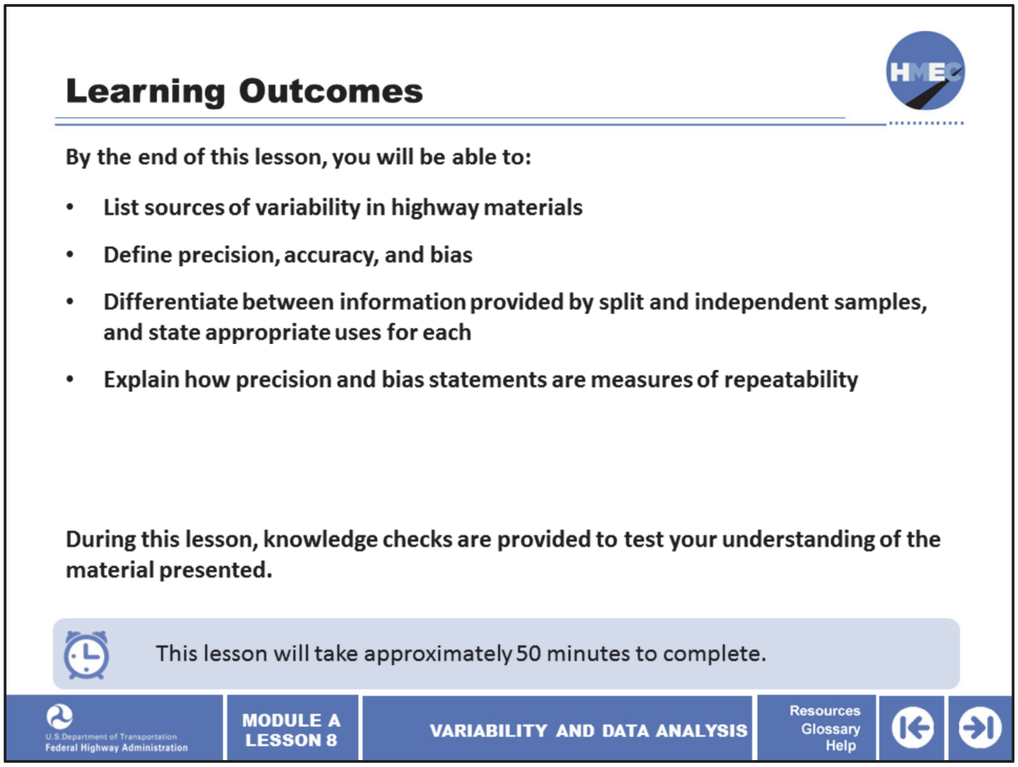
Welcome to the Highway Materials Engineering Course Module A, Lesson 8: Variability and Data Analysis. This lesson will discuss variability in highway materials and provide an understanding of precision, accuracy, and bias.

A printer-friendly version of the lesson materials can be downloaded by selecting the paperclip icon. A copy of the slides and narration are provided for download.

If you need technical assistance during the training, please select the Help link in the upper right-hand corner of the screen.

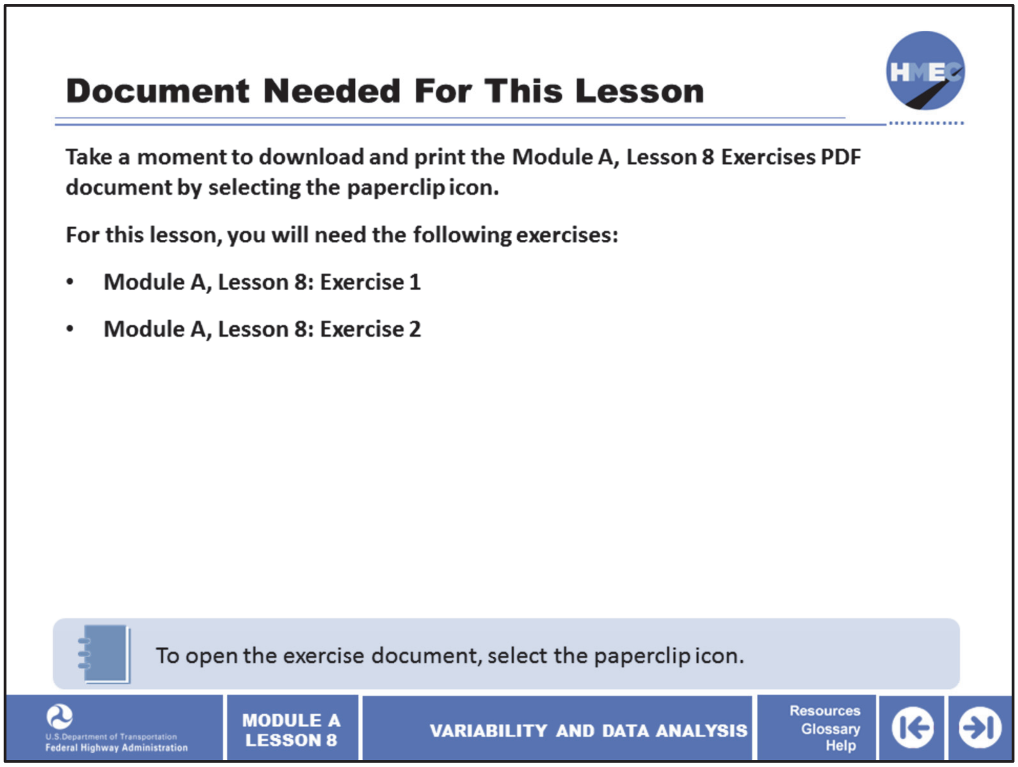


By the end of this lesson, you will be able to:

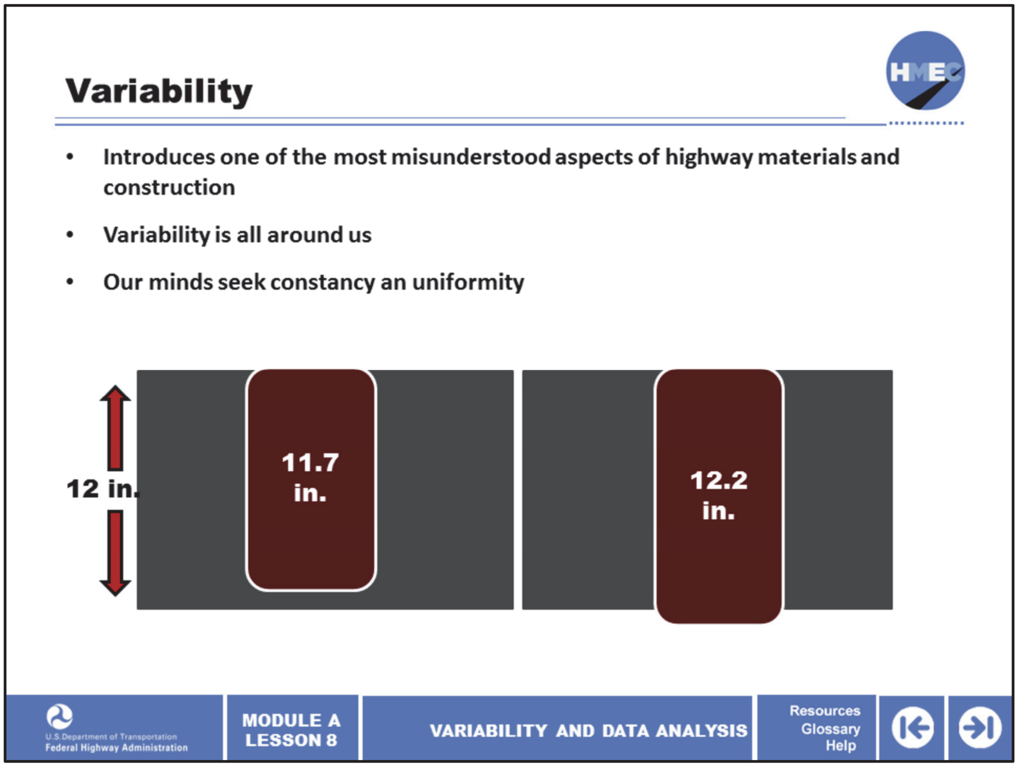
- List sources of variability in highway materials;
- Define precision, accuracy, and bias;
- Differentiate between information provided by split and independent samples, and state appropriate uses for each; and
- Explain how precision and bias statements are measures of repeatability.

During this lesson, knowledge checks are provided to test your understanding of the material presented.

This lesson will take approximately 50 minutes to complete.



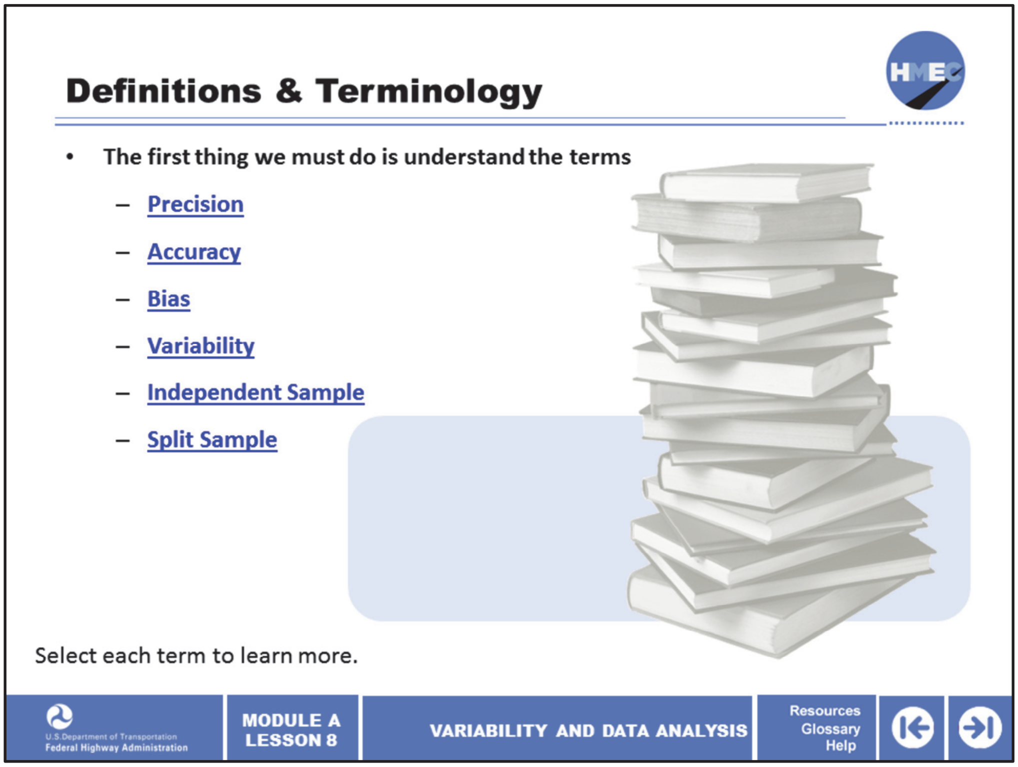
During this lesson, you will be prompted to reference the lesson exercise document. The document referenced during this lesson is attached to the lesson in the paperclip icon. Please take a moment to open and print the document.



Let's get started. This lesson is very important as it introduces one of the most misunderstood aspects of highway materials and construction. We know from experience that variability exists—it is all around us. In fact, constancy is an exception. However, when it comes to highway materials and construction, our minds seek constancy and uniformity.

On the other hand, if we take a sample from a concrete pavement that is designed at 12 in. and the core is 11.7 in., we get concerned that the entire pavement does not meet specifications. Whereas if the core were 12. 2 in. thick, we would be happy that the entire pavement met specifications. Both of these assumptions are incorrect. Because variability exists in everything, we can't get a picture of a population with a sample size of one because there is no measure of variability.

Image description: Sample from a concrete pavement.

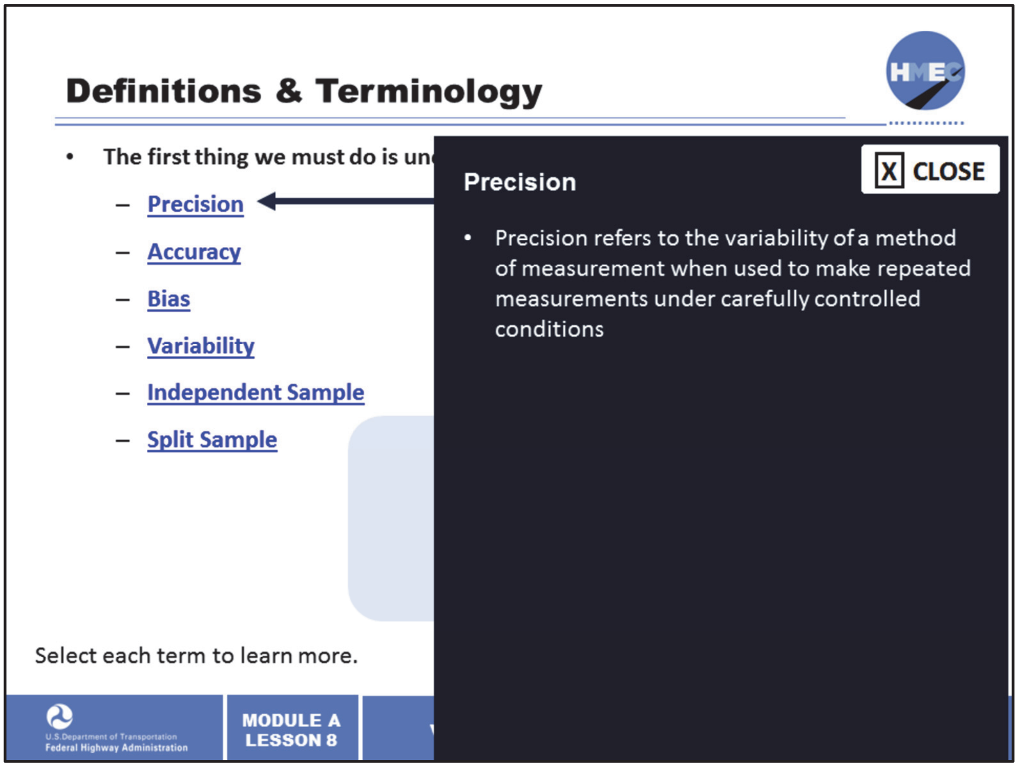


The first thing we must do is understand the terms we will use. You may be familiar with many of these but it is important we're working with the same definition throughout this module. The terms are:

- Precision;
- Accuracy;
- Bias;
- Variability;
- Independent Sample; and
- Split Sample.

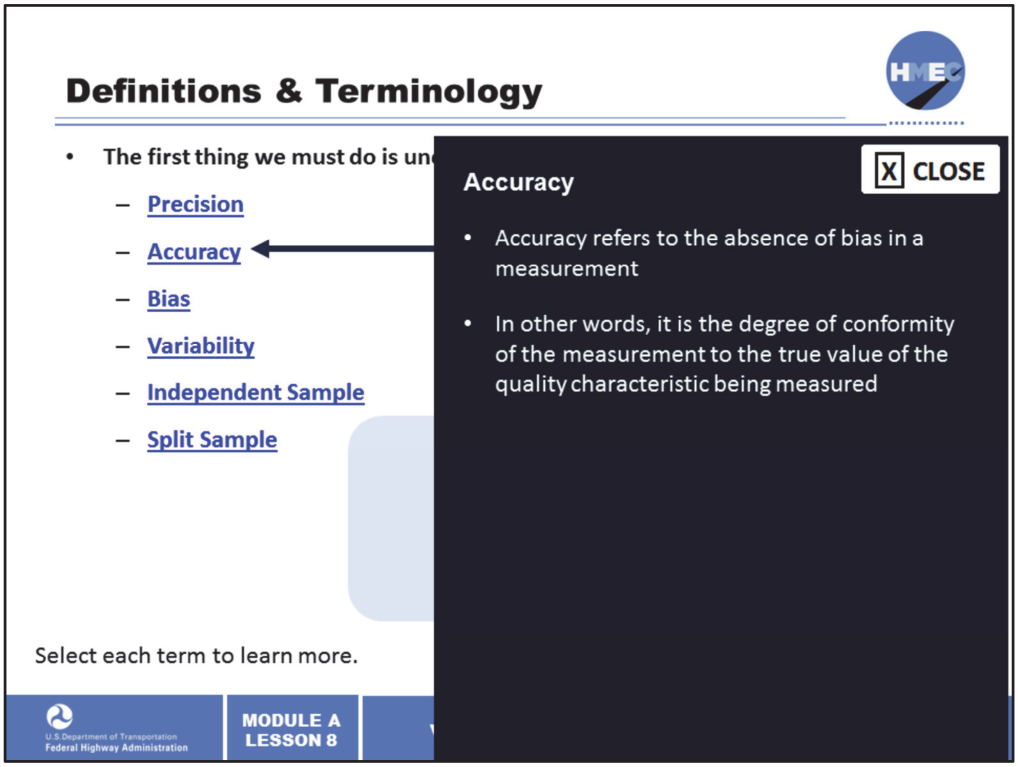
Select each term to learn more.

Image description: Stack of books.



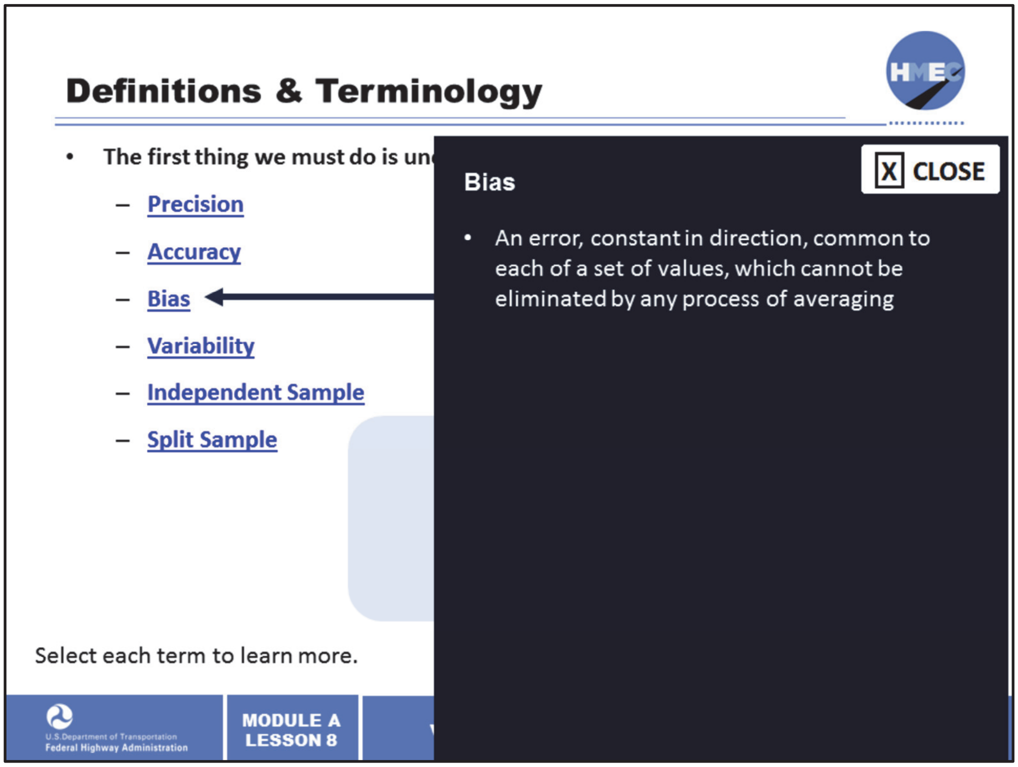
Precision refers to the variability of a method of measurement when used to make repeated measurements under carefully controlled conditions.

Please note that precision and accuracy are often incorrectly used interchangeably. We must remind ourselves that they are not the same.

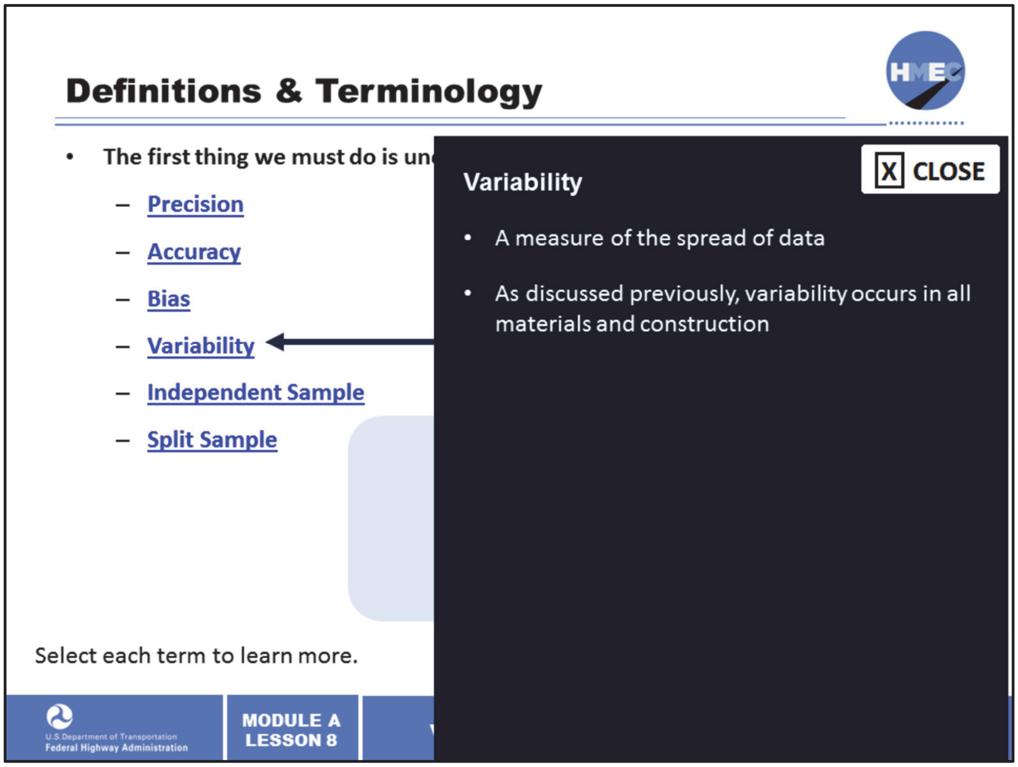


Accuracy refers to the absence of bias in a measurement. In other words, it is the degree of conformity of the measurement to the true value of the quality characteristic being measured.

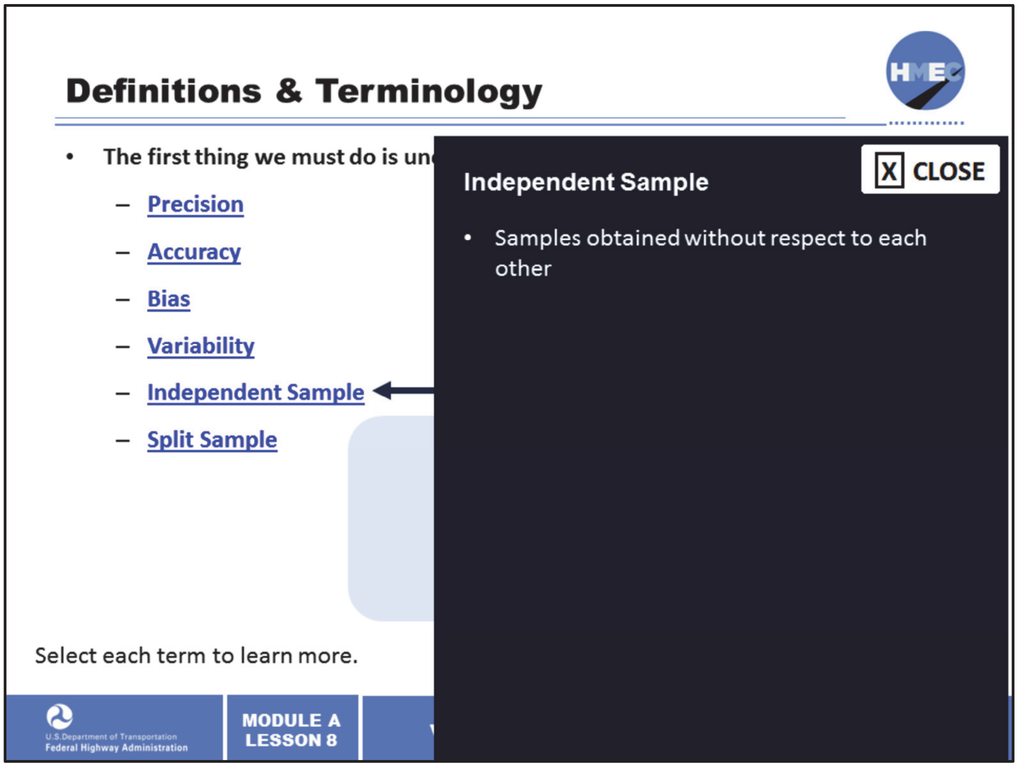
Again, note that precision and accuracy are often incorrectly used interchangeably.



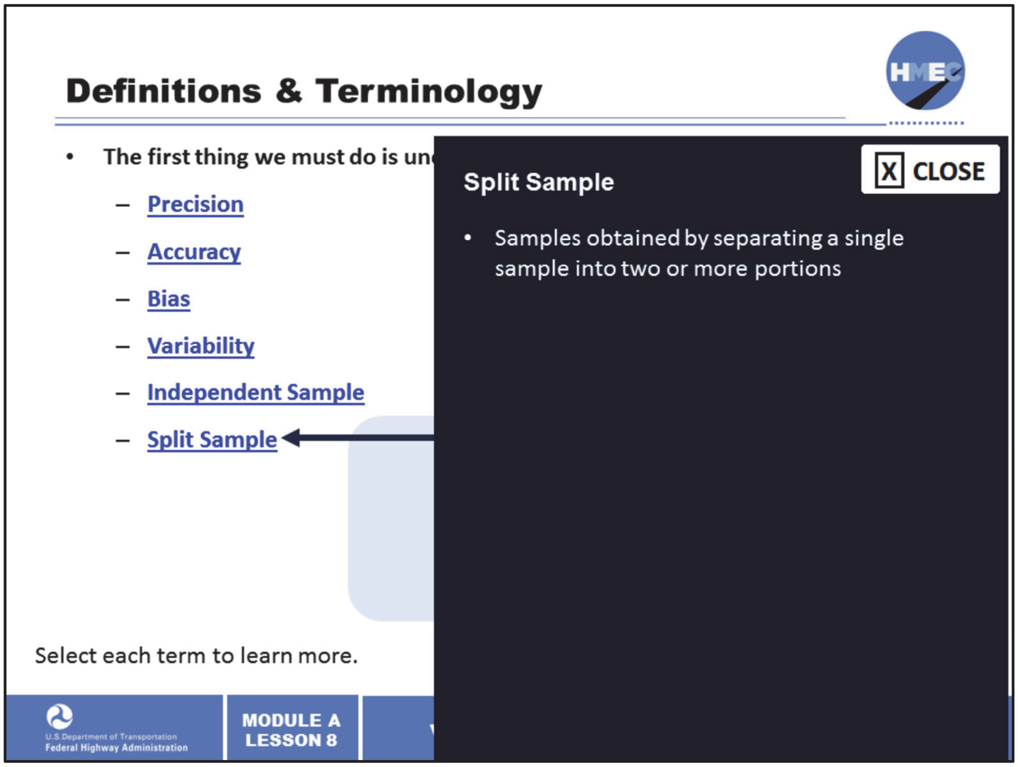
Bias is an error, constant in direction, common to each of a set of values, which cannot be eliminated by any process of averaging. These first three definitions can be most easily visualized by the analogy of a target, which we will discuss in more detail in this lesson.



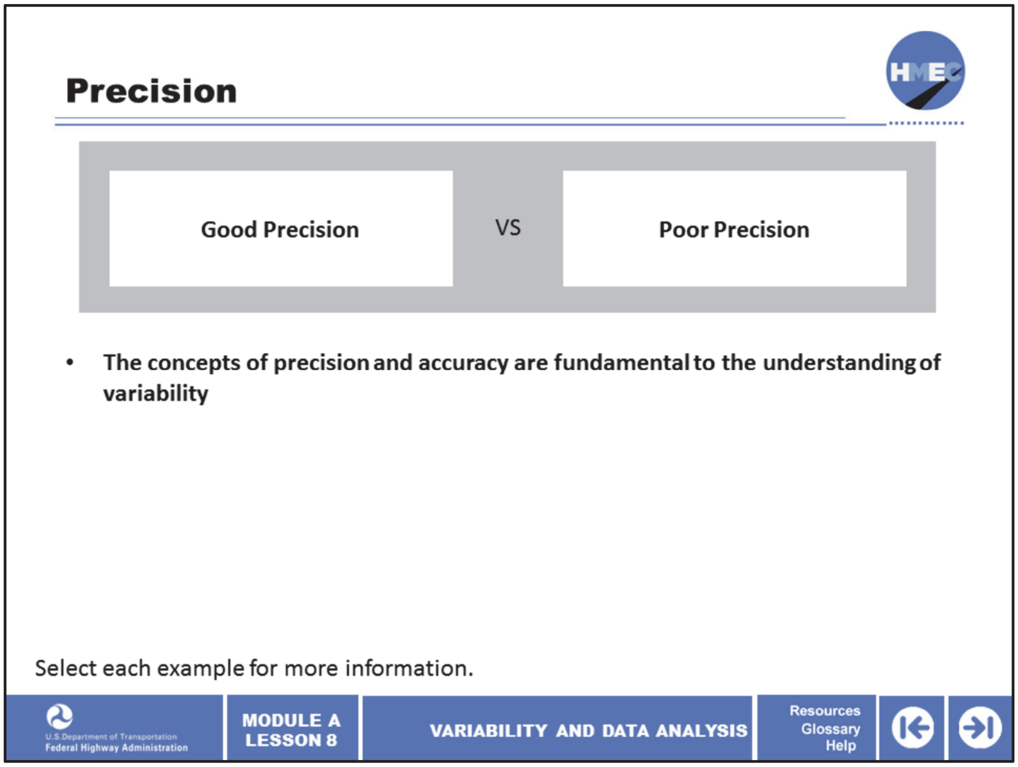
Variability is the measure of the spread of data. As discussed previously, variability is inevitable and it occurs in all materials and construction.



Independent samples are obtained without respect to each other.

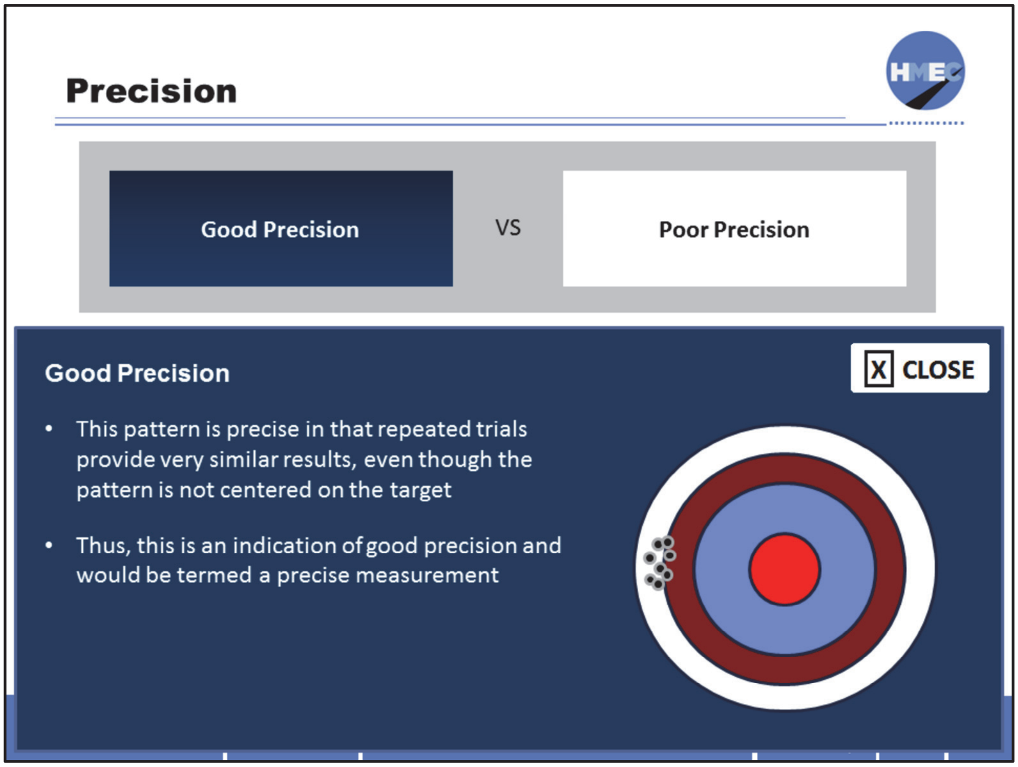


Split samples are obtained by separating a single sample into two or more portions.



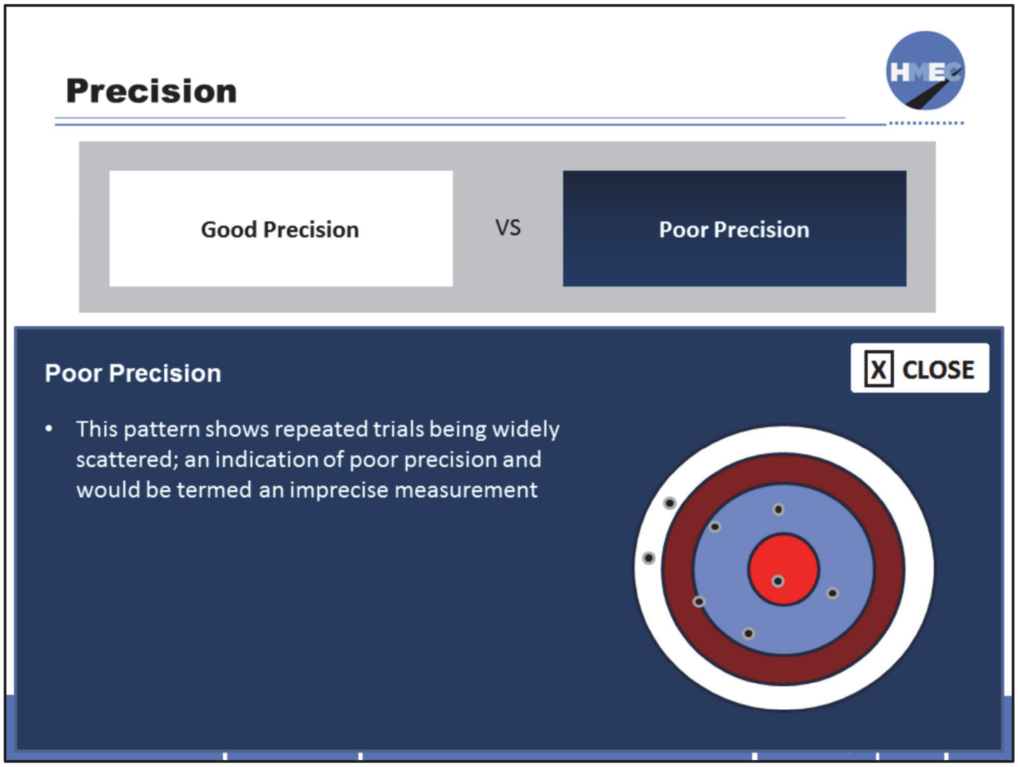
The concepts of precision and accuracy are fundamental to the understanding of variability. As we defined previously, precision refers to the variability of repeat measurements under carefully controlled conditions. In other words, precision is related to the repeatability of a process. The analogy of a target is useful in explaining this definition. The use of this analogy is applicable to a test method.

Select each example for more information.



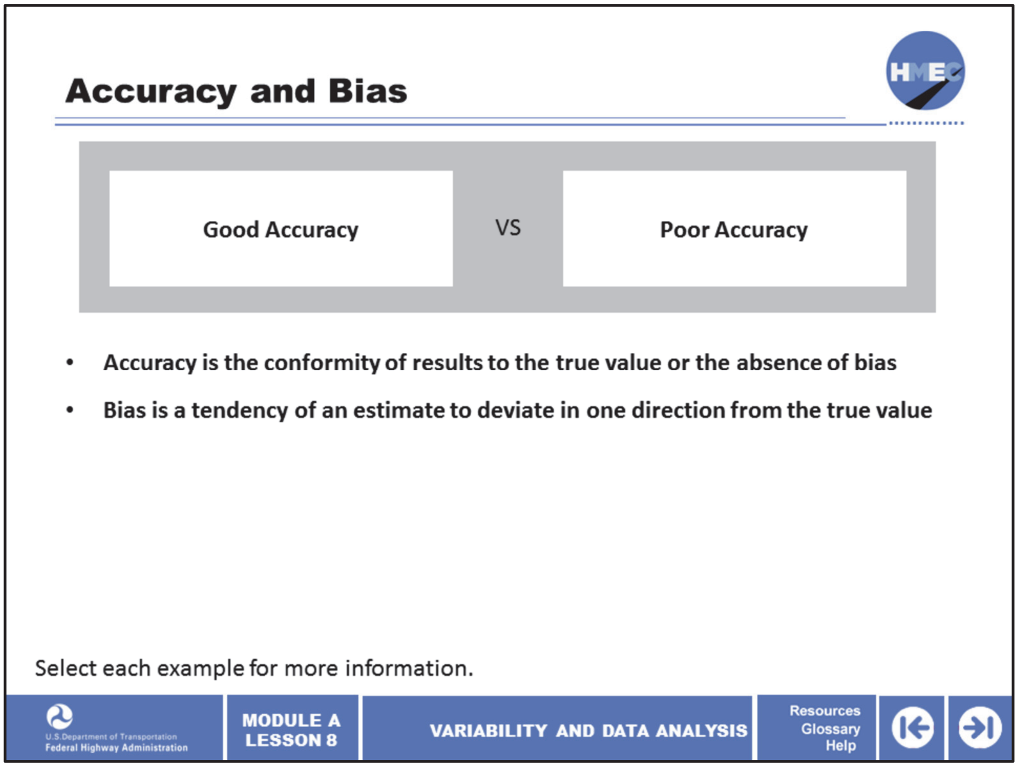
This pattern is precise in that repeated trials provide very similar results, even though the pattern is not centered on the target. Thus, this is an indication of good precision and would be termed a precise measurement.

Image description: Target.



This pattern shows repeated trials being widely scattered; an indication of poor precision and would be termed an imprecise measurement.

Image description: Target.

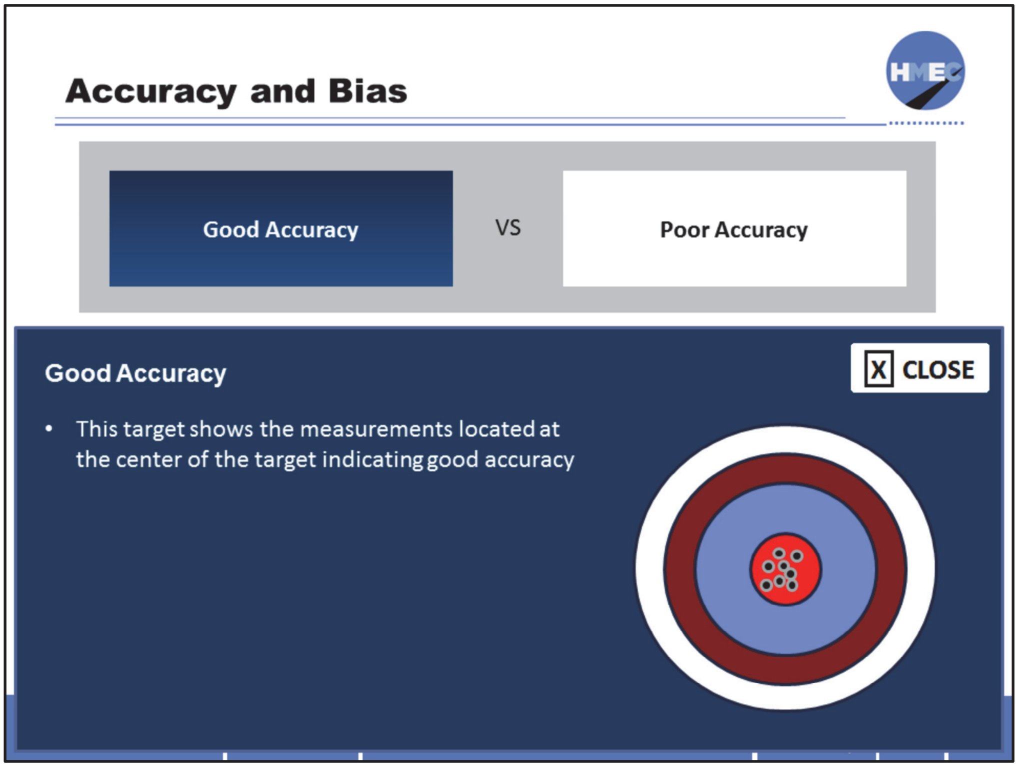


Accuracy is the conformity of results to the true value or the absence of bias. Bias is a tendency of an estimate to deviate in one direction from the true value. Once again, a target analogy is a practical way of understanding the relationship between accuracy and bias.

A practical difficulty in measuring accuracy is that the true value must be determined. Ideally, the true value should be determined by a method of high precision and with as little bias as possible. In reality, few engineering properties have true values because the value is often determined by the test.

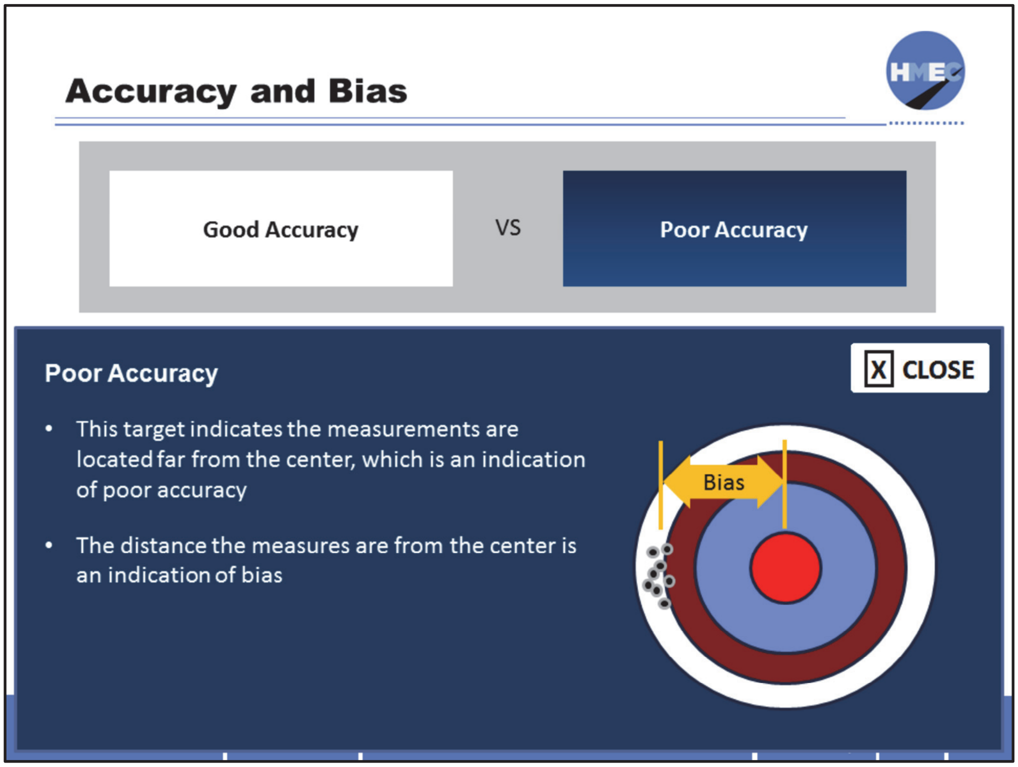
Take the measurement of asphalt content as an example. Variability is introduced by the test method; did we use an extraction method, an ignition furnace, or some other procedure? Each has a different variability associated with it so we never have a measure of the true value. We often use an established procedure, such as density determined by water displacement as the “true” measure and correlate other test results, such as those provided by a nuclear device to it, to determine a correction factor. However, we should keep in mind this is an engineering convenience not a measure of bias according strictly to the definitions.

Select each example for more information.



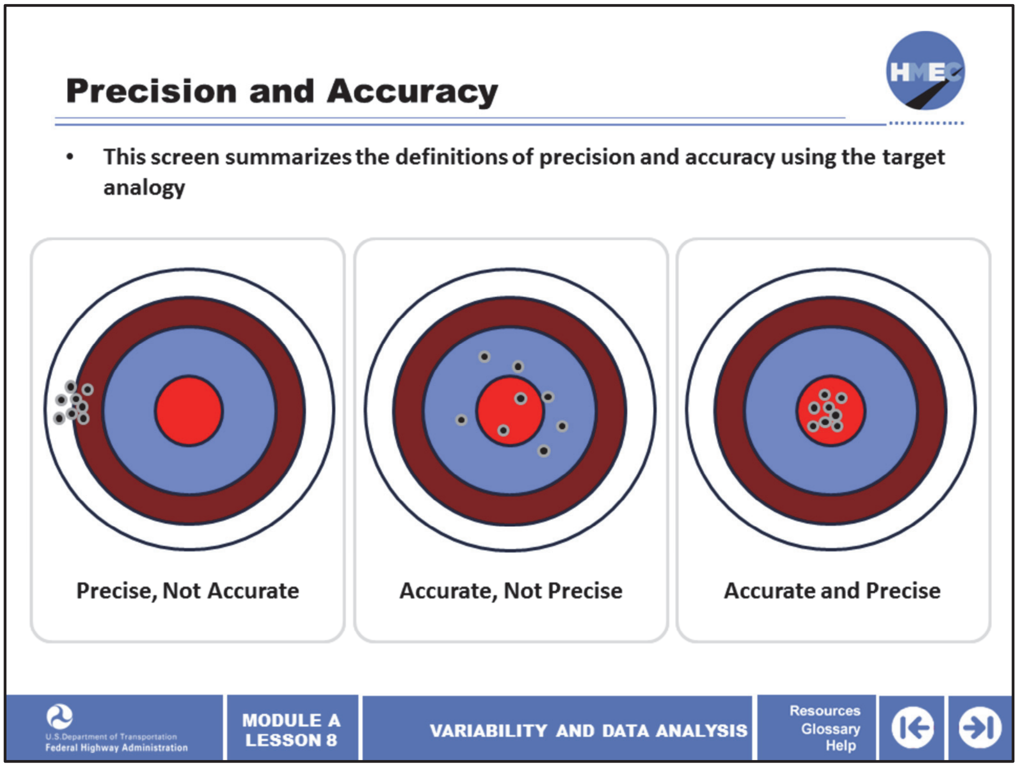
This target shows the measurements located at the center of the target indicating good accuracy.

Image description: Target.



This target indicates the measurements are located far from the center, which is an indication of poor accuracy. The distance the measures are from the center is an indication of bias.

Image description: Target highlighting bias with a two ended arrow.

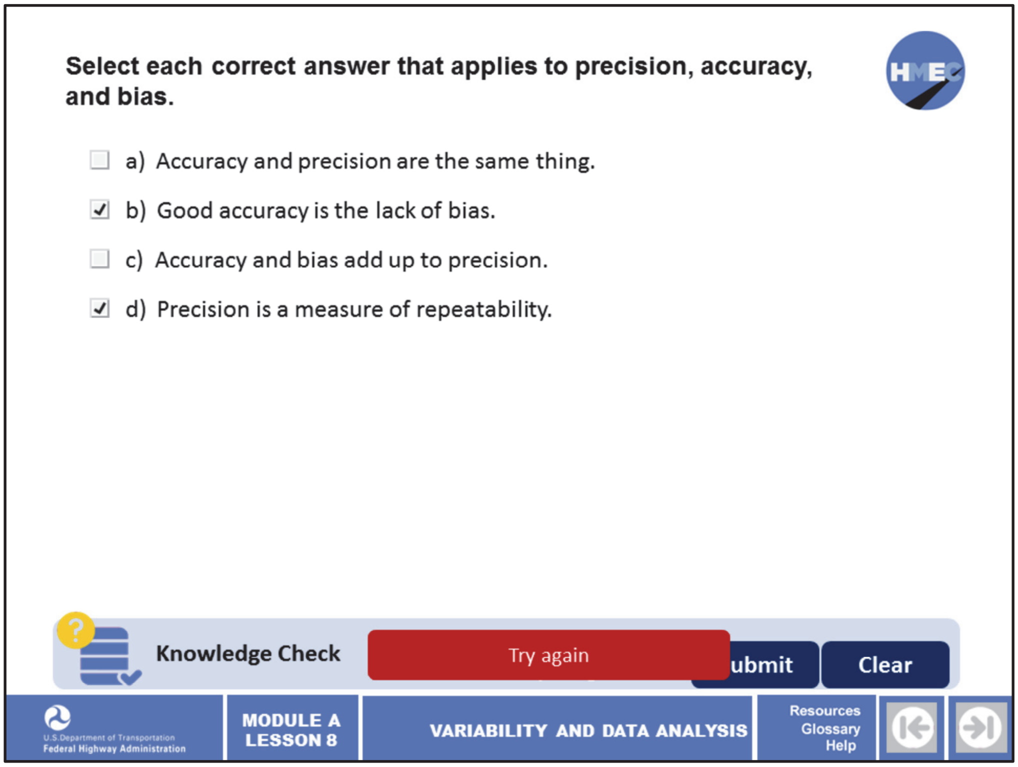


This screen summarizes the definitions of precision and accuracy using the target analogy. These targets illustrate that it is possible to be precise without being accurate (the target on the left), or accurate without being precise (the target in the middle). Ideally, we would like for a production or measurement process to be both accurate and precise (the target on the right).

Image description: Target.

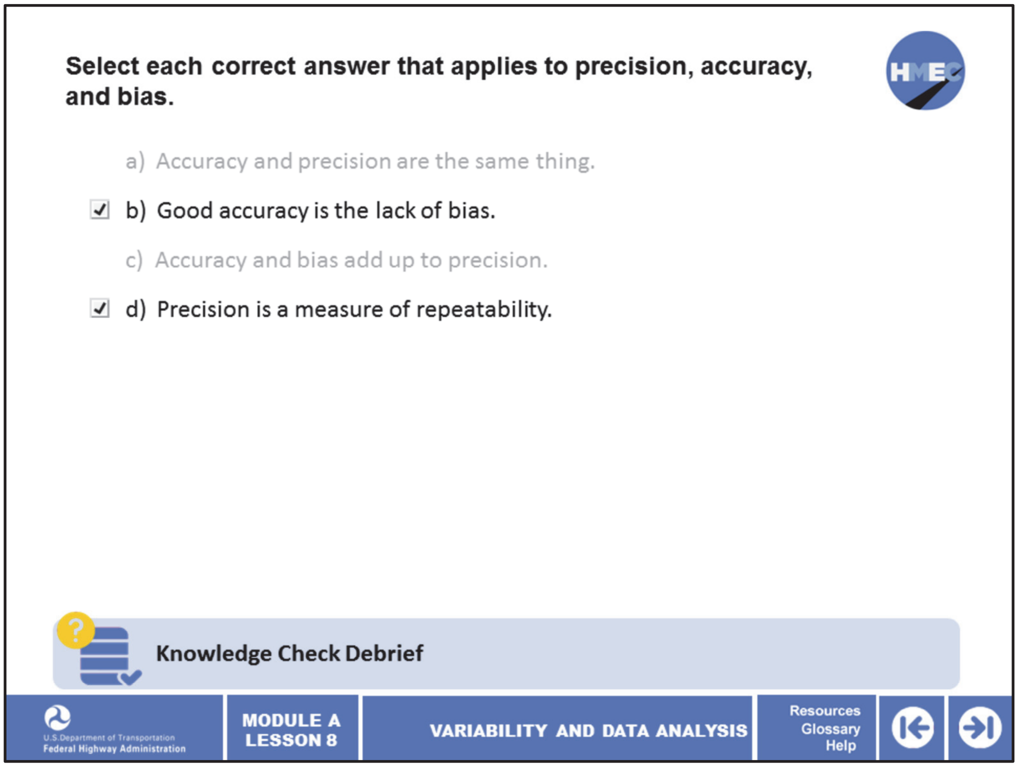
Image description: Target.

Image description: Target.

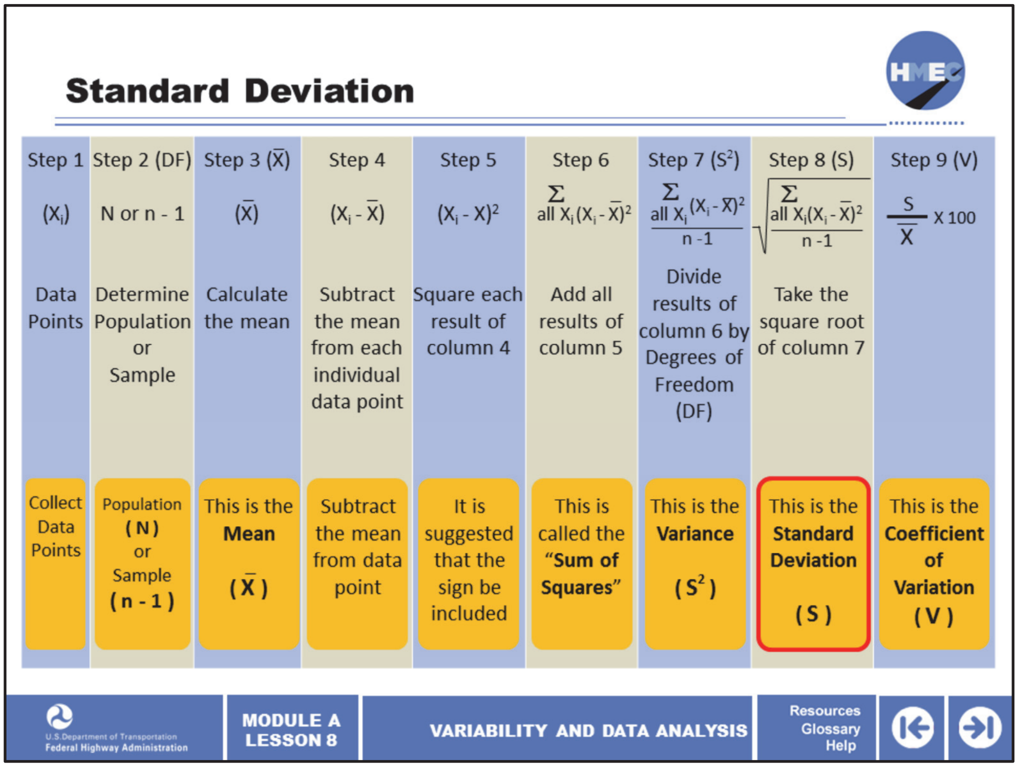


Select each correct answer that applies to precision, accuracy, and bias.

a) Accuracy and precision are the same thing;
b) Good accuracy is the lack of bias;
c) Accuracy and bias add up to precision; and/or
d) Precision is a measure of repeatability.



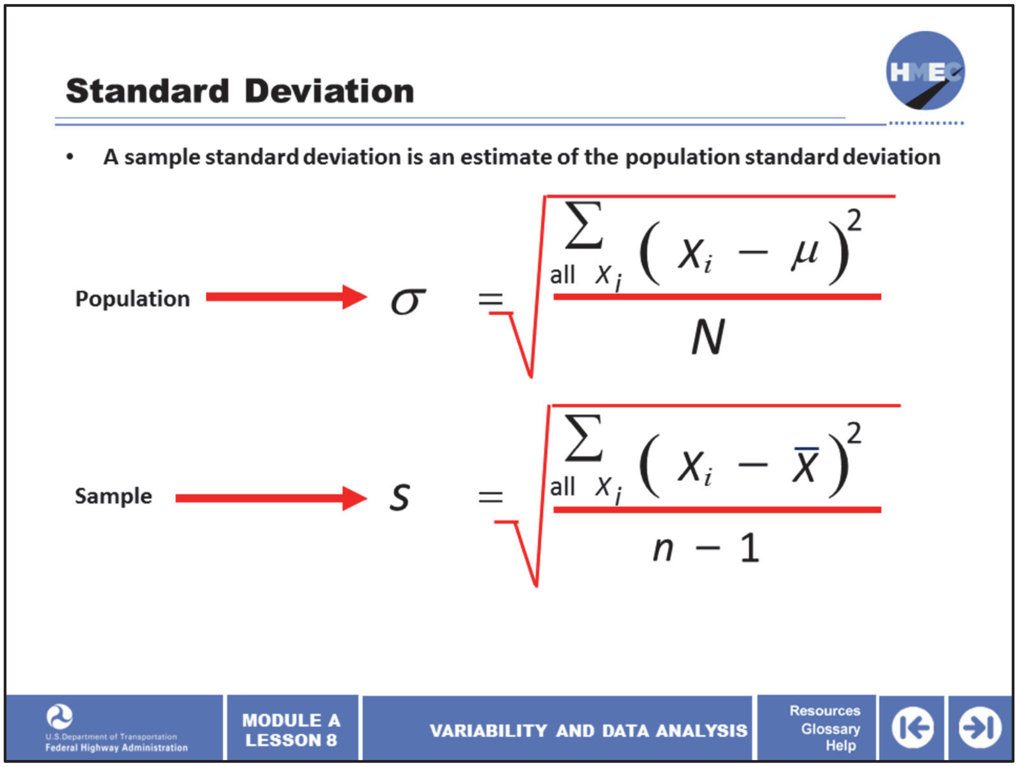
The correct answers are b) Good accuracy is the lack of bias and d) Precision is a measure of repeatability.



This screen highlights the equations for the sample standard deviation that were introduced in an earlier lesson. Standard deviation measures the dispersion of a set of numbers from the average or mean—in other words, their variability.

It is essential to understand that all highway products are inherently variable. Once this is understood, it is important to understand that there are various sources of variability and that variability should be considered an important parameter in the quality of the product.

The source, or component, of variability is a straightforward mathematical relationship based on the Pythagorean theorem. The previously discussed variability, *σ* and *s*, is termed overall variability.



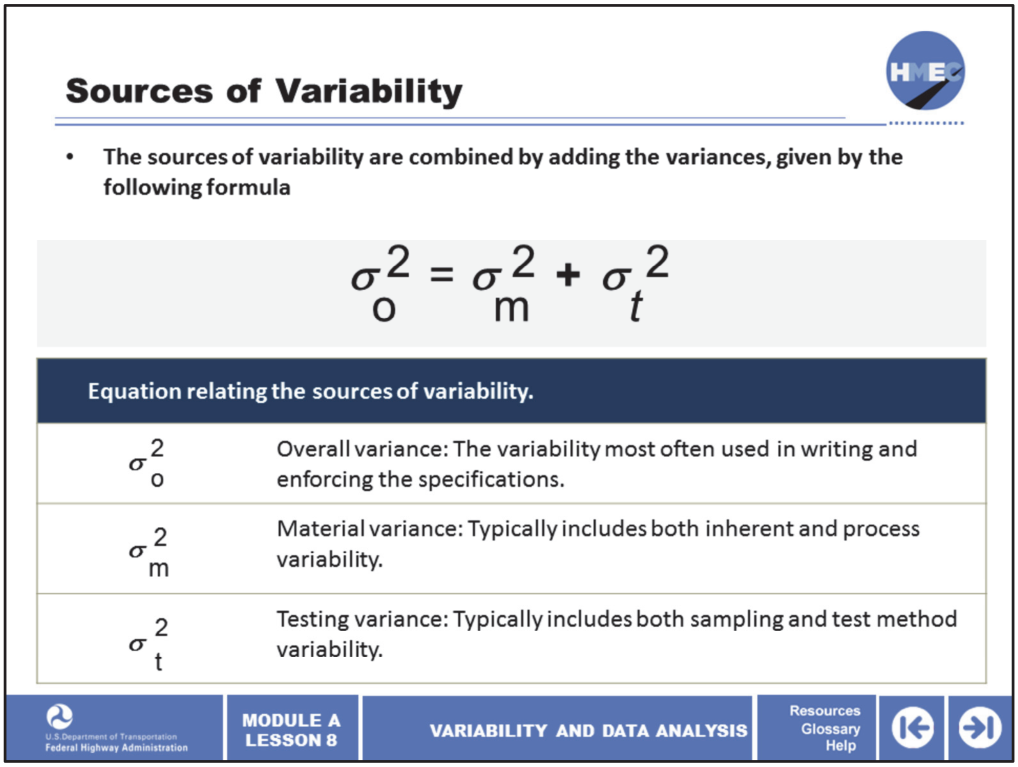
As discussed earlier, *s* is used to designate the standard deviation for a sample, and uses the degrees of freedom *n* - 1 in the denominator under the radical (square root sign).

A sample standard deviation is an estimate of the population standard deviation. However, for the standard deviation of a population, *σ* is used to indicate that the variability is known from a great deal of data. In the equations on the screen, "*n*" is used to denote sample size, while "*N*" represents the total number in the population.

It is obvious that, as the sample size increases, *s* approaches *σ* because of the smaller percentage difference between *n* - 1 and *n*. As a rule of thumb, when *n* is 30 or larger, *s* is approximately equal to *σ* and *n* can be used to calculate standard deviation. However, if a calculator or computer is used, it is preferred to use the sample key.

Image description: Population equation.

Image description: Sample equation.



This screen discusses the sources of variability, meaning the variability can come from many different sources; statisticians call them “errors”—sampling error or testing error, for example. These terms mean sampling variability and testing variability, not mistakes. Sources of variability are combined by the use of the basic measure of variability, the variance, *σ* squared.

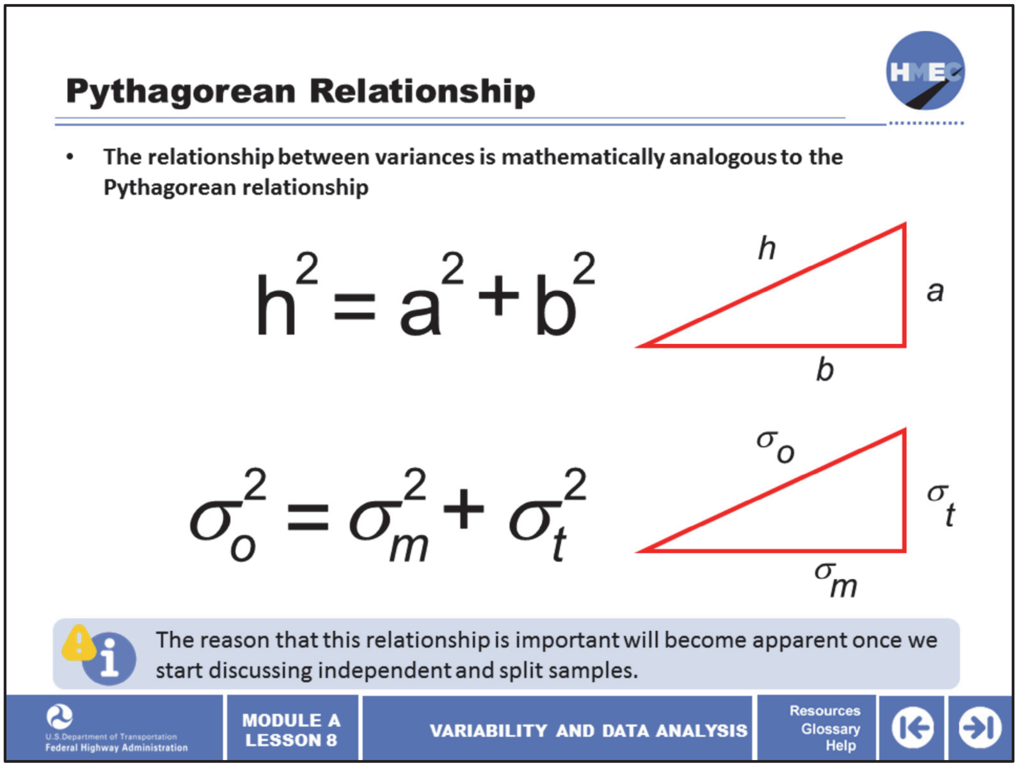
The sources of variability are combined by adding the variances given by the formula on the screen:

Sigma sub o squared equals sigma sub m squared plus sigma sub t squared. Where:

• Sigma sub o squared = overall variance, which is the variability most often used in writing and enforcing the specifications;
• Sigma sub m squared = material variance, which typically includes both inherent and process variability; and
• Sigma sub t squared = testing variance, which typically includes both sampling and test method variability.

As an example of these sources of variability, take aggregate production. When the aggregate is quarried, there is inherent variability in the rock in-place. When the aggregate is crushed, another source of variability is introduced. This also occurs when it is stockpiled, when it is hauled to the job site, and when it is placed. Lastly, when it is sampled and tested, two more sources of variability are introduced. Most likely, the only standard deviation identified will be the overall variability, σ_o, because of the time-consuming process that is required to identify the internal sources of variability.

Image description: Equation.



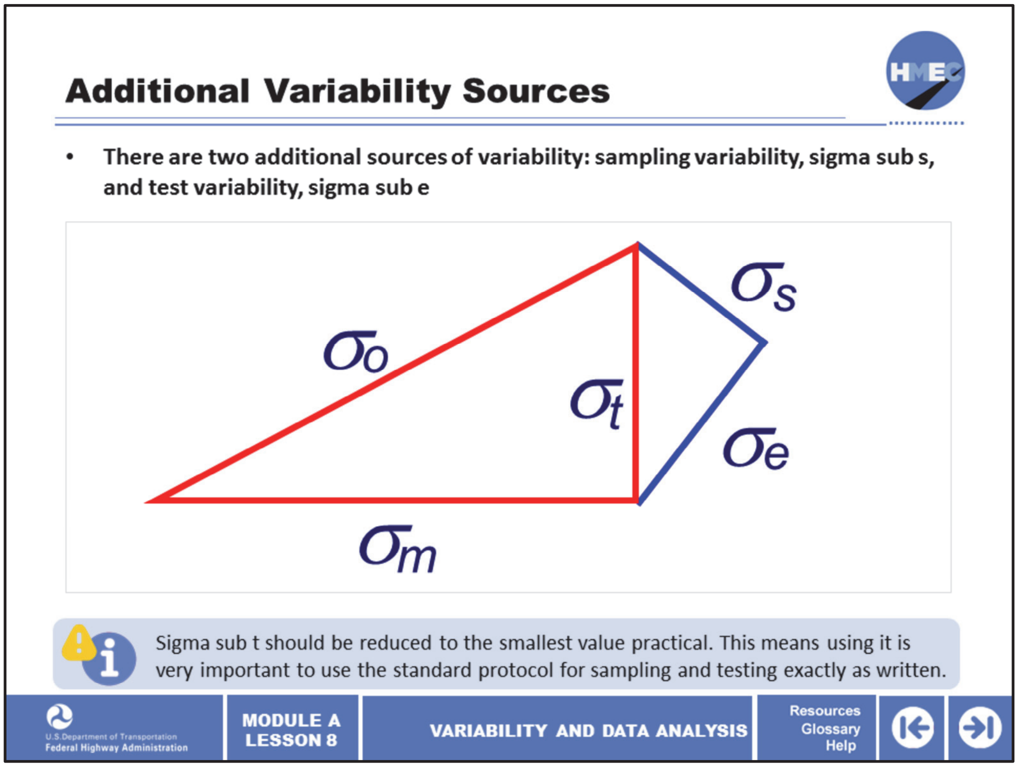
The relationship between variances is mathematically analogous to the Pythagorean relationship, which says that in a right triangle, the sum of the sides squared , a^2 plus b^2 , equals the hypotenuse squared, h^2.

Applying this equation to our aggregate analogy, one side of the triangle, sigma sub m squared, would represent the variability of our quarried, crushed, stockpiled, hauled, and placed aggregate, and the other side would represent variability of the sampling and testing, sigma sub t squared. These two sides squared equal the overall variance, sigma sub o squared.

The reason that this relationship is important will become apparent once we start discussing independent and split samples.

Image description: Equation and right triangle.

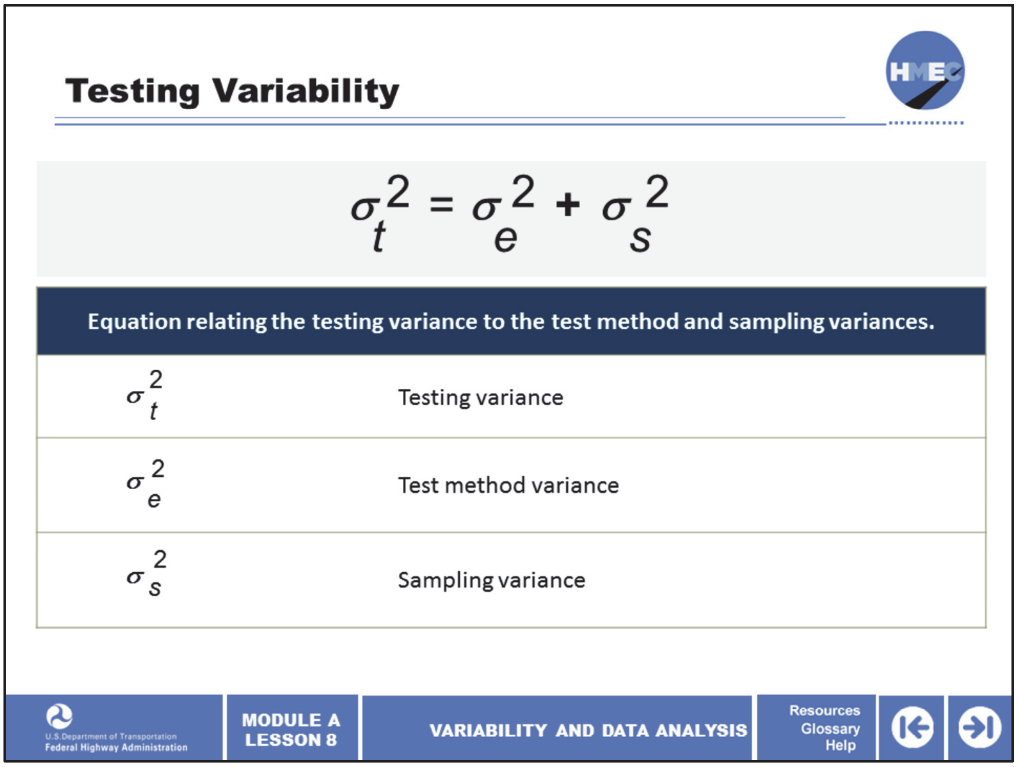
Image description: Equation and right triangle.



There are two additional sources of variability: sampling variability, sigma sub s, and test variability, sigma sub e. These make up sigma sub t in the previous screen; it is obvious that sigma sub t becomes the hypotenuse of the blue triangle with sigma sub s and sigma sub e being the legs. Since sigma sub m can't be found directly, it must be backed into, so to speak.

Although it is a laborious process requiring a lot of testing, sigma sub o and sigma sub t can be measured. From these two sources of variability, sigma sub m is determined. Since determining sigma sub o is relatively easy to determine, sigma sub t is mathematically subtracted, using variances from sigma sub o to determine sigma sub m. The contractor has primary control over sigma sub m, or at least the processing portion of it. But sigma sub t is necessary to obtain sigma sub o. Therefore, sigma sub t should be reduced to the smallest value practical. This means that it is very important to use the standard protocol for sampling and testing exactly as written.

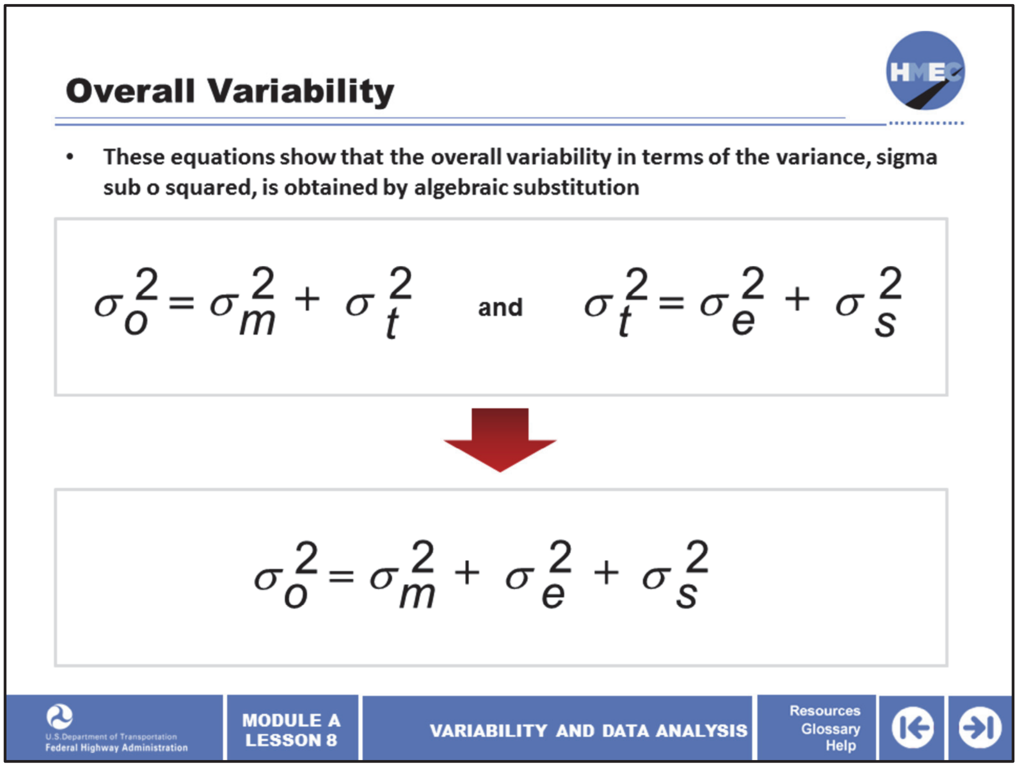
Image description: Right triangle.



Looking at the last screen in more detail provides the equation relating the testing variance to the test method and sampling variances. The additional legs of sigma sub e squared and sigma sub s squared are similarly related to sigma sub t squared, where:

- Sigma sub t squared = testing variance;
- Sigma sub e squared = test method variance; and
- Sigma sub s squared = sampling variance.

Image description: Equation.

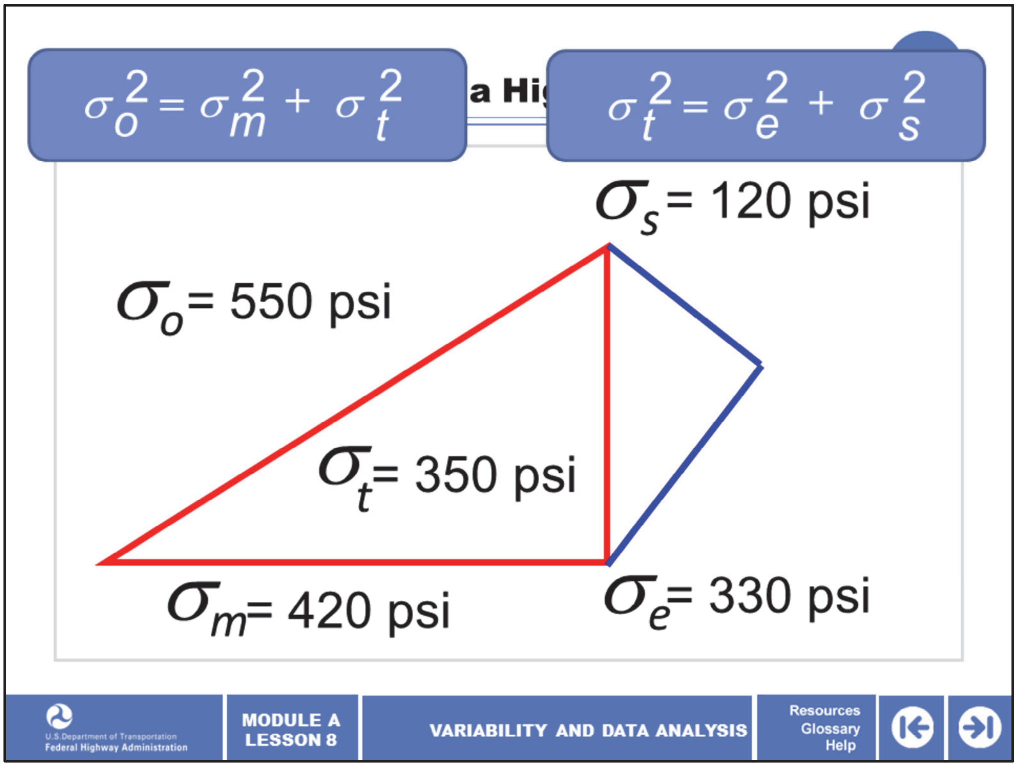


These equations show that the overall variability in terms of the variance, sigma sub o squared, is obtained by algebraic substitution. The relationship that relates the overall variability to material, test method, and sampling variability is given by the equation on the screen: Sigma sub o squared equals sigma sub m squared plus sigma sub t squared and sigma sub t squared equals sigma sub e squared plus sigma sub s squared; so by substitution, sigma sub o squared equals sigma sub m squared plus sigma sub e squared plus sigma sub s squared.

This equation shows that the variance of sampling and testing directly affects the overall variance. So any deviation from the sampling and testing protocol will affect the overall variability. If the overall variability is larger than what is expected, often the first thing that comes to mind is that the process has gotten more variable. But it may be equally likely that the sampling and testing is the reason for the increase.

Image description: Equation.

Image description: Equation.



It is often easier to understand these equations if we use numbers. This screen shows a typical source of variability relationship for Portland cement concrete (PCC). This is an example from a West Virginia Highway Department (WVHD) study of concrete several years ago to determine the sources of variability in their concrete testing. They found the relationship shown:

• Sigma sub e equals 330 psi; and
• Sigma sub s equals 120 psi.

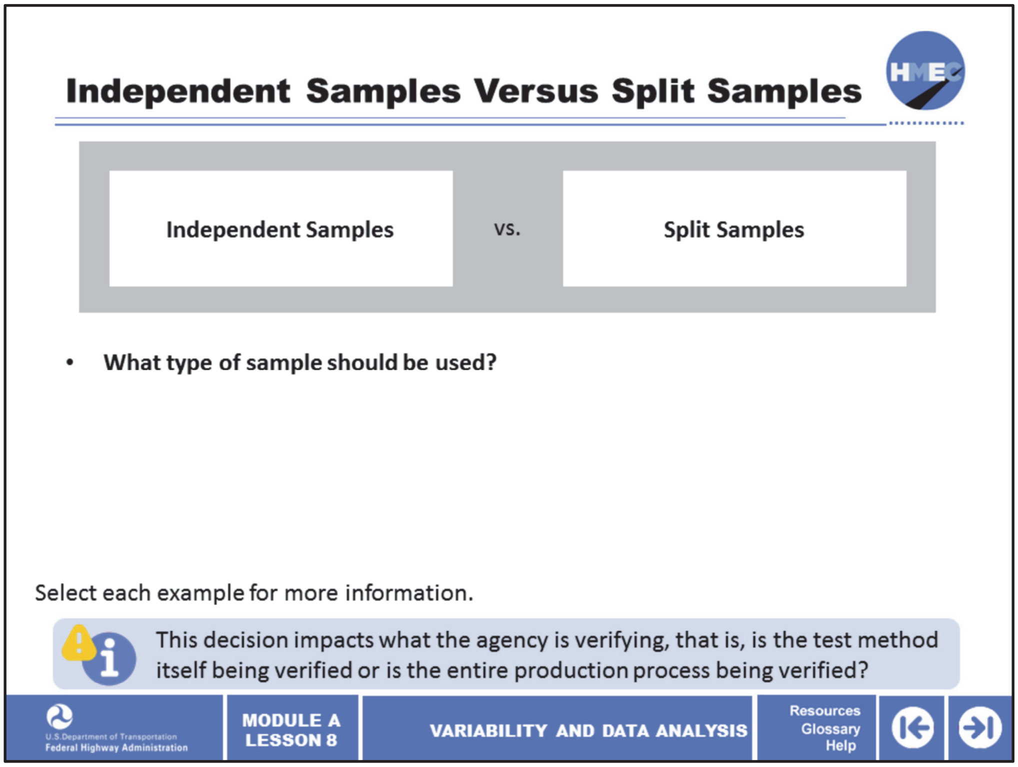
Using the equation for variances, sigma sub t squared equals sigma sub e squared plus sigma sub s squared. This makes sigma sub t squared equal to (330) squared plus (120) squared, which equals (108,900) plus (14,400). This equals 123,300 psi. Taking the square root of this produces sigma sub t equal to 351 psi; WVHD rounded it to 350 psi.

Then sigma sub o squared equals sigma sub m squared plus sigma sub t squared. Substituting values in the equation, sigma sub o squared equals (420) squared plus (350) squared, which equals 176,400 plus 122,500. This equals 299,700. Taking the square root of this determines the sigma sub o to be 547 psi, rounded to 550 psi.

Image description: Equation.

Image description: Equation.

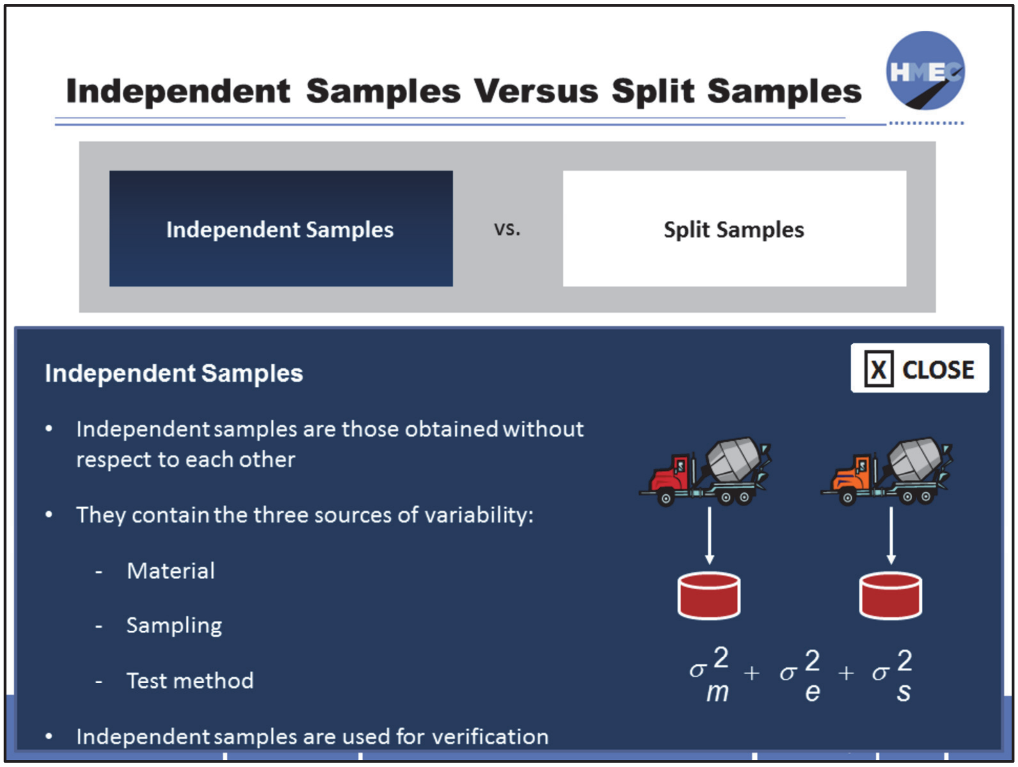
Image description: Right triangle.



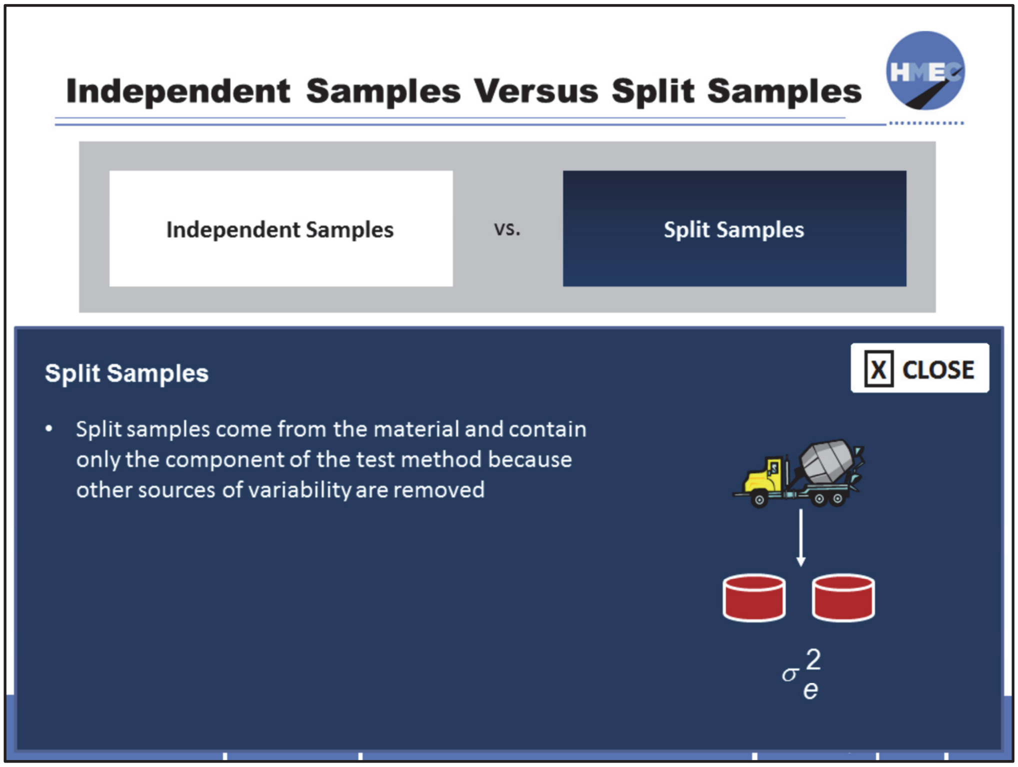
Understanding the sources of variability is very important. For instance, for an agency to verify or validate a contractor's test results, as will be discussed in the lesson on verification and materials testing dispute resolution, the decision must be made whether to use split or independent samples. This decision impacts what the agency is verifying, that is, is the test method itself being verified or is the entire production process being verified?

So, what type of sample should be used: independent samples or split samples?

Select each example for more information.

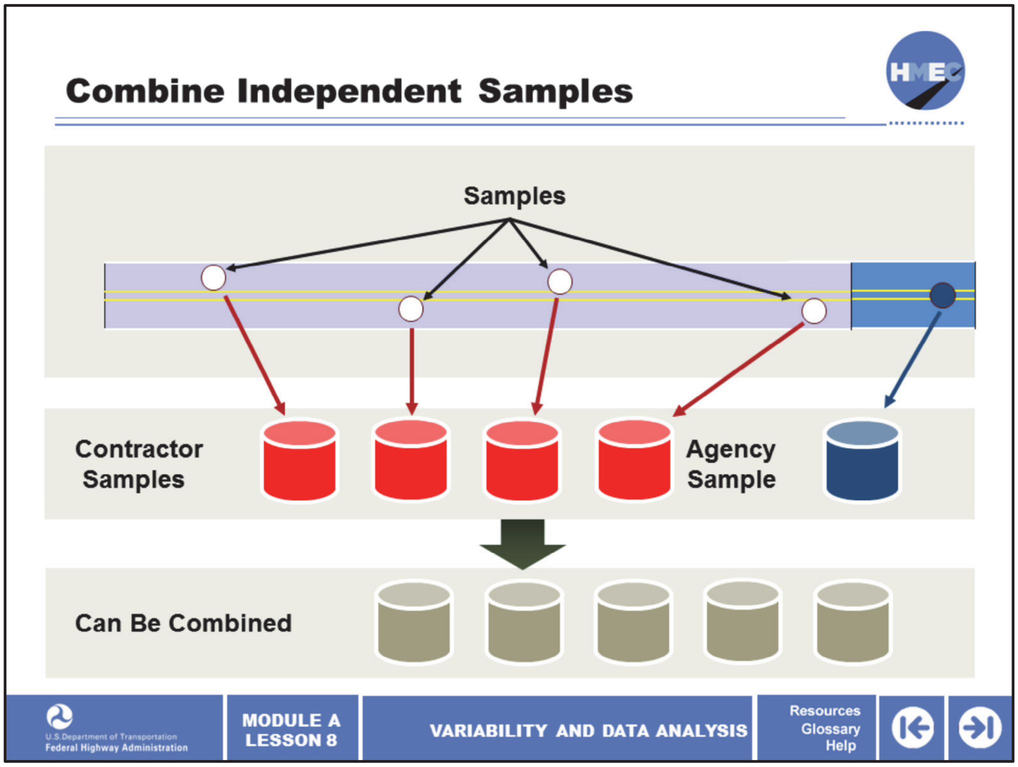


Independent samples are those obtained without respect to each other. They contain the three sources of variability: material, sampling, and test method.



Split samples come from the material and contain only the component of the test method because other sources of variability are removed.

Image description:



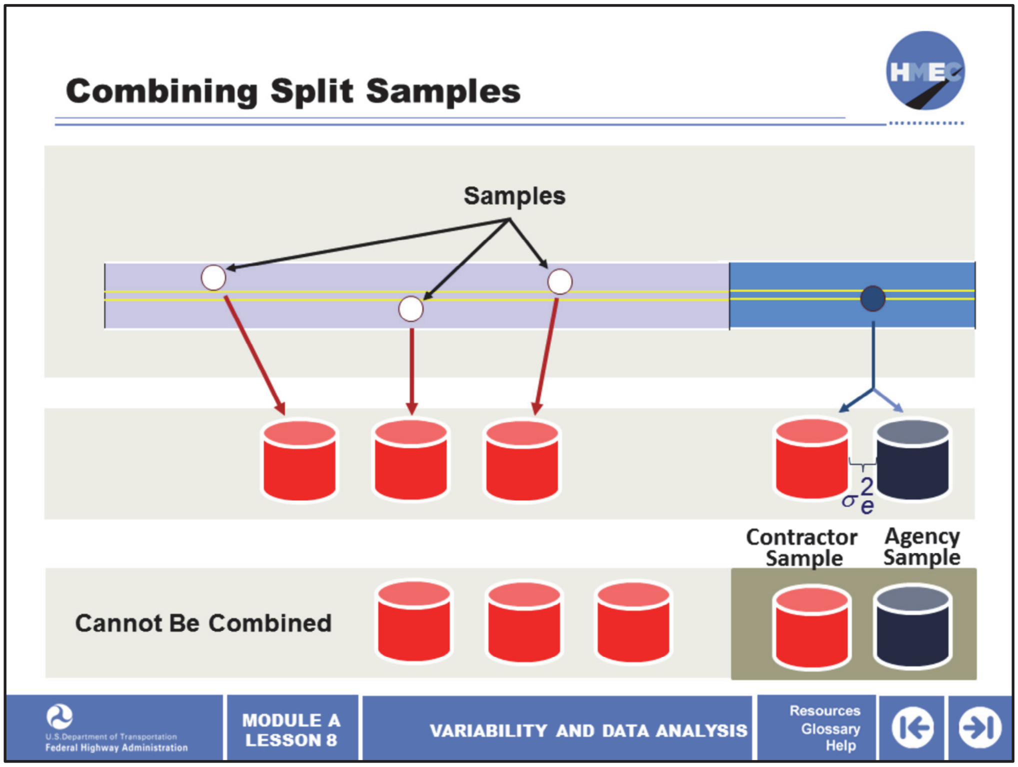
Statistically, can we combine independent samples? If independent samples are used for the agency/contractor test result comparison, then the agency tests are subject to the same overall variation as the contractor tests. Since, in this case, the agency test results are independent data from the contractor's results, the two data sets can be combined to make acceptance decisions from a statistical standpoint.

Although statistical theory allows the results to be combined, in reality, it is rarely a good idea. This is because usually two populations will result from the different sources of the test results. Although the same sources of variability are included in each set of test results, the values of these sources are likely to be different. This is likely to result in two populations, which often results in a bimodal distribution with an increased measure of variability. Some agencies have learned this the hard way.

Image description: Display of random samples from a lot.

Image description: Four contractor core samples and one agency core sample from the lot.

Image description: Five combined core samples.

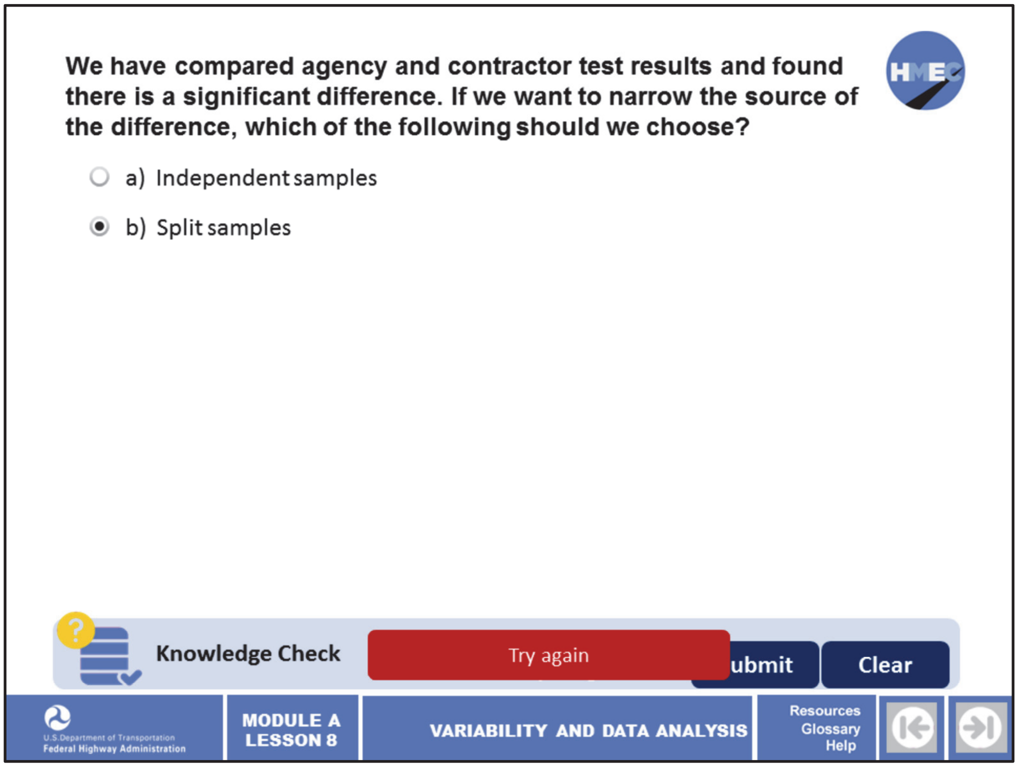


Can split samples be combined? If split samples, on the right of the screen, are used for agency/contractor test result comparison, then the comparisons only consider the agency's and contractor's test procedures, and do not consider the variability associated with material, production, or the sampling process and therefore they cannot be combined. In theory, the agency sample could be combined with the contractor's independent samples, but in practicality, it's not a good idea.

Image description: Display of random samples from a lot.

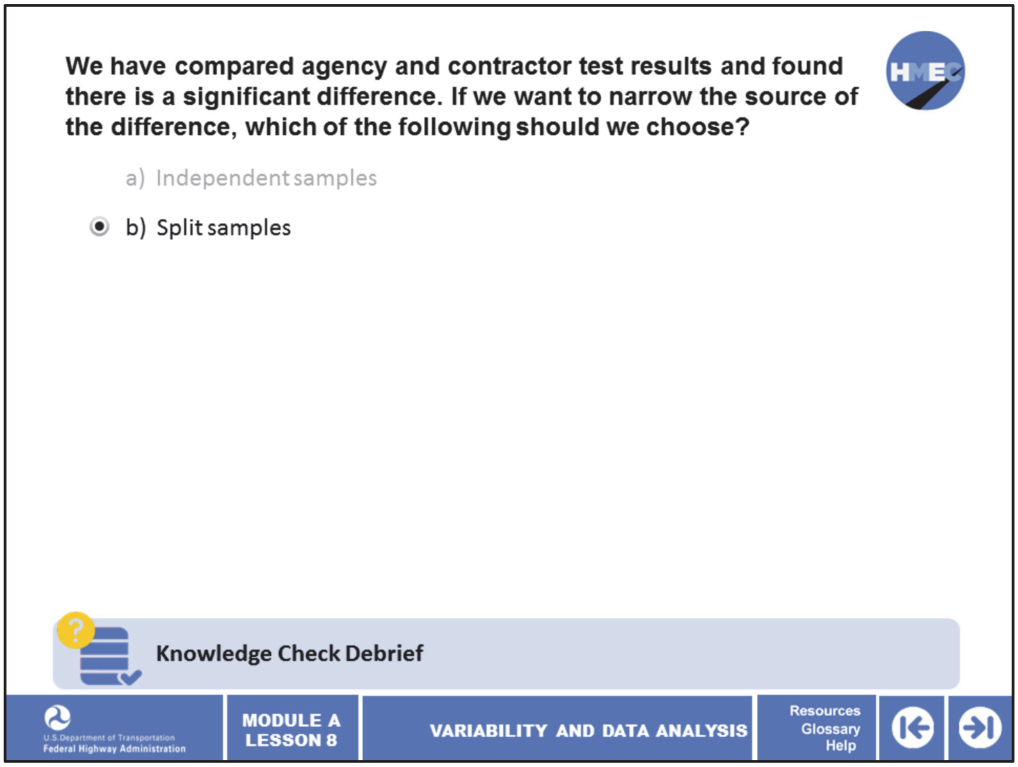
Image description: Five core samples, separating one contractor sample and one agency sample.

Image description: Five core samples that cannot be combined, separating one contractor sample and one agency sample.

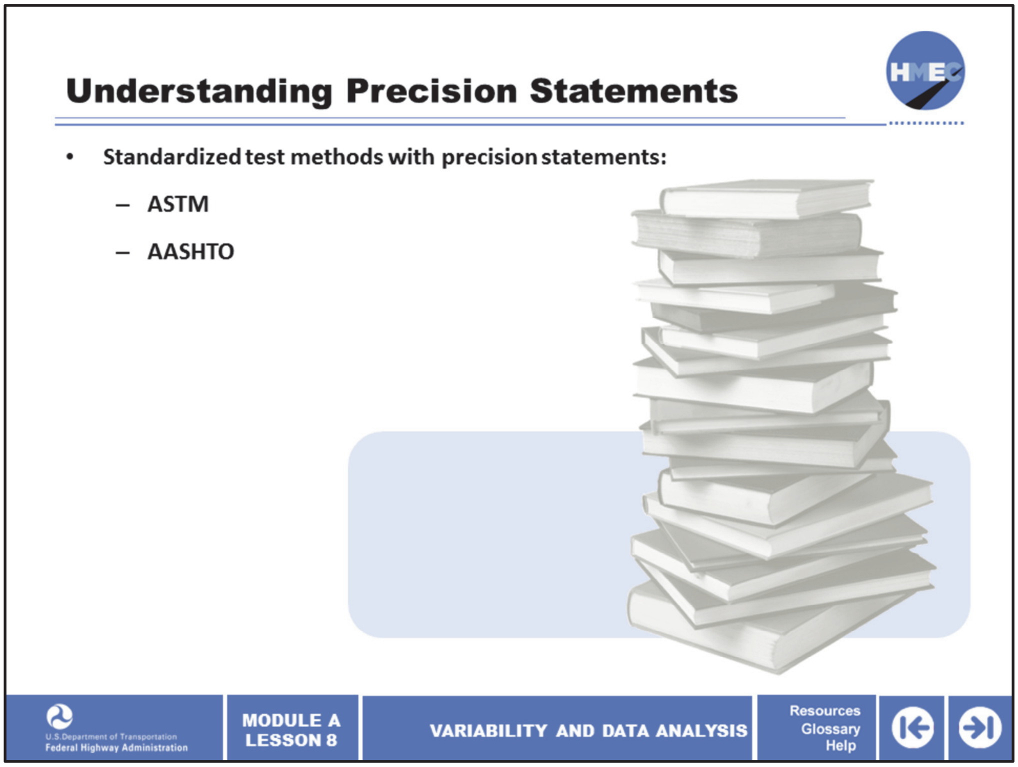


We have compared agency and contractor test results and found there is a significant difference. If we want to narrow the source of the difference, which of the following should we choose?

a) Independent samples; or
b) Split samples.

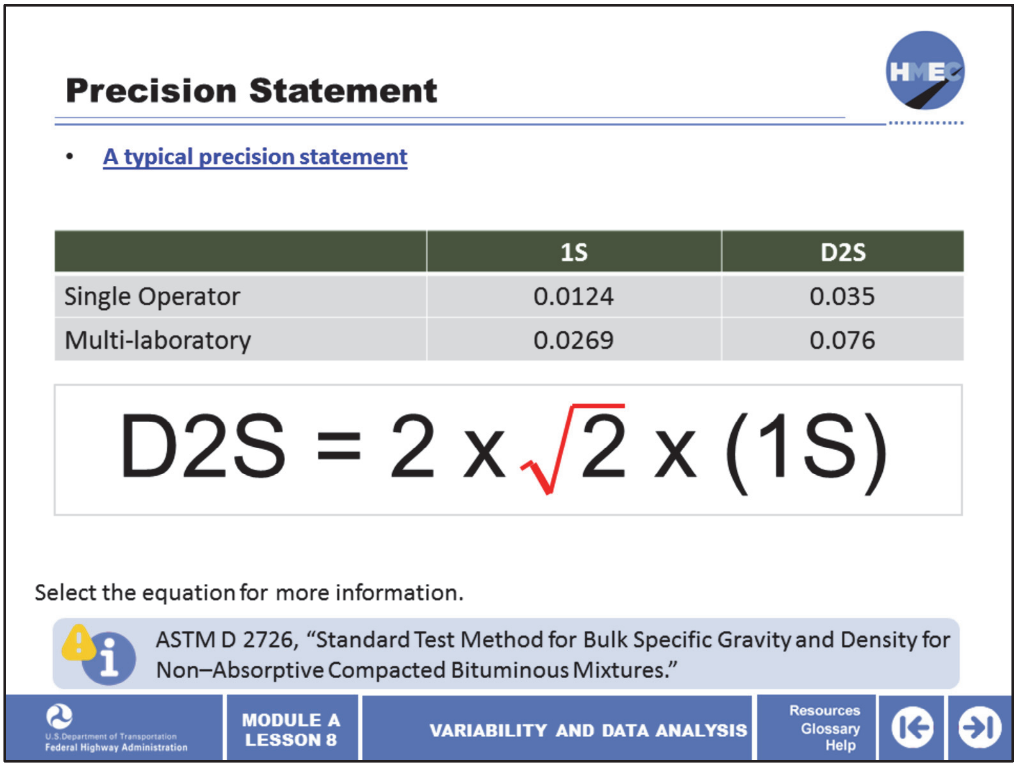


We should choose b) Split samples if we want to narrow the source of the difference between agency and contractor test results.



Many standardized test methods, such as those in ASTM and AASHTO, contain precision and bias statements. It is important to understand the meaning of the terms used in these statements. The precision statement provides a measure of how repeatable the test method is. This is very important when it comes to establishing the creditability of a test method. If these numbers tend to be small, the test method is considered to have good precision. The next screen provides an example.

Image description: Stack of books.



A typical precision statement will have a column with a one-sigma limit (1S) that is used to denote the estimate of the standard deviation of the population of the test method. This limit is usually established for a single operator in the second row, and also for multi–laboratory precision in the third row.

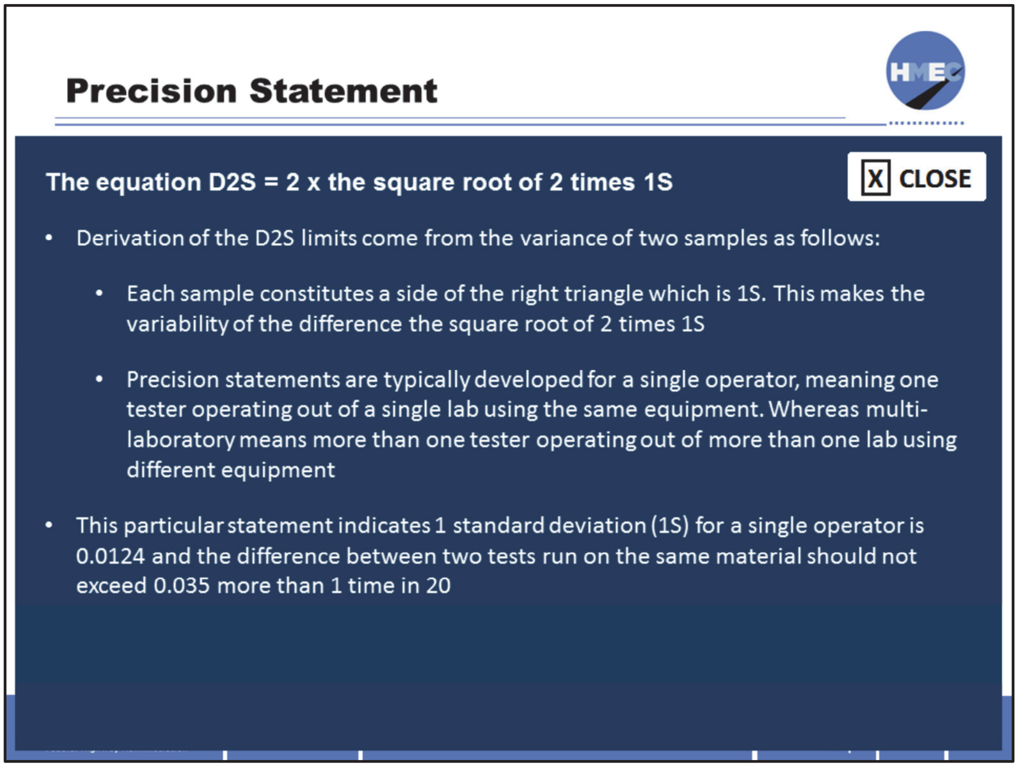
Additionally, it will have a column designated difference 2 sigma limit (D2S) that indicates the maximum difference between two portions of the same material, using split samples, that would be exceeded in the long run in only 1 case in 20.

The values shown on the screen are from ASTM D 2726, "Standard Test Method for Bulk Specific Gravity and Density for Non–Absorptive Compacted Bituminous Mixtures."

Select the equation for more information.

Reference: ASTM D 2726, "Standard Test Method for Bulk Specific Gravity and Density for Non–Absorptive Compacted Bituminous Mixtures."

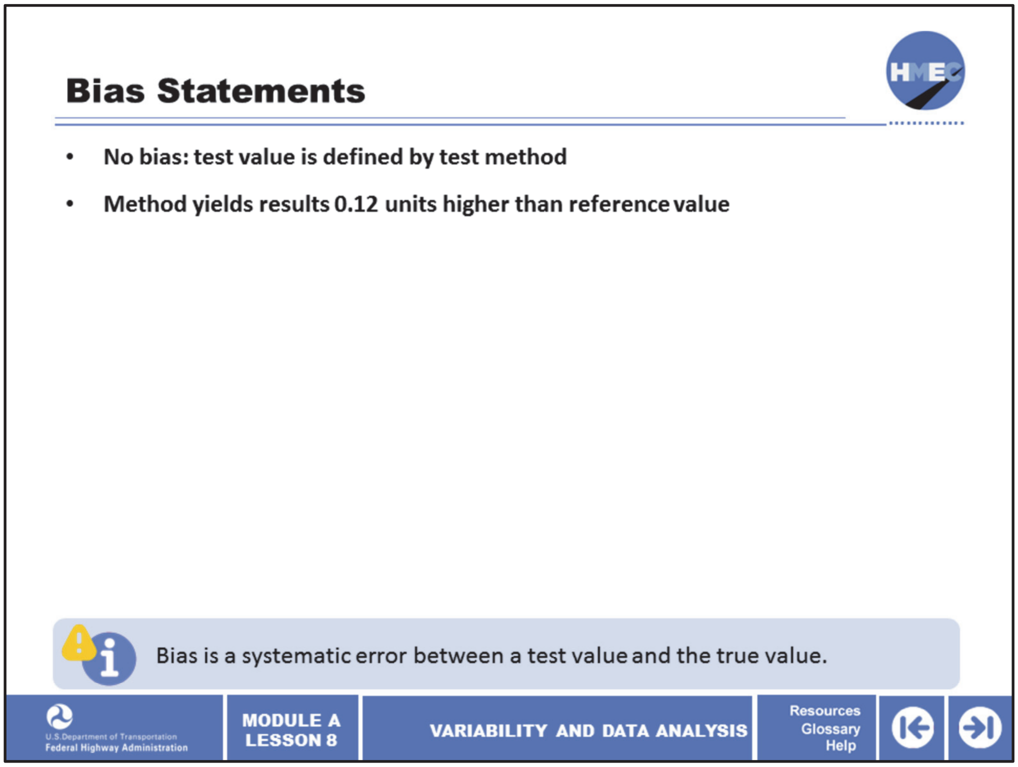
Image description: Equation.



The equation for determining D2S = 2 times the square root of 2 times 1S.

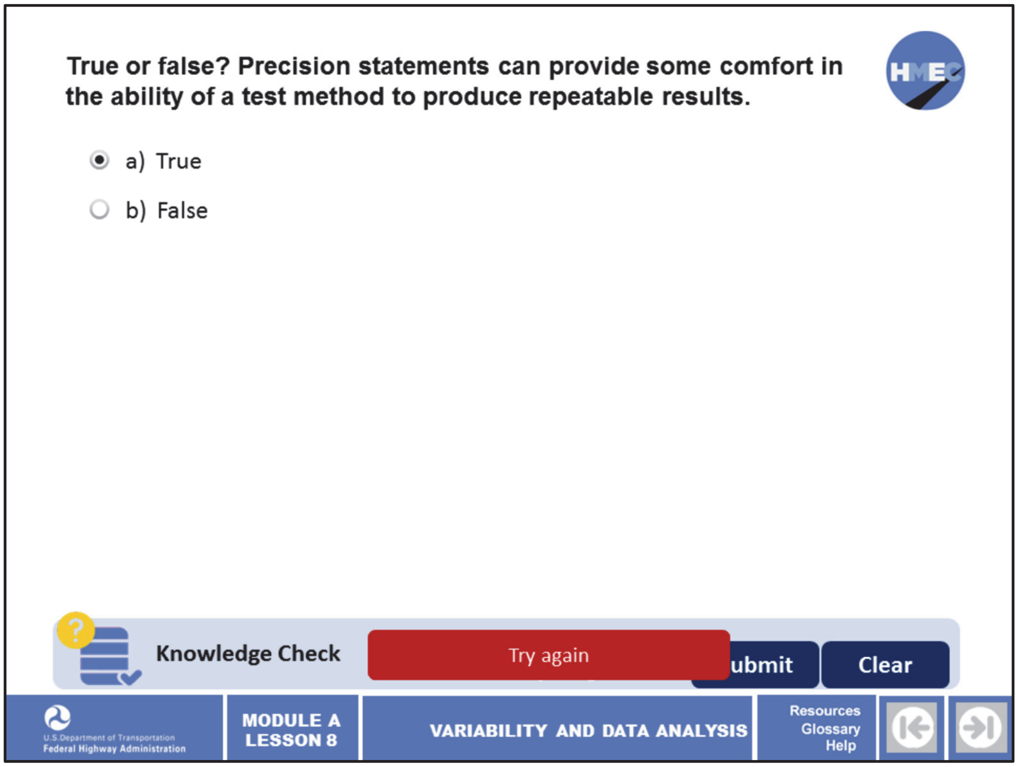
The derivation of the D2S limits comes from the variance of two samples as follows. Each sample constitutes a side of the right triangle, which is 1S. This makes the variability of the difference the square root of 2 times 1S. The D2S stands for difference 2 standard deviations. We know 2 standard deviations includes 95% of the population, which is equivalent to 19 times out of 20, so it would be exceeded only 1 time in 20. Additionally, precision statements are typically developed for a single operator, meaning one tester operating out of a single lab using the same equipment. Whereas multi-laboratory means more than one tester operating out of more than one lab using different equipment. It is obvious that the multi-lab value will be larger than that of a single operator because of the additional sources of variability.

This particular statement indicates 1 standard deviation (1S) for a single operator is 0.0124 and the difference between two tests run on the same material should not exceed 0.035 more than 1 time in 20. Likewise, if more than one lab is being used, 1 standard deviation (1S) for a multi-operators is 0.0269 and the difference between two tests run on the same material in more than one lab should not exceed 0.076 more than 1 time in 20.



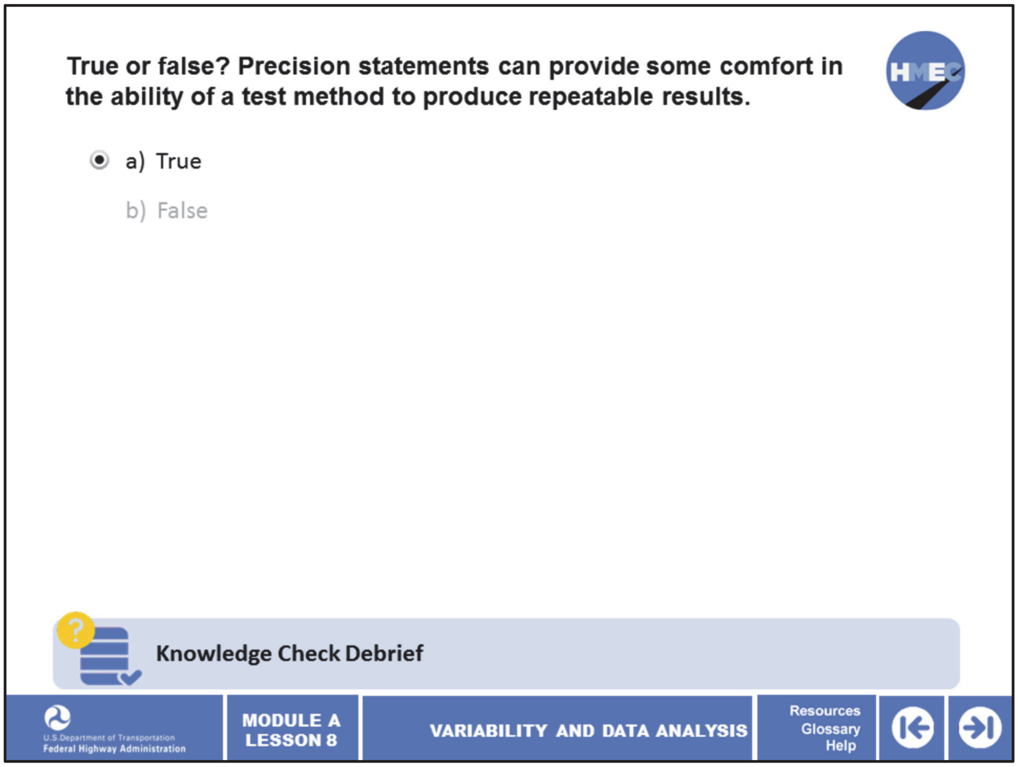
As stated previously, bias is a systematic error between a test value and the true value. Although bias statements are highly desirable, many test methods have no true value. Stated differently, the test values can only be established through the test method. Where bias is found, the method may be adjusted to eliminate the bias.

An example of a bias statement is in a chemical test where there is a reference value, and for instance would say, "the method will yield results 0.12 units higher than the reference value."

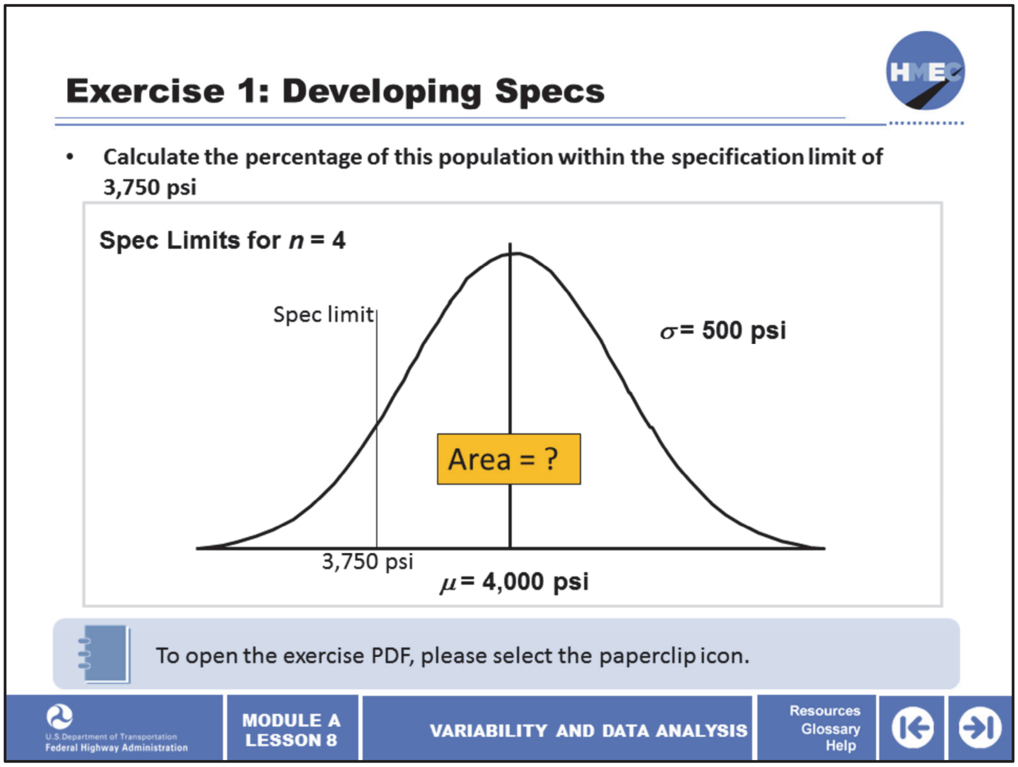


True or false? Precision statements can provide some comfort in the ability of a test method to produce repeatable results.

a) True; or
b) False.



The correct answer is a) True.

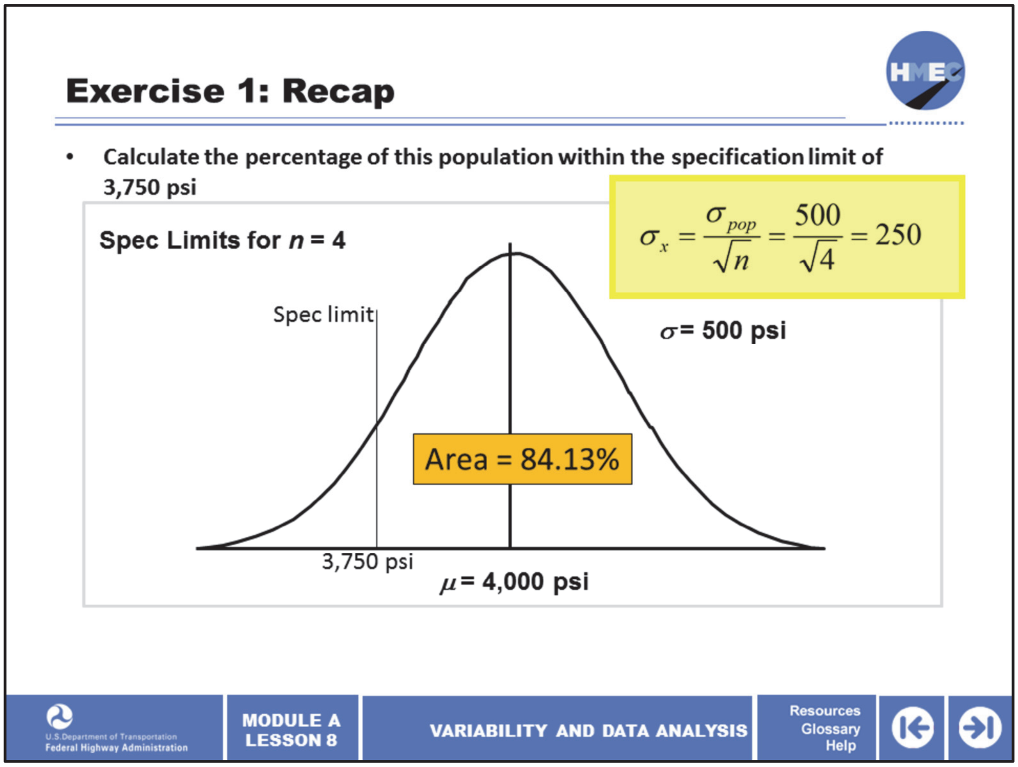


Now let's take a moment for an exercise. Use the exercises document that is provided from the paperclip icon. Please allow 10 minutes to conduct this exercise.

Here is an exercise for you to determine how much of a contractor's population meets a specification limit for concrete strength. The specification requires using the average of a sample size of 4. The contractor's history indicates that he produces an average strength of 4,000 psi with a standard deviation of 500 psi.

Calculate the percentage of this population within the specification limit of 3,750 psi.

Image description: A normal curve.



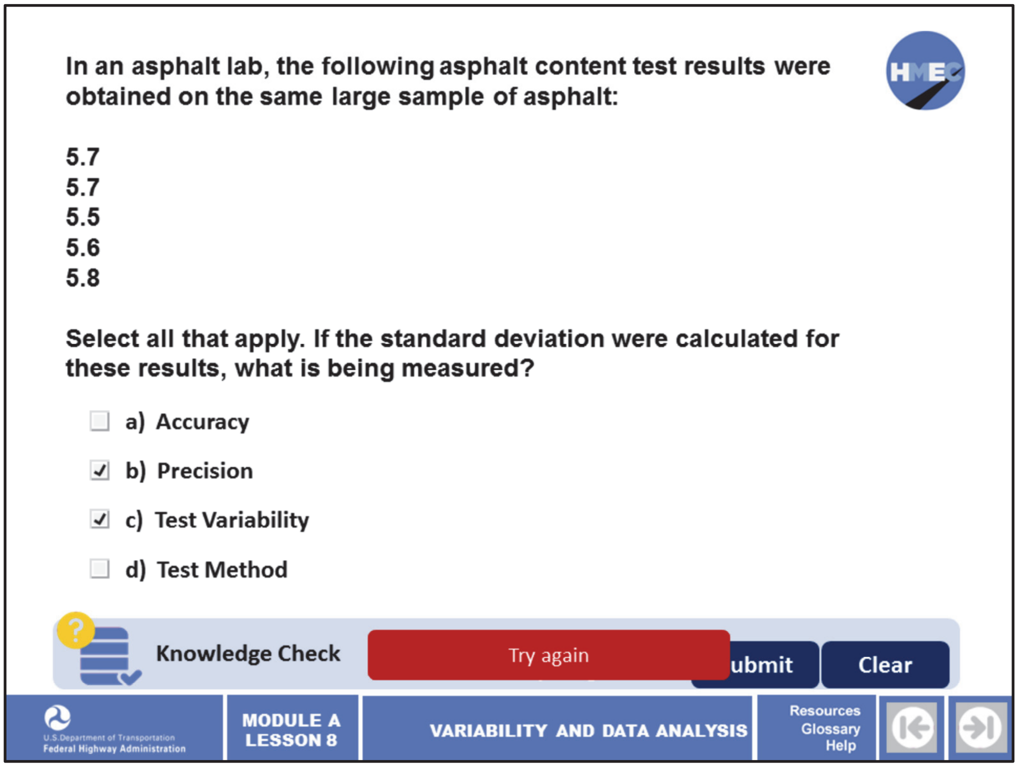
Since the spec is based on the average of a sample size of *n* = 4, we must use the distribution of means concept. The standard deviation of the means is sigma x-bar, which equals sigma divided by the square root of 4; thus, the population sigma, 500 divided by the square root of 4 equals 500 divided by 2 makes sigma x-bar equal to 250 psi.

Next we calculate the z value for the population in relation to the spec limit. Thus Z = (X-mu) /divided by sigma-sub X-bar. This solution is (3,750 – 4,000)/250 = -1.0.

Using the table of areas under the normal curve, we find Z = -1.0 is 0.34.13 or 34.13%. This is added to the 50% of the population above the average, so the percent of the population within the spec limit is 34.13 + 50.00 = 84.13%.

Image description: Equation.

Image description: Normal curve.

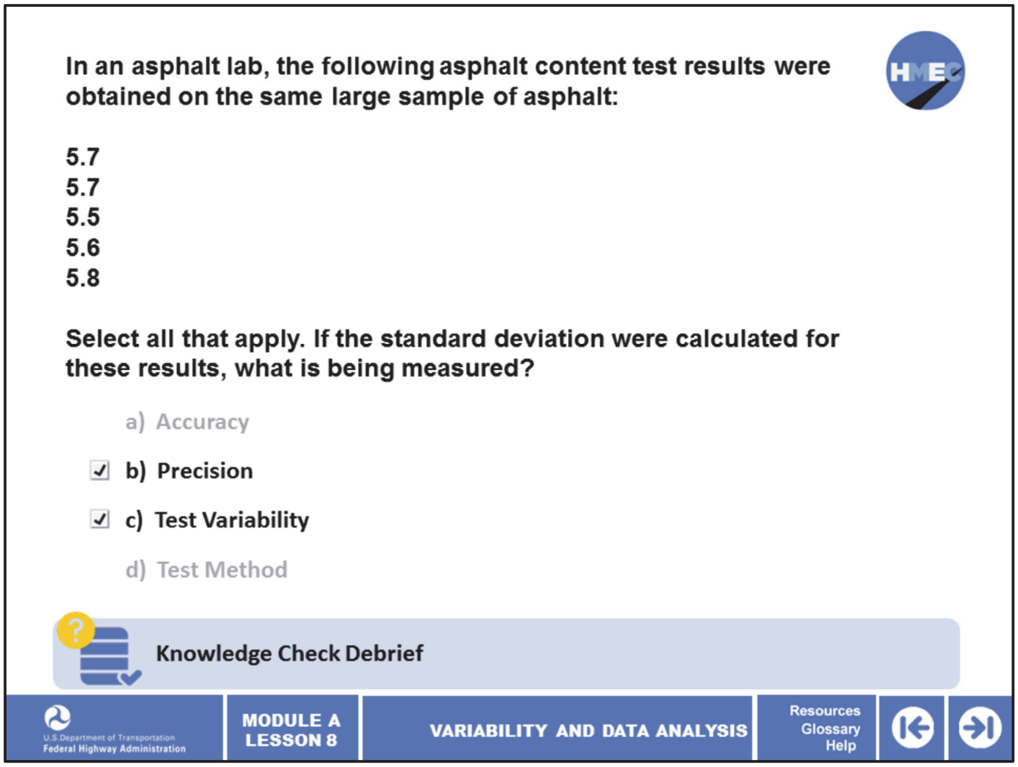


In an asphalt lab, the following asphalt content test results were obtained in the same large sample of asphalt:

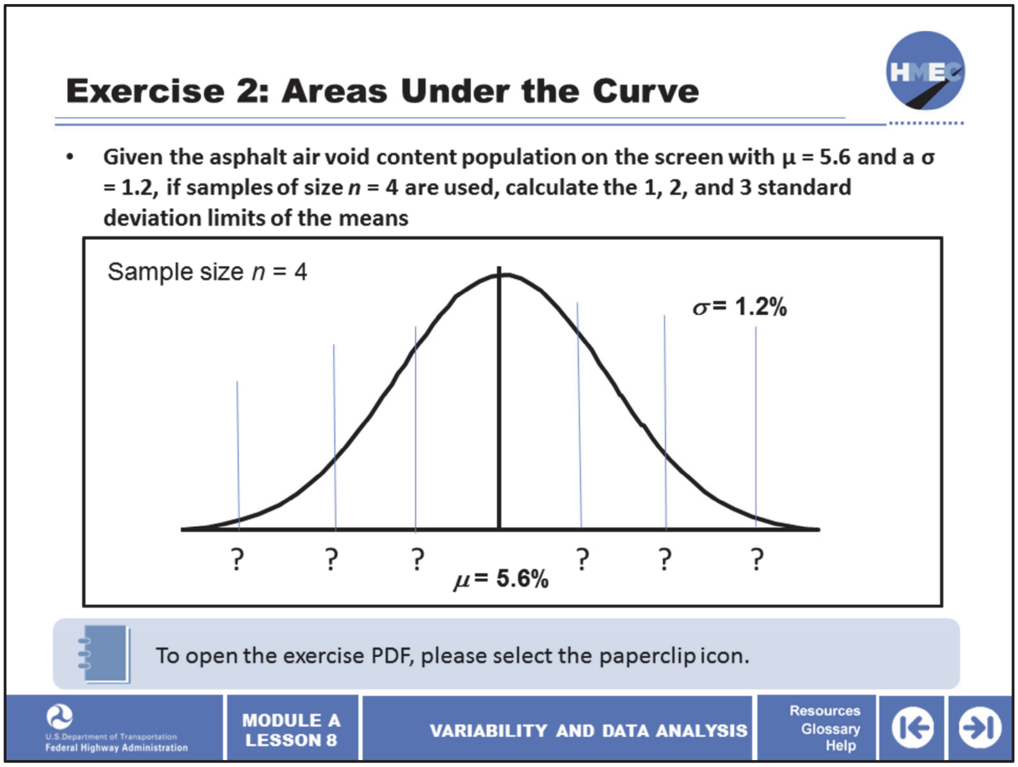
- 5.7;
- 5.7;
- 5.5;
- 5.6; and
- 5.8.

Select all that apply. If the standard deviation were calculated for these results, what is being measured?

a) Accuracy;
b) Precision;
c) Test Variability; and
d) Test Method.



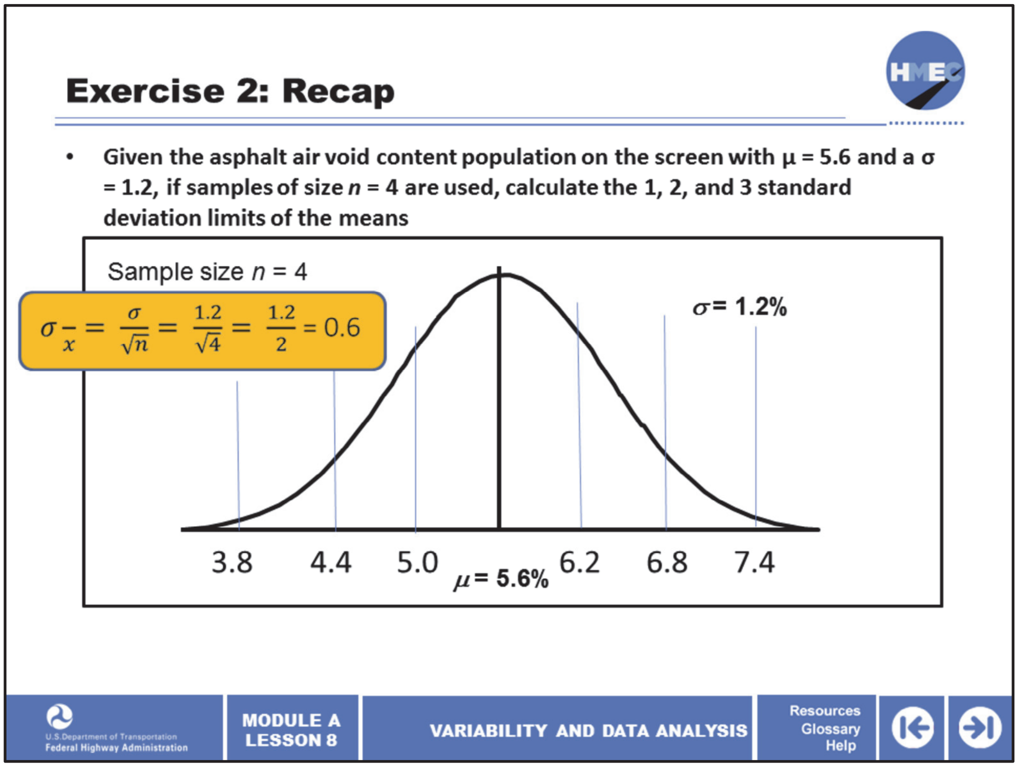
The correct answers are b) Precision; or c) Test Variability.



Now let's take a moment for another exercise. Use the exercises document that is provided from the paperclip icon. Please allow 10 minutes to conduct this exercise.

Given the asphalt air void content population on the screen with $\mu = 5.6$ and a $\sigma = 1.2$, if samples of size n = 4 are used, calculate the 1, 2, and 3 standard deviation limits of the means.

Image description: Normal curve.



Since the question is based on the average of a sample size of *n* = 4, we must use the distribution of means concept.

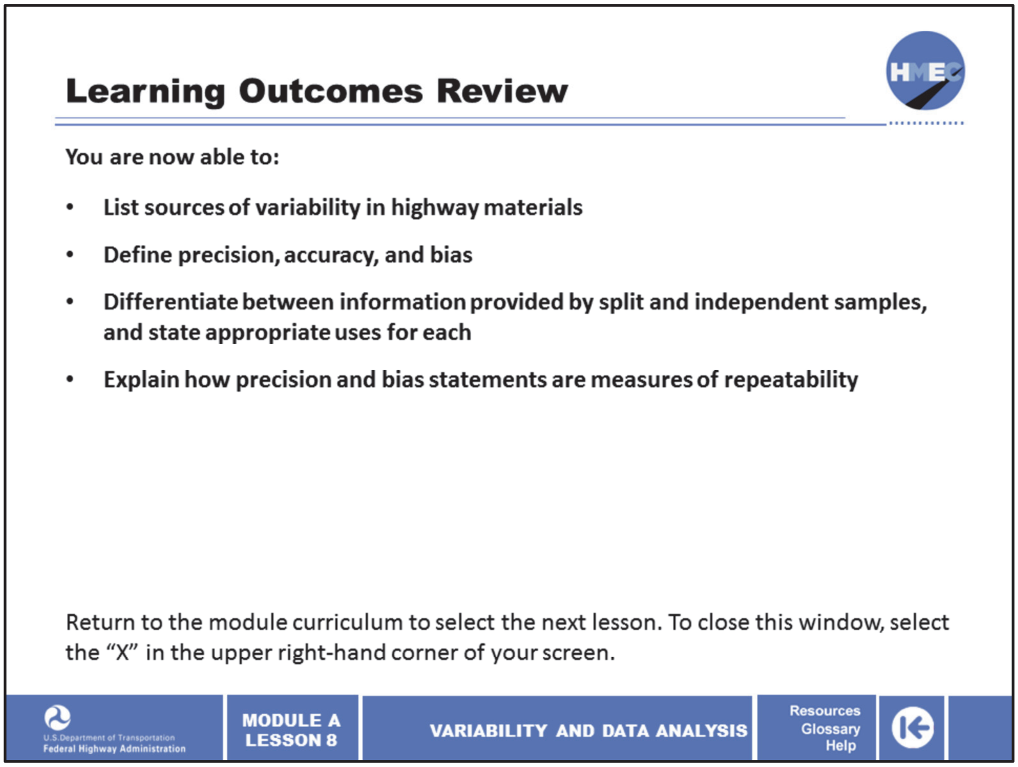
The standard deviation of the means is sigma x-bar which equals sigma divided by the square root of 4; thus, the square root of 4 equals 2 divided into the population sigma of 1.2% makes sigma x-bar equal to 0.6%.

Using the 68-95-99.7 rule for 1, 2, and 3 standard deviations with a sigma x-bar of 0.6 and μ = 5.6,

- 68% of the means will be within 5.6 ± 1 (0.6) = 5.0 and 6.2%;
- 95% of the means will be within 5.6 ± 2 (0.6) = 4.4 and 6.8%; and
- 99.7% of the means will be within 5.6 ± 3 (0.6) = 3.8 and 7.4%.

Image description: Equation.

Image description: A normal curve.



You have completed Module A, Lesson 8: Variability and Data Analysis. You are now able to:

- List sources of variability in highway materials;
- Define precision, accuracy, and bias;
- Differentiate between information provided by split and independent samples, and state appropriate uses for each; and
- Explain how precision and bias statements are measures of repeatability.

Close this lesson, and return to the module curriculum to select the next lesson. To close this window, select the "X" in the upper right-hand corner of your screen.