agreement between the average of one or more test results and an accepted reference value. The word “*accuracy*” alone refers to the accuracy of a test result. The term “*accuracy of a test result*” refers to the closeness of agreement between the test result and the accepted reference value. It depends on the test procedure’s imprecision and bias.

When a test procedure is applied to a large number of samples of a material, that are as nearly alike as possible, the test results will not all have the same value. A measure of the degree of agreement among these test results describes the precision of the test procedure for that material. Numerical measures of the variability between such test results provide inverse measures of the precision of the test method. Greater variability implies smaller precision (poorer) and larger imprecision.

This test procedure is designed to estimate the precision of a test method. When accepted reference values are available for the different property levels, the data
obtained according to this standard test procedure can be used to estimate the bias of a specific test procedure.

1.3.4. Test Units and Test Specimens

A test unit is the total quantity of material needed for obtaining a test result as specified by the specific standard test procedure. The portion of a test unit required to obtain a single test determination is called a test specimen. Usually a separate test specimen is required for each test determination.

1.3.5. Level of Significance

The probability of incorrectly deciding that two sets of data are different when in fact they are from the same population. The level of significance is typically selected as 0.05 or 0.01 which indicates that there is a 5% (5 times out of 100) or 1% (1 time out of 100) chance of deciding that two sets of data are from different populations when in fact they are from the same population.

1.4. Summary of Practice

This standard test procedure consists of four basic steps:

- planning the interlaboratory study,
- guiding the testing phase of the study,
- analyzing the test result data,
- developing precision and bias statements.

The analysis uses tabular, graphical and statistical diagnostic tools for evaluating the consistency of the data so that unusual values can be identified and investigated. It also includes calculations of numerical measures of precision for the standard test procedure’s within-laboratory repeatability and between-laboratory reproducibility.

2. TEST PROCEDURES, LABORATORIES AND MATERIALS

2.1. Test Procedures

All testing performed in the interlaboratory testing program will be done in accordance with Saskatchewan Highways and Transportation Standard Test Procedure(s) unless otherwise noted. The Interlaboratory Testing Program Coordinator will identify the appropriate Standard Test Procedure(s) that are to be used in the interlaboratory study.

2.2. Laboratories

Acceptable test results from a minimum of three laboratories are required to complete an interlaboratory testing program. Acceptable test results from at least six laboratories are required to develop a precision statement for a test procedure.
Any laboratory that is qualified to complete the test described in the specific Standard Test Procedure may participate in the interlaboratory testing program. “ Qualified” implies:

- proper laboratory facilities and testing equipment,
- competent operators,
- familiarity with the test procedure,
- a reputation for reliable testing work,
- sufficient time and interest to do a good job.

It is very important that the operators in the participating laboratories are familiar with the test procedure. If the operators are not familiar with the test procedure, they should be given the opportunity to familiarize themselves with the test procedure and practice its application before the interlaboratory testing program starts.

2.3. Materials

Material designates anything with a property that can be measured. Different materials having the same property may be expected to have different property levels, meaning higher or lower values of the same property. For example, an asphalt concrete mix is a material that has properties, such as asphalt content, that can be measured. Different asphalt concrete mixes can have the same property, asphalt content, at different levels, low asphalt content or high asphalt content.

An interlaboratory testing program should include at least three materials representing different levels of the property being measured by the standard test procedure. For the development of broadly applicable precision statements at least six materials should be included in the program.

The materials involved in the interlaboratory testing program should differ primarily only in the level of the property being measured by the test procedure. If it is known or suspected that different classes of materials will exhibit different levels of precision when tested by the test procedure, consideration should be given to conducting a separate interlaboratory testing program for each class of material.

Material homogeneity is required to remove any potential material variability so that each laboratory is essentially testing an identical material. Thus, each material used in an interlaboratory testing program should be produced to be or selected to be as homogenous as possible before it is subdivided into test units. If the randomization and distribution of individual test specimens (rather than test units) does not conflict with the procedure for preparing the sample for testing, as specified in the test method, greater homogeneity between test units can be achieved by randomizing test specimens. In this case, each test unit would be composed of the required number of randomized test specimens.
2.4. **Number of Test Results per Material**

A sufficient number of test results on each material must be specified in order to obtain a good estimate of the measure of repeatability, usually the repeatability standard deviation. In many cases, the repeatability standard deviation will be a function of the property level being measured. When this happens, the repeatability standard deviation should be determined separately for each property level.

At least three test results from each lab are required on each material. If there is little possibility of losing test units or of obtaining questionable results as few as two test results on each material is sufficient. The number of tests on each material should be increased when the test results are likely to vary considerably.

3. **PROCEDURE**

3.1. **Interlaboratory Testing Program General Information**

One person should be designated as the Interlaboratory Testing Program Coordinator. This person will be the contact for all of the laboratories involved in the interlaboratory testing program.

Each laboratory should be notified in writing to identify their interest in participating. This notification should include a schedule of significant dates for the interlaboratory testing program, as well as some indication of how and when the results will be distributed.

The Standard Test Procedure(s) being evaluated should be clearly identified. If a Standard Test Procedure allows several options in apparatus or procedure, the Coordinator should specify which option or options are to be used in the interlaboratory testing program.

If special calibration procedures are required before every test result, they should be identified in the Standard Test Procedure. If the Standard Test Procedure requires calibration only daily or less frequently, the Interlaboratory Testing Program Coordinator will decide if calibration is to be completed before obtaining each test result. Requiring calibration may eliminate calibration drift and help ensure relative independence of the test results, changes in calibration may increase the variability between test results.

Any special circumstances that must be addressed in order to implement repeatability conditions, such as the period of time between obtaining test results for the same material (not less than in normal testing and not so long as to likely permit significant changes in the test material, equipment or environment) must be specified by the Interlaboratory Testing Program Coordinator.
The Interlaboratory Testing Program Coordinator must specify the required care, handling and conditioning of the materials. The coding system used in identifying the materials and the distinction between test units and test specimens must be explained to the participating laboratories.

Data sheets to record raw observations for each material should be supplied. Instructions should be given on the number of significant digits that are to be recorded when obtaining the raw observations. Test result sheets on which test results can be calculated and reported should also be supplied. In most cases it may be possible to combine the result sheet with the raw data sheet. The number of significant digits to be used in reporting the test results should also be identified. The participating laboratories should be instructed to return the raw data and test result sheets as soon as testing is completed, and at least weekly if testing will continue over several weeks.

Each laboratory should be instructed to keep a record of any special events that arise during any part of the testing. This record should be returned to the Interlaboratory Testing Program Coordinator. It may be useful in dealing with unusual data and in revising the Standard Test Procedure being studied.

The laboratories should be instructed to notify the Interlaboratory Testing Program Coordinator whenever an error occurs while following (or in following) the test procedure. A decision must be made as to whether a new set of test units should be sent to the laboratory for a complete retest of the material.

A questionnaire should be included with the information given to the participating laboratories. It should be designed to obtain information on specific aspects of the apparatus, reagents, calibration, and procedure. It should capture any other information that might assist in dealing with data inconsistencies. The questionnaire can be used to ensure that the laboratory complied with the requirements of the test method.

3.2. **Trial Run**

It may be useful, especially with new test procedures, to conduct a trial run of the interlaboratory test program with only one or two materials to determine if the standard test procedure(s) is clear. This can also serve to familiarize laboratories that have little experience using the standard test procedure(s).

Even though it is just a test run, the entire process that is to be followed in the full scale test program should be carried out.

3.3. **Full Scale Run**

There should be enough material prepared to supply 50% more than is needed by the number of laboratories taking part in the interlaboratory testing program. Each test unit or test specimen should be labeled with a letter for the material and a sequential number.
If ten laboratories were to each complete 2 tests, the test units would be labeled from B1 to B30, 10 extra samples (50% more) are required for retesting, lost samples, etc.

The specified number of test units per material should be randomly distributed to all of the labs. A random number table or a software program using a random number generator should be used to ensure that the samples are distributed at random.

The test units must be packaged properly to arrive in the desired condition. When the material is sensitive to the conditions to which it is exposed (light, heat, humidity, etc.) place special directions for opening the package on the outside of the package. Clearly indicate the name of the person who has been designated as the contact person for that laboratory on the address of each package. Once the test units have been shipped, the Interlaboratory Testing Program Coordinator should call each laboratory to confirm that all of the test units have arrived safely.

As the interlaboratory testing program progresses, a laboratory may discover that the standard test procedure was not used properly on some test units. The affected laboratory must discuss this with the Interlaboratory Testing Program Coordinator who may use his discretion to decide whether to send a replacement set of test units or leave the results as they were obtained.

From time to time, at intervals appropriate to the magnitude of the interlaboratory testing program, the Coordinator should contact each laboratory to determine how the testing is progressing. By comparing the progress of all laboratories, the Coordinator can determine whether some laboratories are lagging considerably behind the others and advise them accordingly.

The completed data sheets should be examined by the Coordinator as soon as they are received in order to detect unusual values or other deficiencies that should be questioned. Replacement sets of test units may be issued when there is missing or obviously erroneous data. The Coordinator will decide whether or not the additional data should be used in the estimation of the Standard Test Procedure’s precision.

4. **CALCULATIONS**

4.1. **Statistical Calculations**

The analysis and treatment of the interlaboratory testing program data has three purposes:

- to determine whether the collected data are consistent enough to form the basis for a test procedure precision statement,
- to investigate and act on any data that is considered to be inconsistent,
- to obtain the precision statistics on which the precision statement can be based.
The statistical analysis of the data is a one way analysis of variance (within and between laboratories) carried out separately for each property level. This type of analysis can be severely affected by severe outliers. Therefore, it is necessary to examine the data consistency first. This will be explained through the use of an example.

4.2. Calculations

4.2.1. Sample Data

The calculations required will be illustrated through the use of an example. For the purpose of the calculations, assume that 50 Blow Hand Compacted Marshall Density tests are conducted on three different materials in six different laboratories. Each laboratory completed three tests on each material. The test results were recorded in Table 1:

<table>
<thead>
<tr>
<th>Material</th>
<th>Test Number</th>
<th>Laboratory 1</th>
<th>Laboratory 2</th>
<th>Laboratory 3</th>
<th>Laboratory 4</th>
<th>Laboratory 5</th>
<th>Laboratory 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.0</td>
<td>2,359.0</td>
<td>2,375.0</td>
<td>2,352.0</td>
<td>2,361.0</td>
<td>2,334.0</td>
<td>2,357.0</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>2,359.0</td>
<td>2,360.0</td>
<td>2,365.0</td>
<td>2,368.0</td>
<td>2,336.0</td>
<td>2,336.0</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>2,368.0</td>
<td>2,363.0</td>
<td>2,356.0</td>
<td>2,361.0</td>
<td>2,336.0</td>
<td>2,360.0</td>
</tr>
<tr>
<td>B</td>
<td>1.0</td>
<td>2,356.0</td>
<td>2,370.0</td>
<td>2,353.0</td>
<td>2,388.0</td>
<td>2,357.0</td>
<td>2,339.0</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>2,352.0</td>
<td>2,368.0</td>
<td>2,365.0</td>
<td>2,380.0</td>
<td>2,358.0</td>
<td>2,348.0</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>2,354.0</td>
<td>2,379.0</td>
<td>2,363.0</td>
<td>2,375.0</td>
<td>2,358.0</td>
<td>2,353.0</td>
</tr>
<tr>
<td>C</td>
<td>1.0</td>
<td>2,350.0</td>
<td>2,358.0</td>
<td>2,339.0</td>
<td>2,347.0</td>
<td>2,350.0</td>
<td>2,349.0</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>2,352.0</td>
<td>2,356.0</td>
<td>2,348.0</td>
<td>2,353.0</td>
<td>2,351.0</td>
<td>2,349.0</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>2,346.0</td>
<td>2,349.0</td>
<td>2,350.0</td>
<td>2,354.0</td>
<td>2,351.0</td>
<td>2,347.0</td>
</tr>
</tbody>
</table>

4.2.2. Initial Preparation of Test Data

The test data should be separated into individual tables for each material as shown in Table 2, Table 3, and Table 4.

<table>
<thead>
<tr>
<th>Laboratory Number</th>
<th>Test Results, x</th>
<th>$\bar{x}$</th>
<th>s</th>
<th>$s^2$</th>
<th>d</th>
<th>$d^2$</th>
<th>h</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,359.0, 2,359.0, 2,368.0</td>
<td>2,366.0000</td>
<td>5.1962</td>
<td>27.0005</td>
<td>4.7778</td>
<td>22.8274</td>
<td>0.43</td>
<td>1.02</td>
</tr>
<tr>
<td>2</td>
<td>2,375.0, 2,360.0, 2,363.0</td>
<td>2,366.0000</td>
<td>7.9373</td>
<td>63.0007</td>
<td>8.7778</td>
<td>77.0498</td>
<td>0.79</td>
<td>1.56</td>
</tr>
<tr>
<td>3</td>
<td>2,352.0, 2,365.0, 2,356.0</td>
<td>2,357.6667</td>
<td>6.6583</td>
<td>44.3330</td>
<td>0.4444</td>
<td>0.1975</td>
<td>0.04</td>
<td>1.31</td>
</tr>
<tr>
<td>4</td>
<td>2,361.0, 2,368.0, 2,361.0</td>
<td>2,363.3333</td>
<td>4.0415</td>
<td>16.3337</td>
<td>6.1111</td>
<td>37.3455</td>
<td>0.55</td>
<td>0.80</td>
</tr>
<tr>
<td>5</td>
<td>2,334.0, 2,336.0, 2,336.0</td>
<td>2,335.3333</td>
<td>1.1547</td>
<td>1.3333</td>
<td>-21.8889</td>
<td>479.1239</td>
<td>-1.97</td>
<td>0.23</td>
</tr>
<tr>
<td>6</td>
<td>2,357.0, 2,360.0, 2,360.0</td>
<td>2,359.0000</td>
<td>1.7321</td>
<td>3.0002</td>
<td>1.7778</td>
<td>3.1606</td>
<td>0.16</td>
<td>0.34</td>
</tr>
</tbody>
</table>

$\Sigma \bar{x} = 14,143.3333 \quad \Sigma s^2 = 155.0014 \quad \Sigma d^2 = 619.7047$
### Table 3: Initial Preparation of Test Result Data for 50 Blow Hand Compaction Marshall Density for Material B

<table>
<thead>
<tr>
<th>Laboratory Number</th>
<th>Test Results, x</th>
<th>Mean (X)</th>
<th>Standard Deviation (s)</th>
<th>Variance (s²)</th>
<th>Deviation (d)</th>
<th>Squared Deviation (d²)</th>
<th>Correction Factor (h)</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,356.0 2,352.0 2,354.0</td>
<td>2,354.0000</td>
<td>2.0000</td>
<td>4.0000</td>
<td>-8.0000</td>
<td>64.0000</td>
<td>-0.64</td>
<td>0.37</td>
</tr>
<tr>
<td>2</td>
<td>2,370.0 2,368.0 2,379.0</td>
<td>2,372.3333</td>
<td>5.8595</td>
<td>34.3337</td>
<td>10.3333</td>
<td>106.7771</td>
<td>0.82</td>
<td>1.09</td>
</tr>
<tr>
<td>3</td>
<td>2,353.0 2,365.0 2,363.0</td>
<td>2,360.3333</td>
<td>6.4291</td>
<td>41.3333</td>
<td>-1.6667</td>
<td>2.7779</td>
<td>-0.13</td>
<td>1.20</td>
</tr>
<tr>
<td>4</td>
<td>2,388.0 2,380.0 2,375.0</td>
<td>2,381.0000</td>
<td>6.5574</td>
<td>42.9995</td>
<td>19.0000</td>
<td>361.0000</td>
<td>1.51</td>
<td>1.22</td>
</tr>
<tr>
<td>5</td>
<td>2,357.0 2,358.0 2,358.0</td>
<td>2,357.6667</td>
<td>6.5574</td>
<td>42.9995</td>
<td>19.0000</td>
<td>361.0000</td>
<td>1.51</td>
<td>1.22</td>
</tr>
<tr>
<td>6</td>
<td>2,339.0 2,348.0 2,353.0</td>
<td>2,346.6667</td>
<td>7.0946</td>
<td>50.3333</td>
<td>-15.3333</td>
<td>235.1101</td>
<td>-1.22</td>
<td>1.32</td>
</tr>
</tbody>
</table>

\[ \sum X = 14,172.0000 \]
\[ \sum s^2 = 173.3333 \]
\[ \sum d^2 = 788.4426 \]

### Table 4: Initial Preparation of Test Result Data for 50 Blow Hand Compaction Marshall Density for Material C

<table>
<thead>
<tr>
<th>Laboratory Number</th>
<th>Test Results, x</th>
<th>Mean (X)</th>
<th>Standard Deviation (s)</th>
<th>Variance (s²)</th>
<th>Deviation (d)</th>
<th>Squared Deviation (d²)</th>
<th>Correction Factor (h)</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,350.0 2,352.0 2,346.0</td>
<td>2,349.3333</td>
<td>5.8595</td>
<td>34.3337</td>
<td>-0.3333</td>
<td>0.1111</td>
<td>-0.10</td>
<td>0.80</td>
</tr>
<tr>
<td>2</td>
<td>2,358.0 2,356.0 2,349.0</td>
<td>2,354.3333</td>
<td>4.7258</td>
<td>22.3332</td>
<td>-4.6667</td>
<td>21.7779</td>
<td>1.47</td>
<td>1.24</td>
</tr>
<tr>
<td>3</td>
<td>2,339.0 2,348.0 2,350.0</td>
<td>2,345.6667</td>
<td>5.8595</td>
<td>34.3337</td>
<td>4.0000</td>
<td>16.0000</td>
<td>-1.26</td>
<td>1.54</td>
</tr>
<tr>
<td>4</td>
<td>2,347.0 2,353.0 2,354.0</td>
<td>2,351.3333</td>
<td>3.7859</td>
<td>14.3330</td>
<td>1.6667</td>
<td>2.7779</td>
<td>0.52</td>
<td>0.99</td>
</tr>
<tr>
<td>5</td>
<td>2,350.0 2,351.0 2,351.0</td>
<td>2,350.6667</td>
<td>0.5774</td>
<td>0.3334</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.31</td>
<td>0.15</td>
</tr>
<tr>
<td>6</td>
<td>2,344.0 2,349.0 2,347.0</td>
<td>2,346.6667</td>
<td>2.5166</td>
<td>6.3333</td>
<td>-3.0000</td>
<td>9.0000</td>
<td>-0.94</td>
<td>0.66</td>
</tr>
</tbody>
</table>

\[ \sum X = 14,098.0000 \]
\[ \sum s^2 = 87.0003 \]
\[ \sum d^2 = 50.6671 \]

#### 4.2.3. Laboratory Statistics

##### 4.2.3.(i) Laboratory Average, \( \bar{x} \)

Compute the average test result for each laboratory for each material using the following equation:

\[
\bar{x} = \frac{\sum x}{n}
\]

Where:

- \( \bar{x} \) = the average of the test results for a material for each laboratory,
- \( x \) = individual test result for a material in each laboratory,
- \( n \) = the number of test results for a material in each laboratory.

For the testing done by Laboratory 1 on Material A:
4.2.3.(ii) Laboratory Standard Deviation, s

Compute the standard deviation for testing done by each laboratory on each material using the following equation:

\[
 s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

Where:
- \( s \) = the laboratory standard deviation for a material,
- \( \bar{x} \) = the average of the test results for a material for each laboratory,
- \( x_i \) = individual test result for a material in each laboratory,
- \( n \) = the number of test results for a material in each laboratory.

For the testing done by Laboratory 1 on Material A:

\[
 s = \sqrt{\frac{1}{2} \sum_{i=1}^{3} (2,359.0 - 2,362.0000)^2 + (2,359.0 - 2,362.0000)^2 + (2,368.0 - 2,362.0000)^2} = 5.1962
\]

4.2.4. Intermediate Statistics

4.2.4.(i) Average of the Laboratory Averages, \( \bar{x} \)

Calculate the average of the laboratory averages for one material using the following formula:

\[
 \bar{x} = \frac{1}{p} \sum_{i=1}^{p} \bar{x}_i
\]

Where:
- \( \bar{x} \) = the average of the laboratory averages for one material,
- \( \bar{x}_i \) = the average of the test results for a material for each laboratory,
- \( p \) = the number of laboratories involved in the interlaboratory testing program.

For Material A:
Laboratory Deviation, \( d \)

Calculate each laboratory's deviation by subtracting the laboratory average from the average of the laboratory averages using the following equation:

\[
d = \bar{x} - \bar{x}
\]

Where:
- \( d \) = the deviation for each laboratory from the average of laboratory averages for one material,
- \( \bar{x} \) = the average of the laboratory averages for one material,
- \( \bar{X} \) = the average of the test results for a material for each laboratory.

For the testing done by Laboratory 1 on Material A:

\[
d = \bar{x} - \bar{x} = 2,362.0000 - 2,357.2222 = 4.7778
\]

Standard Deviation of the Laboratory Averages, \( s_{\bar{x}} \)

Calculate the standard deviation of the laboratory averages for each material using the following equation:

\[
s_{\bar{x}} = \sqrt{\frac{\sum d^2}{p}}
\]

Where:
- \( s_{\bar{x}} \) = the standard deviation of the laboratory averages for one material,
- \( d \) = the deviation for each laboratory from the average of laboratory averages for one material,
- \( p \) = the number of laboratories involved in the interlaboratory testing program.

For Material A:

\[
s_{\bar{x}} = \sqrt{\frac{\sum d^2}{p}} = \sqrt{\frac{(4.7778)^2 + (8.7778)^2 + (0.4444)^2 + (6.1111)^2 + (-21.8889)^2 + (1.7778)^2}{6}} = 11.1329
\]
4.2.5. Precision Statistics

The fundamental precision statistics of an interlaboratory testing program are the repeatability standard deviation and the reproducibility standard deviation. Other statistics are introduced later in this testing procedure.

4.2.5.(i) Repeatability Standard Deviation, $s_r$

The repeatability standard deviation is a measure of the variability that can be expected within a laboratory under repeatability conditions.

The laboratory standard deviation, $s$, is a measure of the within laboratory variability of each individual laboratory for each material. All laboratories are assumed to have essentially the same level of variability when following the specified repeatability conditions. This assumption is not always accurate. The shorter the period of time in which the test results for a particular material are obtained, by the laboratories, the more likely this assumption is accurate. The laboratory variances can be pooled by averaging the squares of the laboratory standard deviations. The square root of the average within laboratory variance is the repeatability standard deviation. It is calculated for each material with the following equation:

$$s_r = \sqrt{\frac{\sum_{i=1}^{p} s_i^2}{p}}$$

Where:
- $s_r =$ the repeatability standard deviation for each material,
- $s =$ the laboratory standard deviation for each material,
- $p =$ the number of laboratories involved in the interlaboratory testing program.

For Material A:

$$s_r = \sqrt{\frac{\sum_{i=1}^{6} s_i^2}{6}} = \sqrt{\frac{(5.1962)^2 + (7.9373)^2 + (6.6583)^2 + (4.0415)^2 + (1.1547)^2 + (1.7321)^2}{6}} = 5.0827$$

4.2.5.(ii) Reproducibility Standard Deviation, $s_R$

The reproducibility standard deviation is a measure of between laboratory variability.
The test results obtained on a particular material at any particular laboratory are considered to be part of a population having a normal distribution with a standard deviation equal to the repeatability standard deviation but with a mean that may be different for each laboratory. The laboratory means are also assumed to vary according to a normal distribution, whose mean is estimated by the average of all the interlaboratory test results for a given material and with a standard deviation that is dependent upon $s_r$ and $s_x$. 

The variance among individual test results obtained in different laboratories is the sum of the within laboratory variance and the between laboratory variance of the laboratory means. It is calculated for each material by first computing an interim calculation:

$$ s_{R}^{*} = \sqrt{(s_{x})^2 + \frac{(s_{r})^2(n-1)}{n}} $$

When $s_{R}^{*}$ is calculated to be less than $s_r$, then $s_{R}$ is equal to $s_r$:

If $s_{R}^{*} \leq s_r$, then $s_{R} = s_r$

When $s_{R}^{*}$ is calculated to be more than $s_r$, then $s_{R}$ is equal to $s_{R}^{*}$:

If $s_{R}^{*} > s_r$, then $s_{R} = s_{R}^{*}$

Where:

- $s_R =$ the reproducibility standard deviation for each material,
- $s_{R}^{*} =$ interim reproducibility standard deviation calculation,
- $s_{r} =$ the repeatability standard deviation for each material,
- $s_{x} =$ the standard deviation of the laboratory averages for one material,
- $n =$ the number of test results for a material in each laboratory.

For Material A:

$$ s_{R}^{*} = \sqrt{(s_{x})^2 + \frac{(s_{r})^2(n-1)}{n}} = \sqrt{(11.1329)^2 + \frac{(5.0827)^2(3-1)}{3}} = 11.8812 $$

Since $s_{R}^{*} = 11.8812 > s_r = 5.0827$, then $s_{R} = s_{R}^{*} = 11.8812$.

4.2.6. Consistency Statistics

Consistency statistics are calculated to evaluate how one laboratory compares to all of the other laboratories involved in the interlaboratory testing program. The
“h” consistency statistic and the “k” consistency statistic are computed for each laboratory for each material.

4.2.6.(i) Between Laboratory Consistency Statistic, $h$

The $h$ consistency statistic is an indicator of how one laboratories testing average for a particular material compares with the average of the other laboratories (excluding the one being considered). The $h$ consistency statistic is computed for each laboratory for each material using the following formula:

$$h = \frac{d}{s_x}$$

Where:
- $h$ = the between laboratory consistency statistic,
- $d$ = the deviation for each laboratory from the average of laboratory averages for one material,
- $s_x$ = the standard deviation of the laboratory averages for one material.

For the testing done by Laboratory 1 on Material A:

$$h = \frac{4.7778}{11.1329} = 0.43$$

4.2.6.(ii) Critical Between Laboratory Consistency Statistic, $h_{crit}$

The $h$ consistency statistic can be compared to a critical value, $h_{crit}$. The critical value is determined from a student’s $t$-test at a 0.5% level of significance using the following formula:

$$h_{crit} = \frac{(p - 1) t}{\sqrt{p(t^2 + p - 2)}}$$

Where:
- $h_{crit}$ = the critical between laboratory consistency statistic,
- $p$ = the number of laboratories involved in the interlaboratory testing program,
- $t$ = Student’s $t$ value. Determined from Table 9 for a 0.5% two-tailed level of significance and $\nu$ degrees of freedom.
- $\nu$ = degrees of freedom = $p - 2$

For the testing done on Material A with six laboratories:

$$\nu = p - 2 = 6 - 2 = 4$$
For $\nu = 4$ and a 0.5% two tailed level of significance from Table 9:

$$t = 5.5975$$

Then $h_{crit}$ can be found by:

$$h_{crit} = \frac{(p-1) t}{\sqrt{pt^2 + p - 2}} = \frac{(6-1) 5.5975}{\sqrt{6[(5.5975)^2 + 6 - 2]}} = \frac{27.9875}{\sqrt{211.9920}} = 1.92$$

4.2.6.(iii) Within Laboratory Consistency Statistic, $k$

The $k$ consistency statistic is an indicator of how one laboratories within laboratory variability, under repeatability conditions, on a particular material compares with the within laboratory variability of all the other laboratories combined. The $k$ consistency statistic is computed for each laboratory for each material using the following formula:
\[ k = \frac{s}{s_r} \]

Where:
\( k \) = the within laboratory consistency statistic,
\( s \) = the laboratory standard deviation for each material,
\( s_r \) = the repeatability standard deviation for each material.

For the testing done by Laboratory 1 on Material A:

\[ k = \frac{5.1962}{5.0827} = 1.02 \]

4.2.6.(iv) Critical Within Laboratory Consistency Statistic, \( k_{crit} \)

Values of \( k \) greater than 1 indicate greater within laboratory variability than the average within laboratory variability for all laboratories. This type of variation among laboratories is expected. Critical \( k \) values can be calculated to determine if the standard deviation for one laboratory on a particular material is sufficiently different from the rest of the laboratories that it should be investigated. The \( k \) consistency statistic can be compared to the critical value, \( k_{crit} \), that is determined from an F-test-test at a 0.5% level of significance. The critical \( k \) value is calculated using the following formula:

\[
k_{crit} = \frac{p}{\sqrt{1 + \left( \frac{p-1}{F} \right)}}
\]

Where:
\( k_{crit} \) = the critical within laboratory consistency statistic,
\( p \) = the number of laboratories involved in the interlaboratory testing program,
\( F \) = Value for the F-ratio. Determined from Table 10 for a 0.5% level of significance with \( \nu_1 \) and \( \nu_2 \) degrees of freedom,
\( \nu_1 \) = degrees of freedom 1 = \( n - 1 \)
\( \nu_2 \) = degrees of freedom 2 = \( (p - 1)(n - 1) \)
\( n \) = the number of test results for a material in each laboratory.

The degrees of freedom are dependent on \( p \), the number of laboratories involved, and \( n \), the number of test results in each laboratory for that material. For the testing done on Material A with six laboratories and three tests in each laboratory:
\[ v_1 = n - 1 = 3 - 1 = 2 \]
\[ v_2 = (p - 1)(n - 1) = (6 - 1)(3 - 1) = 10 \]

For \( v_1 = 2 \) and \( v_2 = 10 \) and a 0.5% level of significance from Table 10:

\[ F = 9.4269 \]

Then \( k_{crit} \) can be found by:

\[
 k_{crit} = \sqrt{\frac{p}{1 + \frac{(p-1)}{F}}} = \sqrt{\frac{6}{1 + \frac{(6-1)}{9.4269}}} = 1.98
\]

4.3. Analyzing Results

4.3.1. Tabular and Graphical Display of Statistics

The data from the interlaboratory testing program can be arranged into tables and graphs so that it is easier to interpret.

4.3.1.(i) Tables

The \( h \) and \( k \) consistency statistics from Table 2, Table 3 and Table 4 should be tabulated into separate tables for the \( h \) and \( k \) consistency statistics as shown in Table 5 and Table 6. It is best to arrange the materials in increasing value of property level, as indicated by \( \bar{x} \), from left to right. The laboratories should be arranged in order of laboratory code number.

<table>
<thead>
<tr>
<th>Laboratory</th>
<th>Material C</th>
<th>Material A</th>
<th>Material B</th>
<th>( h_{crit} )</th>
<th>( h_{crit} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.10</td>
<td>0.43</td>
<td>-0.64</td>
<td>1.92</td>
<td>-1.92</td>
</tr>
<tr>
<td>2</td>
<td>1.47</td>
<td>0.79</td>
<td>0.82</td>
<td>1.92</td>
<td>-1.92</td>
</tr>
<tr>
<td>3</td>
<td>-1.26</td>
<td>0.04</td>
<td>-0.13</td>
<td>1.92</td>
<td>-1.92</td>
</tr>
<tr>
<td>4</td>
<td>0.52</td>
<td>0.55</td>
<td>1.51</td>
<td>1.92</td>
<td>-1.92</td>
</tr>
<tr>
<td>5</td>
<td>0.31</td>
<td>-1.97</td>
<td>-0.35</td>
<td>1.92</td>
<td>-1.92</td>
</tr>
<tr>
<td>6</td>
<td>-0.94</td>
<td>0.16</td>
<td>-1.22</td>
<td>1.92</td>
<td>-1.92</td>
</tr>
</tbody>
</table>

Critical Value, \( h_{crit} = 1.92 \)

<table>
<thead>
<tr>
<th>Laboratory</th>
<th>Material C</th>
<th>Material A</th>
<th>Material B</th>
<th>( k_{crit} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.80</td>
<td>1.02</td>
<td>0.37</td>
<td>1.98</td>
</tr>
<tr>
<td>2</td>
<td>1.24</td>
<td>1.56</td>
<td>1.09</td>
<td>1.98</td>
</tr>
<tr>
<td>3</td>
<td>1.54</td>
<td>1.31</td>
<td>1.20</td>
<td>1.98</td>
</tr>
<tr>
<td>4</td>
<td>0.99</td>
<td>0.80</td>
<td>1.22</td>
<td>1.98</td>
</tr>
<tr>
<td>5</td>
<td>0.15</td>
<td>0.23</td>
<td>0.11</td>
<td>1.98</td>
</tr>
<tr>
<td>6</td>
<td>0.66</td>
<td>0.34</td>
<td>1.32</td>
<td>1.98</td>
</tr>
</tbody>
</table>

Critical Value, \( k_{crit} = 1.98 \)

4.3.1.(ii) Graphs

Bar graphs for the \( h \) and \( k \) statistics should be prepared in two ways. The materials should be grouped by laboratories as shown in Figure 1 and Figure 2 and the laboratories should be grouped by materials as shown in Figure 3 and Figure 4. The laboratories and materials within
each grouping should be arranged in the same order as Table 5 and Table 6.

**Figure 1: h Consistency Statistic by Laboratory**

![Figure 1: h Consistency Statistic by Laboratory](image)

**Figure 2: k Consistency Statistic by Laboratory**

![Figure 2: k Consistency Statistic by Laboratory](image)
4.3.2. Checking for Inconsistent Results

4.3.2.(i) Critical Values of the Consistency Statistics

The critical values of the $h$ and $k$ consistency statistics can be calculated using the procedure previously described. The $h$ critical values depend on the number of laboratories, $p$, participating in the study. The $k$ critical values depend on both the number of laboratories,
The number of test results per laboratory per material, \( n \), and the number of test results per laboratory per material, \( p \). The 0.5\% level of significance was chosen on the basis that too many laboratories are flagged at the 1.0\% level of significance and not enough laboratories are flagged at the 0.1\% level of significance. At the 0.5\% level of significance there is a 0.5\% chance (1 time in 200) that a laboratory will be identified as having different test results from the other labs (some type of testing problem) when in fact its test results are the same as the other labs (there is no testing problem).

After the critical \( h \) and \( k \) values are calculated, examine the \( h \) and \( k \) statistics reported in Table 5 and Table 6. Underline any values that exceed the critical values and circle any values that approach the critical values. A horizontal line representing the critical value should be drawn on each of the graphs. There will be one line for the \( k \) statistic graphs and two lines (critical values can be either positive or negative) on the \( h \) statistic graphs. Any bars that exceed or approach the critical values should be investigated.

In this example, the testing done by Laboratory 5 on Material A should be investigated because the \( h \) value exceeds the \( h_{crit} \) value.

4.3.2.(ii) \( h \) Consistency Statistic Graph by Laboratory (Figure 1)

The four general patterns to look for in this type of graph are:

1. All the laboratories have both positive and negative \( h \) values among the materials. This pattern is acceptable and no investigation is required.
2. The \( h \) values for individual laboratories tend to be either positive or negative for all the materials and the number of negative laboratories equals the number of positive laboratories, more or less. This pattern is acceptable and no investigation is required.
3. One laboratory with all positive (or negative) \( h \) values is opposed to all the other laboratories, with substantially all the \( h \) values negative (or positive). This pattern is not acceptable and that laboratory must be investigated.
4. In one laboratory, the \( h \) values for low property levels are of one sign and for high property levels are of the opposite sign. This laboratory must be investigated if the values are extreme.

In Figure 1, Laboratory 2 and Laboratory 4 both have all positive test results for each material. This could indicate some type of testing problem and should be investigated.

Again referring to Figure 1, the testing done by Laboratory 5 on Material A results in strongly negative \( h \) values. The rest of the
laboratories have all positive $h$ values for Material A. This indicates that there is some type of testing problem and should be investigated.

In Figure 1, the $h$ values change in sign (positive to negative or negative to positive) for different property levels in all the laboratories except Laboratory 2 and Laboratory 4. However, the sign change is not from one extreme value to another extreme value, so there is no need to take any action in these laboratories.

4.3.2.(iii) $k$ Consistency Statistic Graph by Laboratory (Figure 2)

The primary pattern to look for in this type of graph is:

1. One laboratory has large $k$ values (or very small $k$ values) for all or most of the materials. High $k$ values represent within laboratory imprecision. Very small $k$ values may indicate a very insensitive measurement scale or other measurement problem. In either case, the laboratory should be investigated.

In this example, Laboratory 5 has very low $k$ values compared to all of the other laboratories, Figure 2. This may indicate some type of testing problem and should be investigated.

4.3.2.(iv) $h$ or $k$ Consistency Statistic Graphs by Material (Figure 3 and Figure 4)

When a graph by laboratory (Figure 1 or Figure 2) shows several $h$ or $k$ values near the critical value line, the corresponding material graph (Figure 3 or Figure 4) should be reviewed to see how that laboratory differs from the rest of the laboratories for a given material. The patterns to look for in these graphs are:

1. A high value that seems strong in the graph by laboratory, because of its relation to the values for the other materials, will turn out to be relatively consistent with the other laboratories for the same material. This pattern is acceptable and no investigation is required.

2. One laboratory has an $h$ or $k$ value that is strongly different from the other laboratories’ values. This laboratory should be investigated.

In this example, the $h$ statistic graph by laboratory, Figure 1, indicated that the testing done by Laboratory 5 on Material A exceeded the critical level. The $h$ statistic graph by material, Figure 3, revealed that $h$ value for Laboratory 5 is larger than the other laboratories for Material A. This laboratory should be investigated.

In the $k$ statistic graph by laboratory, Figure 2, indicated that Laboratory 5 had low $k$ values for all of the materials. The $k$ statistic graph by material, Figure 4, shows that the $k$ value for Laboratory 5 is
lower than the other laboratories for each material. This laboratory should be investigated.

5. INVESTIGATIONS, CORRECTIONS, PRECISION AND BIAS

5.1. Investigation of Errors

5.1.1. Clerical and Sampling Errors

Examine the laboratory report for each flagged $h$ or $k$ value. Attempt to determine where each test result begins to deviate from the others. Is it in the original observations? Are the data rounded prematurely? Are the calculations correct?

Look for signs of mislabeling test units to see if the test results for one material were reported as belonging to another material. These errors should be checked with the laboratories involved.

5.1.2. Procedural Errors

The laboratory records should be studied to identify any deviations from either the standard testing procedure or the protocol established for the interlaboratory testing program. For example, variations in the number of significant digits reported in the test results may be a sign of incorrect rounding or that equipment in one laboratory is different from the rest. The event log should be studied for any special comments relating to the flagged $h$ or $k$ consistency statistics.

5.1.3. Dealing with Errors

If the investigation disclosed no clerical, sampling or procedural errors, the unusual data should be retained and the precision statistics based on them should be published.

If the laboratory clearly and seriously deviated from the standard test procedure, the test results for that laboratory must be removed from the calculations for the interlaboratory testing program. It may be appropriate to ask the laboratory to retest one or more materials following the correct standard test procedure. There is a danger that the laboratory may apply the standard test procedure in the same manner the second time. If the laboratory is allowed to retest, the data will change and it will be necessary to recalculate all the $h$ and $k$ consistency statistics and reexamine the data for inconsistent results.

If the investigation can not determine a reason for some unusual test results for a particular laboratory and material, it may be appropriate to delete the test results for that laboratory and material. This will change the data, so it will be necessary to recalculate all the $h$ and $k$ consistency statistics for the affected materials and reexamine the data for inconsistent results. It should be noted that data should only be deleted when there is a large number of laboratories involved in the
interlaboratory testing program. It is difficult to state the exact number of laboratories that can be considered large enough to support deletion of data without an identified cause. Generally, if more than 5% of the interlaboratory testing program data is discarded, it is likely that the precision statistics will be incorrect under routine test procedure application.

During a laboratory investigation, the standard test procedure should be evaluated for vagueness that could allow a wide range of interpretation which could lead to a loss of precision. Specific elements that could cause problems are:

- lack of measurement tolerances,
- diversity of testing equipment,
- insufficient direction for operator technique.

These problems can be the basis for a revising the standard test procedure.

5.2. **Precision and Bias**

5.2.1. **Repeatability and Reproducibility Limits**

After the test results are corrected and the statistics are recalculated, the 95% repeatability and reproducibility limits\(^2\) may be calculated for each material using the following equations:

\[
\begin{align*}
  r &= 2.8 \times s_r \\
  R &= 2.8 \times s_R
\end{align*}
\]

Where:
- \(r\) = 95% repeatability limit (within a laboratory).
- \(s_r\) = the repeatability standard deviation for each material.
- \(R\) = 95% reproducibility limit (between laboratories).
- \(s_R\) = the reproducibility standard deviation for each material.

The multiplier (2.8) is based on the Difference "Two"-Standard Deviation Limit \((d2s)\). Approximately 95% of all pairs of test results from laboratories similar to those in the interlaboratory testing program can be expected to differ in absolute value by less than \(1.96 \times \sqrt{2} \times s = 2.77 \times s = 2.8 \times s\). This index is also known as the 95% limit on the difference between two test results.\(^2\)

Another method that is commonly used is the "Two"-Standard Deviation Limits \((2s)\). Approximately 95% of individual test results from laboratories similar to those in the interlaboratory testing program can be expected to differ in absolute value from their average value by less than \(1.96 \times s\).\(^3\)
The 1.96 multiplier assumes that the test results being compared are normally distributed. For methods in which the average of several test results is reported as a single test result the assumption of normality is usually valid. When normality cannot be assumed, it is usually acceptable to use the 1.96 multiplier recognizing that the actual probability level may differ slightly from the nominal 95% limit.

The use of the 1.96 multiplier requires that the sample standard deviation (s) be equal to the population standard deviation (σ). No within or between laboratory study will yield a sample standard deviation (s) equal to the population standard deviation (σ) unless there are at least 30 laboratories included in the study. The use of the t multiplier, from a student’s t distribution, does not solve this problem. In order to resolve this problem, a range of probabilities around 95% must be accepted as defining the “95% limit”. It has been shown that 1.96 is the best multiplier for achieving the desired (but approximate) 95% coverage. The multiplier is independent of the number of laboratories in the testing program or the number of test results obtained by each laboratory. However, a within or between laboratory testing program must be of reasonably large size in order to provide reliable information on which to base precision statements.

For the testing done on Material A:

\[ r_A = 2.8 \times s_r = 2.8 \times 5.083 = 14.2 \]

\[ R_A = 2.8 \times s_R = 2.8 \times 11.881 = 33.3 \]

5.2.2. Bias

5.2.2.(i) Bias of a Test Procedure

When evaluating the bias of a test procedure, the effect of random measurement error should be minimized. The average of at least 30 or more test results, measured independently, at each property level for each of several relatively uniform materials are required to determine the bias of a test procedure. In addition, an accepted reference value must be established for each material using one of the following methods:

1. a theoretical or established value based on scientific principles,
2. an assigned value based on experimental work of some national or international organization such as the US National Institute of Standards and Technology,
3. a consensus value based on collaborative experimental work under the auspices of a scientific or engineering group
4. an agreed upon value obtained using an accepted reference method.
The bias of a test procedure for a specific material can be calculated by comparing the average of all the test results for that material with the accepted reference value for that material (determined using one of the methods listed above). If there is no accepted reference value, bias can not be determined. However, a maximum value for the bias of a test procedure can be estimated by analyzing the effect of equipment and procedure tolerances on the test results.

5.2.2.(ii) Bias of a Specific Laboratory Relative to the Other Laboratories

For a given material, the bias of a specific laboratory relative to the other laboratories can be calculated by averaging the test results obtained in that laboratory and comparing the result with the average of all the test values for the same material, as shown in Table 7.

The calculations for bias of a particular laboratory compared to the other laboratories are illustrated below, for the testing by Laboratory 1 on Material A:

$$\text{Bias}_{1A} = \frac{2,362.0 - 2,357.2}{2,357.2} \times 100 = 0.20\%$$

<table>
<thead>
<tr>
<th>Material A</th>
<th>Material B</th>
<th>Material C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias A</td>
<td>Bias B</td>
<td>Bias C</td>
</tr>
<tr>
<td>Laboratory 1</td>
<td>2,362.0</td>
<td>-0.34%</td>
</tr>
<tr>
<td>Laboratory 2</td>
<td>2,366.0</td>
<td>0.03%</td>
</tr>
<tr>
<td>Laboratory 3</td>
<td>2,357.7</td>
<td>-0.07%</td>
</tr>
<tr>
<td>Laboratory 4</td>
<td>2,363.3</td>
<td>0.80%</td>
</tr>
<tr>
<td>Laboratory 5</td>
<td>2,335.3</td>
<td>-0.18%</td>
</tr>
<tr>
<td>Laboratory 6</td>
<td>2,359.0</td>
<td>-0.65%</td>
</tr>
</tbody>
</table>

In this case, Laboratory 2 and Laboratory 4 both seem to exhibit a tendency toward higher test results than the other labs. The results of this analysis can not be used to establish a bias for these two laboratories because there is no accepted reference value to which these test results can be compared. This analysis only indicates that these two laboratories tend to report higher test results than the other laboratories. This information which may be useful in determining if a particular laboratory needs to investigate its testing process.
5.2.3. Precision Statement

The precision statement for the test procedure being studied should include the following:

- A brief description of the interlaboratory testing program on which the statement is based. The statement should include:
  1. What materials were tested.
  2. The number of laboratories.
  3. The number of test results per laboratory per material.
  4. The interlaboratory practice (this standard test procedure in most cases).

- A description of any deviation from complete adherence to the standard test procedure being studied for each test result.

- The number of test determinations and their combination to form a test result, if not clearly defined in the body of the standard test procedure.

- A statement of the precision between test results expressed in terms of the 95\% repeatability and 95\% reproducibility limit including any variation of these statistics with the property level (material). Report the repeatability and reproducibility standard deviations among the test results.

- A statement describing what is known about bias, including how the method has been modified to adjust for what is known about bias and that it is now without known bias. If the value of the property being measured can be defined only in terms of the test method (there is no accepted reference value), state this and whether the method is generally accepted as a reference method. The statement may take one of several forms:
  - Since there is no accepted reference material, method, or laboratory suitable for determining the bias for the (insert name of the test procedure being studied), no statement on bias is being made.
  - The test results were compared to the appropriate reference method (list the reference method), and found to give results which were XX\% high/low, as theoretical considerations (list reference to theoretical considerations) would suggest. An adjustment for bias is made in Section X.X, so that the final result is without known bias.
  - Error analysis shows that the absolute value of the maximum systematic error that could result from instrument and other tolerances specified in the test procedure is XX \% of the test result.

An example of a precision statement for the test procedure to determine the 50 Blow Hand Compaction Marshall Density follows:

**Precision Statement for 50 Blow Hand Compaction Marshall Density**
Interlaboratory Test Program - An interlaboratory study was run in which randomly drawn test specimens of three materials were tested for 50 Blow Hand Compaction Marshall Density in each of six laboratories, with each laboratory testing three specimens of each material. Saskatchewan Highways and Transportation Standard Test Procedure XXXX was followed for the design and analysis of the data.

Test Results - The precision information given below in kg/m³ is for the comparison of two test results each of which is the average of three test determinations.

Precision:

<table>
<thead>
<tr>
<th>Material</th>
<th>Average Test Value</th>
<th>95% Repeatability Limit (within laboratory)</th>
<th>95% Reproducibility Limit (between laboratories)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2,357.2 kg/m³</td>
<td>14.2 kg/m³</td>
<td>33.3 kg/m³</td>
</tr>
<tr>
<td>B</td>
<td>2,362.0 kg/m³</td>
<td>15.0 kg/m³</td>
<td>37.2 kg/m³</td>
</tr>
<tr>
<td>C</td>
<td>2,349.7 kg/m³</td>
<td>12.5 kg/m³</td>
<td>12.5 kg/m³</td>
</tr>
</tbody>
</table>

The respective standard deviations among test results may be obtained by dividing the above limit values by 2.8. More detailed precision statistics are shown in Table 8.

Table 8: 50 Blow Hand Compaction Marshall Density Precision Statistics (kg/m³)

<table>
<thead>
<tr>
<th>Material</th>
<th>( \bar{X} )</th>
<th>( s_{\bar{X}} )</th>
<th>( s_r )</th>
<th>( s_R )</th>
<th>( r )</th>
<th>( R )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2,357.2</td>
<td>11.1</td>
<td>5.1</td>
<td>11.9</td>
<td>14.2</td>
<td>33.3</td>
</tr>
<tr>
<td>B</td>
<td>2,362.0</td>
<td>12.6</td>
<td>5.4</td>
<td>13.3</td>
<td>15.0</td>
<td>37.2</td>
</tr>
<tr>
<td>C</td>
<td>2,349.7</td>
<td>4.4</td>
<td>4.4</td>
<td>4.4</td>
<td>12.5</td>
<td>12.5</td>
</tr>
</tbody>
</table>

Bias - Since there is no accepted reference material, method, or laboratory suitable for determining the bias for the 50 Blow Hand Compaction Marshall Density, no statement on bias is being made.

5.2.4. Interpretation

The precision described above means that for materials with the property levels found in Materials A and B, approximately 95% of all pairs of test results from within a laboratory similar to that in the study, can be expected to differ in absolute value by less than 15 kg/m³. For materials with the property level of Material A or B, two test results from the same laboratory would be considered suspect if they differed in absolute value by more than 15 kg/m³.
For materials with the property level found in Material C, approximately 95% of all pairs of test results from within a laboratory similar to that in the study can be expected to differ in absolute value by less than 12.5 kg/m³. For materials with the property level of Material C, two test results from the same laboratory would be considered suspect if they differed in absolute value by more than 12.5 kg/m³.

For materials with the property levels found in Materials A and B, approximately 95% of all pairs of test results from between laboratories similar to those in the study, can be expected to differ in absolute value by less than 37 kg/m³. For materials with the property level of Material A or B, two test results from different laboratories would be considered suspect if they differed in absolute value by more than 37 kg/m³.

For materials with the property level found in Material C, approximately 95% of all pairs of test results from between laboratories similar to those in the study can be expected to differ in absolute value by less than 12.5 kg/m³. For materials with the property level of Material C, two test results from different laboratories would be considered suspect if they differed in absolute value by more than 12.5 kg/m³.

6. **ADDED INFORMATION**

6.1. **Limitations**

The precision statistics obtained by an interlaboratory testing program, such as the one described in this Standard Test Procedure, must not be treated as exact mathematical quantities which are applicable to all circumstances and uses. The small number of laboratories and materials used in the typical interlaboratory testing program guarantees that there will be times when differences greater than that predicted by the interlaboratory testing program will arise. This can occur with considerably greater or smaller frequency than the 95% probability level would imply. The repeatability and reproducibility limit should be considered as general guides and the associated 95% probability as only a rough indicator of what can be expected. If more precise information is needed in a specific case, the laboratories directly involved in the material comparison must conduct interlaboratory studies specifically aimed at the material being studied.

6.2. **References**

This test procedure is similar to ASTM E 691-92 *Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method*. It also incorporates components of ASTM E 177 -90a *Standard Practice for the Use of the Terms Precision and Bias in ASTM Test Methods*. 
<table>
<thead>
<tr>
<th>Section:</th>
<th>Subject:</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATISTICAL QUALITY CONTROL PRINCIPLES</td>
<td>INTERLABORATORY TESTING PROGRAMS</td>
</tr>
</tbody>
</table>

C:\WINDOWS\WINWORK\QUALITY\LABCORR\STP304-3.DOC
### Table 9: Student’s t Distribution

<table>
<thead>
<tr>
<th>Degrees of Freedom (ν)</th>
<th>0.1%</th>
<th>0.5%</th>
<th>1.0%</th>
<th>1.5%</th>
<th>2.0%</th>
<th>2.5%</th>
<th>3.0%</th>
<th>3.5%</th>
<th>4.0%</th>
<th>4.5%</th>
<th>5.0%</th>
<th>10.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6356.6</td>
<td>127.32</td>
<td>63.59</td>
<td>42.43</td>
<td>31.82</td>
<td>25.46</td>
<td>21.05</td>
<td>18.69</td>
<td>16.33</td>
<td>15.00</td>
<td>14.12</td>
<td>12.06</td>
</tr>
<tr>
<td>2</td>
<td>31.59</td>
<td>14.09</td>
<td>9.29</td>
<td>6.02</td>
<td>4.64</td>
<td>3.65</td>
<td>3.03</td>
<td>2.68</td>
<td>2.36</td>
<td>2.11</td>
<td>2.00</td>
<td>1.81</td>
</tr>
<tr>
<td>3</td>
<td>12.59</td>
<td>7.45</td>
<td>5.84</td>
<td>4.07</td>
<td>3.48</td>
<td>3.00</td>
<td>2.67</td>
<td>2.42</td>
<td>2.19</td>
<td>2.00</td>
<td>1.89</td>
<td>1.75</td>
</tr>
<tr>
<td>4</td>
<td>8.61</td>
<td>5.59</td>
<td>4.60</td>
<td>3.80</td>
<td>3.29</td>
<td>2.93</td>
<td>2.63</td>
<td>2.39</td>
<td>2.19</td>
<td>2.00</td>
<td>1.88</td>
<td>1.76</td>
</tr>
<tr>
<td>5</td>
<td>6.86</td>
<td>4.78</td>
<td>3.97</td>
<td>3.34</td>
<td>2.90</td>
<td>2.54</td>
<td>2.27</td>
<td>2.08</td>
<td>1.91</td>
<td>1.79</td>
<td>1.69</td>
<td>1.58</td>
</tr>
<tr>
<td>6</td>
<td>5.96</td>
<td>4.32</td>
<td>3.57</td>
<td>3.04</td>
<td>2.62</td>
<td>2.31</td>
<td>2.07</td>
<td>1.89</td>
<td>1.74</td>
<td>1.63</td>
<td>1.53</td>
<td>1.44</td>
</tr>
<tr>
<td>7</td>
<td>5.41</td>
<td>4.03</td>
<td>3.34</td>
<td>2.89</td>
<td>2.49</td>
<td>2.19</td>
<td>1.96</td>
<td>1.79</td>
<td>1.65</td>
<td>1.54</td>
<td>1.45</td>
<td>1.37</td>
</tr>
<tr>
<td>8</td>
<td>5.04</td>
<td>3.83</td>
<td>3.15</td>
<td>2.74</td>
<td>2.36</td>
<td>2.08</td>
<td>1.85</td>
<td>1.70</td>
<td>1.57</td>
<td>1.47</td>
<td>1.39</td>
<td>1.32</td>
</tr>
<tr>
<td>9</td>
<td>4.76</td>
<td>3.67</td>
<td>2.99</td>
<td>2.59</td>
<td>2.22</td>
<td>1.96</td>
<td>1.74</td>
<td>1.60</td>
<td>1.48</td>
<td>1.39</td>
<td>1.31</td>
<td>1.25</td>
</tr>
<tr>
<td>10</td>
<td>4.56</td>
<td>3.53</td>
<td>2.86</td>
<td>2.48</td>
<td>2.12</td>
<td>1.87</td>
<td>1.66</td>
<td>1.53</td>
<td>1.41</td>
<td>1.33</td>
<td>1.26</td>
<td>1.20</td>
</tr>
<tr>
<td>11</td>
<td>4.38</td>
<td>3.41</td>
<td>2.75</td>
<td>2.38</td>
<td>2.03</td>
<td>1.79</td>
<td>1.59</td>
<td>1.46</td>
<td>1.35</td>
<td>1.27</td>
<td>1.21</td>
<td>1.15</td>
</tr>
<tr>
<td>12</td>
<td>4.22</td>
<td>3.31</td>
<td>2.66</td>
<td>2.30</td>
<td>1.96</td>
<td>1.73</td>
<td>1.54</td>
<td>1.42</td>
<td>1.31</td>
<td>1.24</td>
<td>1.18</td>
<td>1.12</td>
</tr>
<tr>
<td>13</td>
<td>4.08</td>
<td>3.23</td>
<td>2.59</td>
<td>2.25</td>
<td>1.92</td>
<td>1.69</td>
<td>1.51</td>
<td>1.39</td>
<td>1.28</td>
<td>1.21</td>
<td>1.15</td>
<td>1.09</td>
</tr>
<tr>
<td>14</td>
<td>3.96</td>
<td>3.16</td>
<td>2.52</td>
<td>2.19</td>
<td>1.86</td>
<td>1.64</td>
<td>1.47</td>
<td>1.35</td>
<td>1.24</td>
<td>1.18</td>
<td>1.12</td>
<td>1.06</td>
</tr>
<tr>
<td>15</td>
<td>3.85</td>
<td>3.09</td>
<td>2.47</td>
<td>2.15</td>
<td>1.82</td>
<td>1.61</td>
<td>1.44</td>
<td>1.33</td>
<td>1.22</td>
<td>1.16</td>
<td>1.10</td>
<td>1.04</td>
</tr>
<tr>
<td>16</td>
<td>3.75</td>
<td>3.02</td>
<td>2.42</td>
<td>2.09</td>
<td>1.77</td>
<td>1.56</td>
<td>1.40</td>
<td>1.29</td>
<td>1.18</td>
<td>1.13</td>
<td>1.07</td>
<td>1.01</td>
</tr>
<tr>
<td>17</td>
<td>3.66</td>
<td>2.96</td>
<td>2.38</td>
<td>2.05</td>
<td>1.73</td>
<td>1.52</td>
<td>1.37</td>
<td>1.26</td>
<td>1.15</td>
<td>1.09</td>
<td>1.04</td>
<td>0.98</td>
</tr>
<tr>
<td>18</td>
<td>3.57</td>
<td>2.90</td>
<td>2.32</td>
<td>1.99</td>
<td>1.68</td>
<td>1.47</td>
<td>1.33</td>
<td>1.22</td>
<td>1.12</td>
<td>1.06</td>
<td>1.01</td>
<td>0.95</td>
</tr>
<tr>
<td>19</td>
<td>3.50</td>
<td>2.85</td>
<td>2.28</td>
<td>1.95</td>
<td>1.64</td>
<td>1.44</td>
<td>1.29</td>
<td>1.19</td>
<td>1.08</td>
<td>1.02</td>
<td>0.97</td>
<td>0.91</td>
</tr>
<tr>
<td>20</td>
<td>3.43</td>
<td>2.80</td>
<td>2.23</td>
<td>1.90</td>
<td>1.60</td>
<td>1.40</td>
<td>1.26</td>
<td>1.16</td>
<td>1.05</td>
<td>0.99</td>
<td>0.94</td>
<td>0.88</td>
</tr>
<tr>
<td>21</td>
<td>3.37</td>
<td>2.75</td>
<td>2.19</td>
<td>1.86</td>
<td>1.56</td>
<td>1.36</td>
<td>1.22</td>
<td>1.12</td>
<td>1.01</td>
<td>0.95</td>
<td>0.90</td>
<td>0.84</td>
</tr>
<tr>
<td>22</td>
<td>3.32</td>
<td>2.70</td>
<td>2.15</td>
<td>1.82</td>
<td>1.52</td>
<td>1.32</td>
<td>1.18</td>
<td>1.08</td>
<td>0.97</td>
<td>0.91</td>
<td>0.86</td>
<td>0.80</td>
</tr>
<tr>
<td>23</td>
<td>3.27</td>
<td>2.66</td>
<td>2.11</td>
<td>1.78</td>
<td>1.48</td>
<td>1.28</td>
<td>1.14</td>
<td>1.04</td>
<td>0.93</td>
<td>0.87</td>
<td>0.82</td>
<td>0.77</td>
</tr>
<tr>
<td>24</td>
<td>3.23</td>
<td>2.62</td>
<td>2.07</td>
<td>1.74</td>
<td>1.44</td>
<td>1.24</td>
<td>1.10</td>
<td>1.00</td>
<td>0.90</td>
<td>0.84</td>
<td>0.79</td>
<td>0.74</td>
</tr>
<tr>
<td>25</td>
<td>3.19</td>
<td>2.58</td>
<td>2.03</td>
<td>1.69</td>
<td>1.39</td>
<td>1.19</td>
<td>1.05</td>
<td>0.96</td>
<td>0.86</td>
<td>0.80</td>
<td>0.75</td>
<td>0.70</td>
</tr>
<tr>
<td>30</td>
<td>3.06</td>
<td>2.47</td>
<td>1.90</td>
<td>1.52</td>
<td>1.24</td>
<td>1.04</td>
<td>0.90</td>
<td>0.81</td>
<td>0.71</td>
<td>0.65</td>
<td>0.60</td>
<td>0.55</td>
</tr>
<tr>
<td>40</td>
<td>2.95</td>
<td>2.39</td>
<td>1.83</td>
<td>1.44</td>
<td>1.15</td>
<td>0.95</td>
<td>0.81</td>
<td>0.72</td>
<td>0.62</td>
<td>0.56</td>
<td>0.51</td>
<td>0.46</td>
</tr>
<tr>
<td>50</td>
<td>2.88</td>
<td>2.35</td>
<td>1.78</td>
<td>1.38</td>
<td>1.09</td>
<td>0.88</td>
<td>0.75</td>
<td>0.66</td>
<td>0.56</td>
<td>0.50</td>
<td>0.45</td>
<td>0.40</td>
</tr>
<tr>
<td>100</td>
<td>2.81</td>
<td>2.31</td>
<td>1.73</td>
<td>1.32</td>
<td>0.94</td>
<td>0.74</td>
<td>0.61</td>
<td>0.52</td>
<td>0.42</td>
<td>0.36</td>
<td>0.31</td>
<td>0.26</td>
</tr>
<tr>
<td>1000</td>
<td>2.76</td>
<td>2.28</td>
<td>1.67</td>
<td>1.25</td>
<td>0.87</td>
<td>0.67</td>
<td>0.54</td>
<td>0.45</td>
<td>0.35</td>
<td>0.29</td>
<td>0.24</td>
<td>0.19</td>
</tr>
</tbody>
</table>
### Table 10: F Distribution

<table>
<thead>
<tr>
<th>Degrees of Freedom 1 ($v_1$)</th>
<th>0.50%</th>
<th>1.00%</th>
<th>5.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>198.520 072</td>
<td>199.012 000</td>
<td>199.157 552</td>
</tr>
<tr>
<td>3</td>
<td>55.519 803</td>
<td>47.463 803</td>
<td>34.116 1164</td>
</tr>
<tr>
<td>5</td>
<td>22.784 713</td>
<td>16.530 136</td>
<td>12.251 343</td>
</tr>
<tr>
<td>6</td>
<td>18.634 616</td>
<td>12.916 601</td>
<td>10.294 666</td>
</tr>
<tr>
<td>8</td>
<td>14.683 112</td>
<td>9.596 552</td>
<td>8.649 438</td>
</tr>
<tr>
<td>9</td>
<td>13.613 108</td>
<td>8.717 616</td>
<td>8.021 425</td>
</tr>
<tr>
<td>10</td>
<td>12.826 926</td>
<td>8.089 009</td>
<td>7.595 563</td>
</tr>
<tr>
<td>11</td>
<td>12.263 912</td>
<td>7.604 092</td>
<td>7.207 637</td>
</tr>
<tr>
<td>12</td>
<td>11.753 807</td>
<td>7.225 727</td>
<td>6.926 569</td>
</tr>
<tr>
<td>13</td>
<td>11.373 816</td>
<td>6.928 592</td>
<td>6.709 476</td>
</tr>
<tr>
<td>14</td>
<td>11.060 921</td>
<td>6.688 684</td>
<td>6.514 757</td>
</tr>
<tr>
<td>15</td>
<td>10.790 700</td>
<td>6.471 671</td>
<td>6.388 588</td>
</tr>
<tr>
<td>16</td>
<td>10.576 736</td>
<td>6.303 504</td>
<td>6.263 649</td>
</tr>
<tr>
<td>17</td>
<td>10.342 735</td>
<td>6.155 735</td>
<td>6.121 629</td>
</tr>
<tr>
<td>18</td>
<td>10.218 724</td>
<td>6.027 855</td>
<td>6.012 599</td>
</tr>
<tr>
<td>19</td>
<td>10.072 709</td>
<td>5.916 516</td>
<td>6.013 599</td>
</tr>
<tr>
<td>20</td>
<td>9.940 6965</td>
<td>5.817 096</td>
<td>5.849 438</td>
</tr>
<tr>
<td>21</td>
<td>9.829 6915</td>
<td>5.734 704</td>
<td>5.704 870</td>
</tr>
<tr>
<td>22</td>
<td>9.727 6806</td>
<td>5.652 652</td>
<td>5.625 719</td>
</tr>
<tr>
<td>23</td>
<td>9.634 6730</td>
<td>5.582 583</td>
<td>5.563 667</td>
</tr>
<tr>
<td>24</td>
<td>9.531 6699</td>
<td>5.519 519</td>
<td>5.516 734</td>
</tr>
<tr>
<td>25</td>
<td>9.475 6592</td>
<td>5.461 561</td>
<td>5.468 675</td>
</tr>
<tr>
<td>26</td>
<td>9.406 6410</td>
<td>5.409 421</td>
<td>5.423 636</td>
</tr>
<tr>
<td>27</td>
<td>9.342 6486</td>
<td>5.361 561</td>
<td>5.488 609</td>
</tr>
<tr>
<td>28</td>
<td>9.283 6440</td>
<td>5.317 635</td>
<td>5.452 549</td>
</tr>
<tr>
<td>29</td>
<td>9.229 6398</td>
<td>5.276 597</td>
<td>5.420 538</td>
</tr>
<tr>
<td>30</td>
<td>9.179 6354</td>
<td>5.238 576</td>
<td>5.390 507</td>
</tr>
<tr>
<td>35</td>
<td>8.976 6179</td>
<td>5.085 419</td>
<td>5.267 435</td>
</tr>
<tr>
<td>40</td>
<td>8.827 6064</td>
<td>4.975 314</td>
<td>5.185 412</td>
</tr>
<tr>
<td>45</td>
<td>8.714 5974</td>
<td>4.891 723</td>
<td>5.103 424</td>
</tr>
<tr>
<td>50</td>
<td>8.625 5916</td>
<td>4.829 719</td>
<td>5.056 449</td>
</tr>
<tr>
<td>60</td>
<td>8.494 5792</td>
<td>4.729 077</td>
<td>4.977 415</td>
</tr>
<tr>
<td>80</td>
<td>8.336 5652</td>
<td>4.611 962</td>
<td>4.880 403</td>
</tr>
<tr>
<td>100</td>
<td>8.240 5589</td>
<td>4.524 893</td>
<td>4.823 393</td>
</tr>
<tr>
<td>120</td>
<td>8.179 5539</td>
<td>4.497 850</td>
<td>4.785 391</td>
</tr>
<tr>
<td>140</td>
<td>8.135 5490</td>
<td>4.452 819</td>
<td>4.760 392</td>
</tr>
<tr>
<td>150</td>
<td>8.117 5489</td>
<td>4.425 806</td>
<td>4.749 391</td>
</tr>
<tr>
<td>10,000,000</td>
<td>7.879 5393</td>
<td>4.274 634</td>
<td>4.605 378</td>
</tr>
</tbody>
</table>