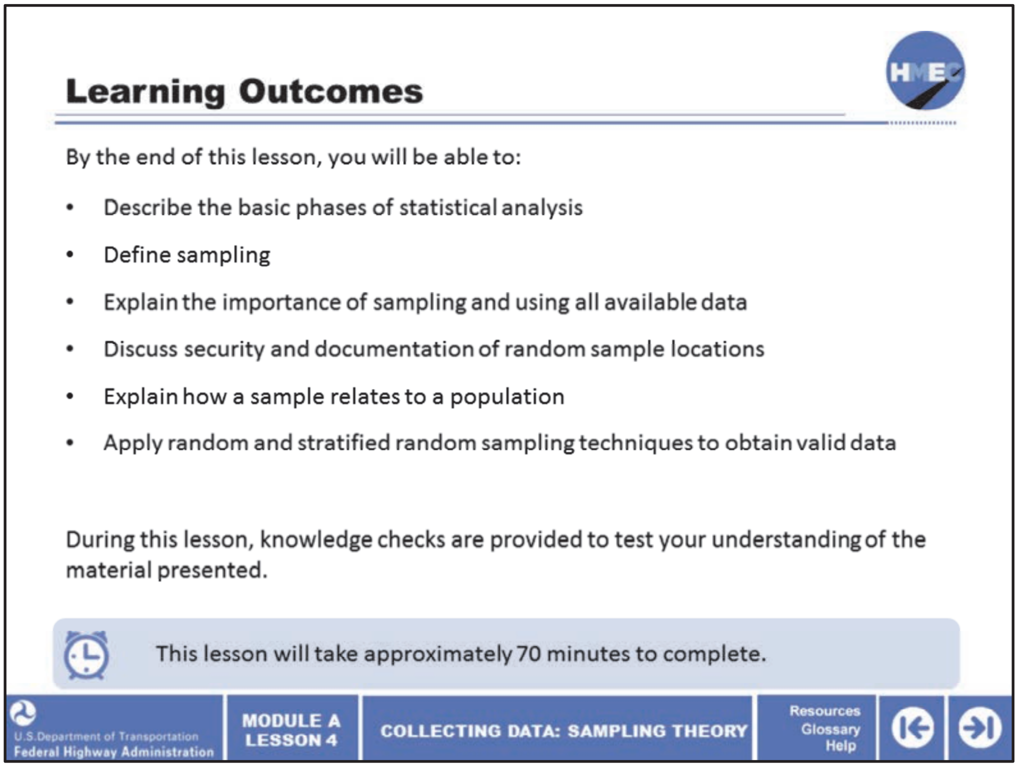
Welcome to the Highway Materials Engineering Course (HMEC) Module A, Lesson 4: Collecting Data: Sampling Theory. This lesson provides an understanding of the basic elements of a statistically-based quality assurance (QA) program and includes an introduction to quality assurance as well as techniques for collecting data.

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.

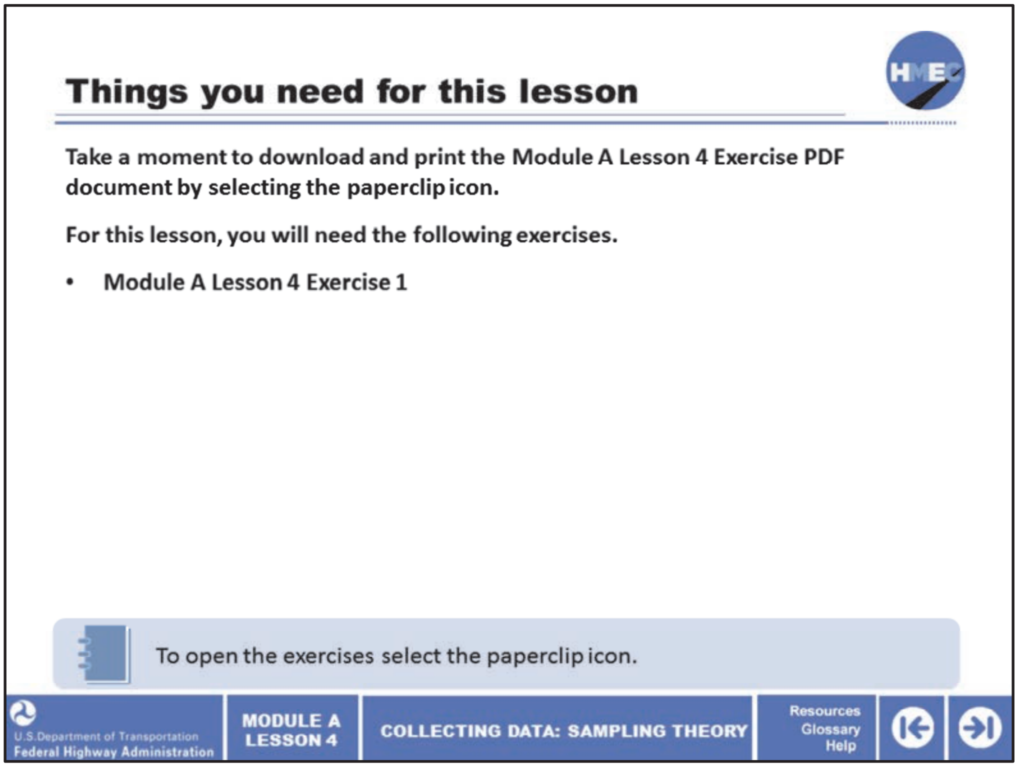


By the end of this lesson, you will be able to:

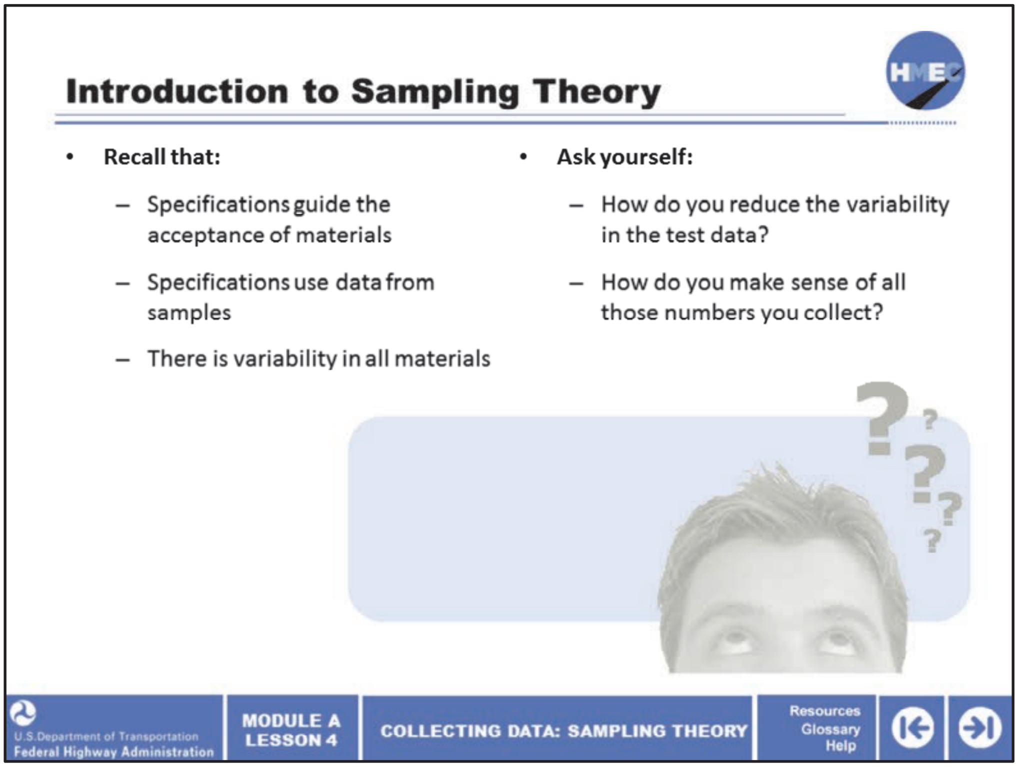
- Describe the basic phases of statistical analysis;
- Define sampling;
- Explain the importance of sampling and using all available data;
- Discuss security and documentation of random sample locations;
- Explain how a sample relates to a population; and
- Apply random and stratified random sampling techniques to obtain valid data.

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

This lesson will take approximately 70 minutes to complete.



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

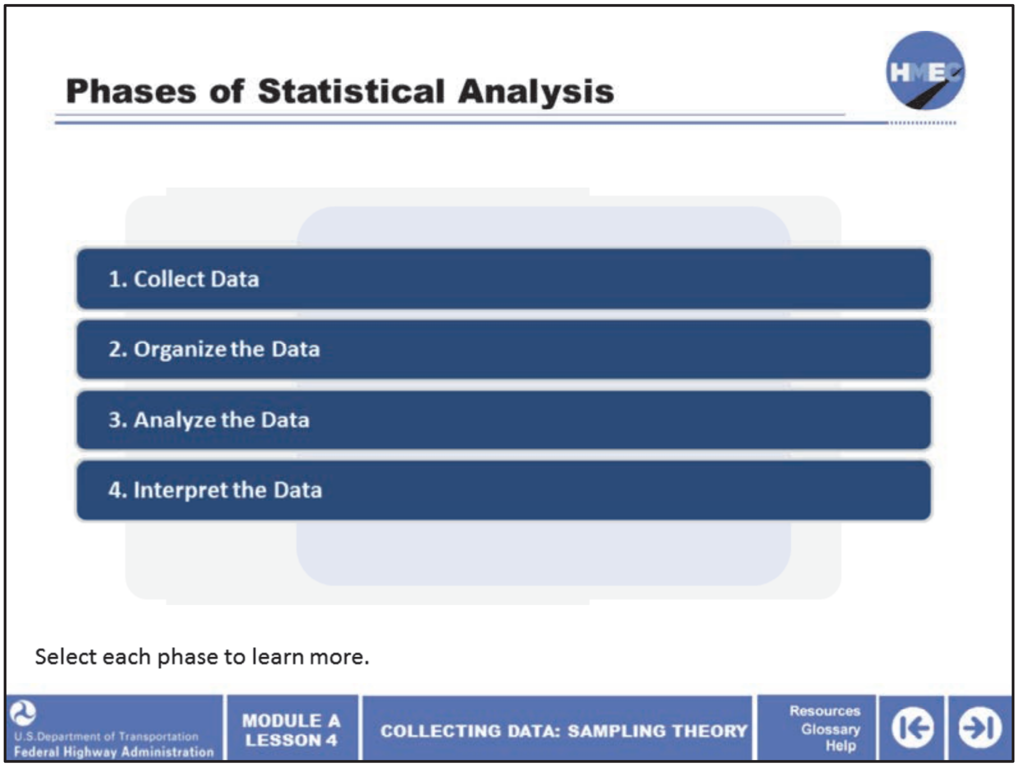


Remember, specifications guide the acceptance of materials by quantifying the risk we are willing to assume. We use data from samples to indicate how close to "normal" or acceptable the materials are. There is variability in all materials, and in sampling and testing as well.

So ask yourself the following questions:

• How do you reduce the variability in the test data? Answer: Make sure to follow the correct sampling and testing procedures.
• How do you make sense of all those numbers you collect? Answer: One way is to organize the data correctly as you will see in Lesson 5.

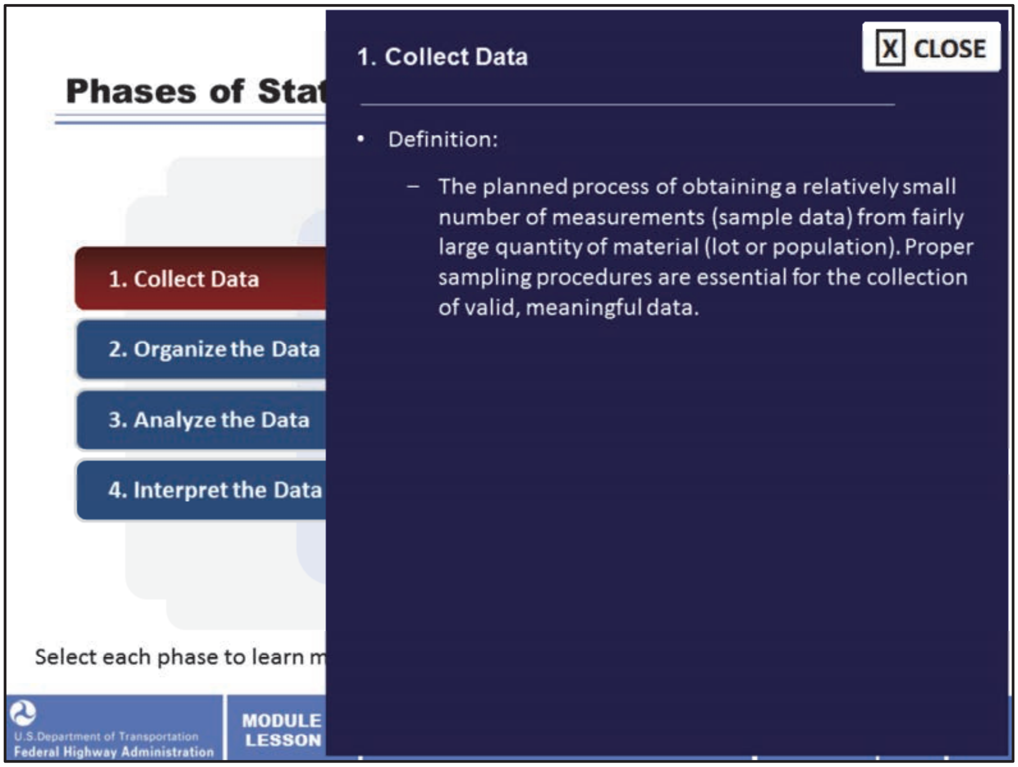
Image description: Man looking up with question marks.



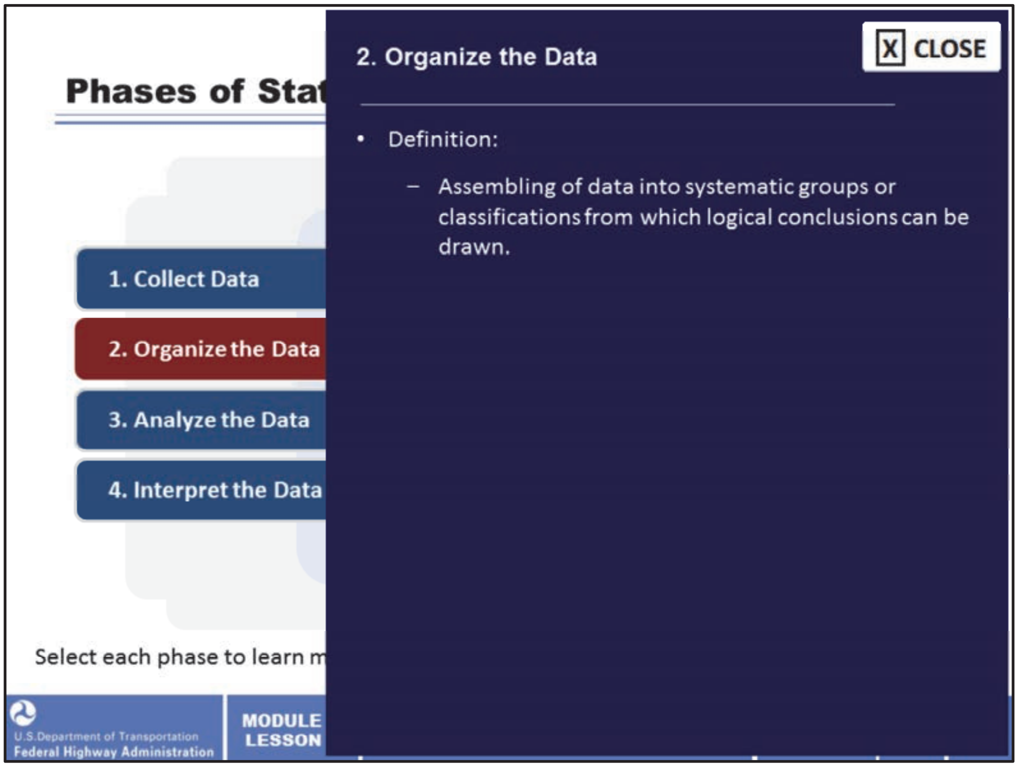
There are four phases of statistical analysis, which include:

1. Collect Data;
2. Organize the Data;
3. Analyze the Data; and
4. Interpret the Data**.**

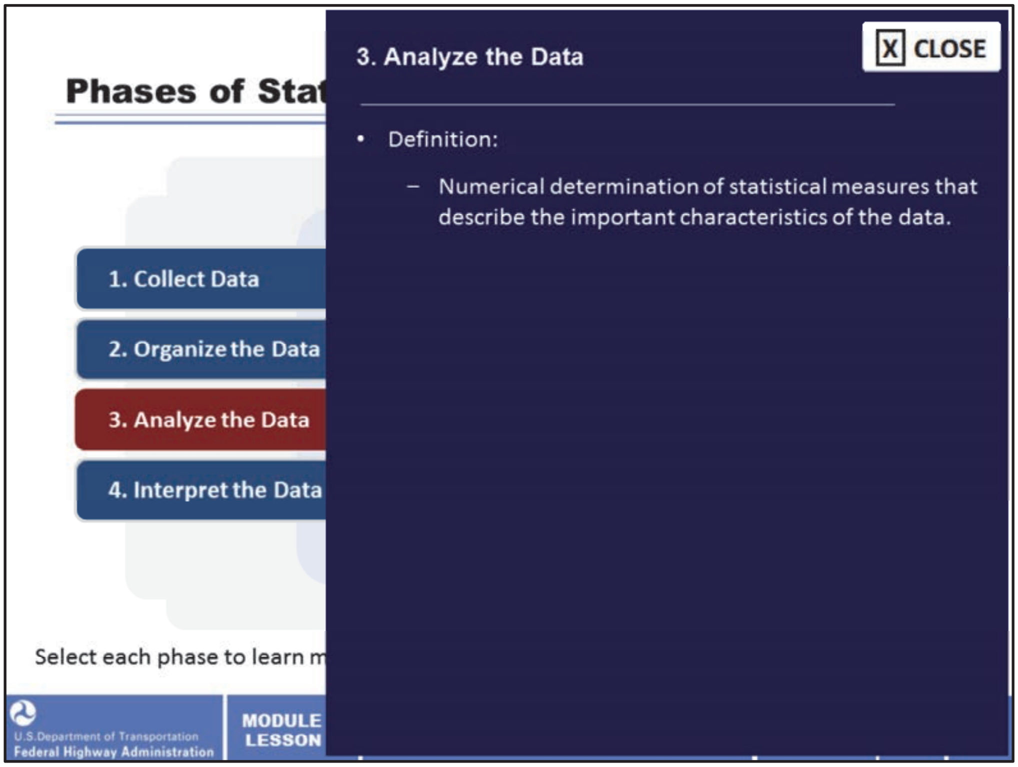
Select each phase to learn more.



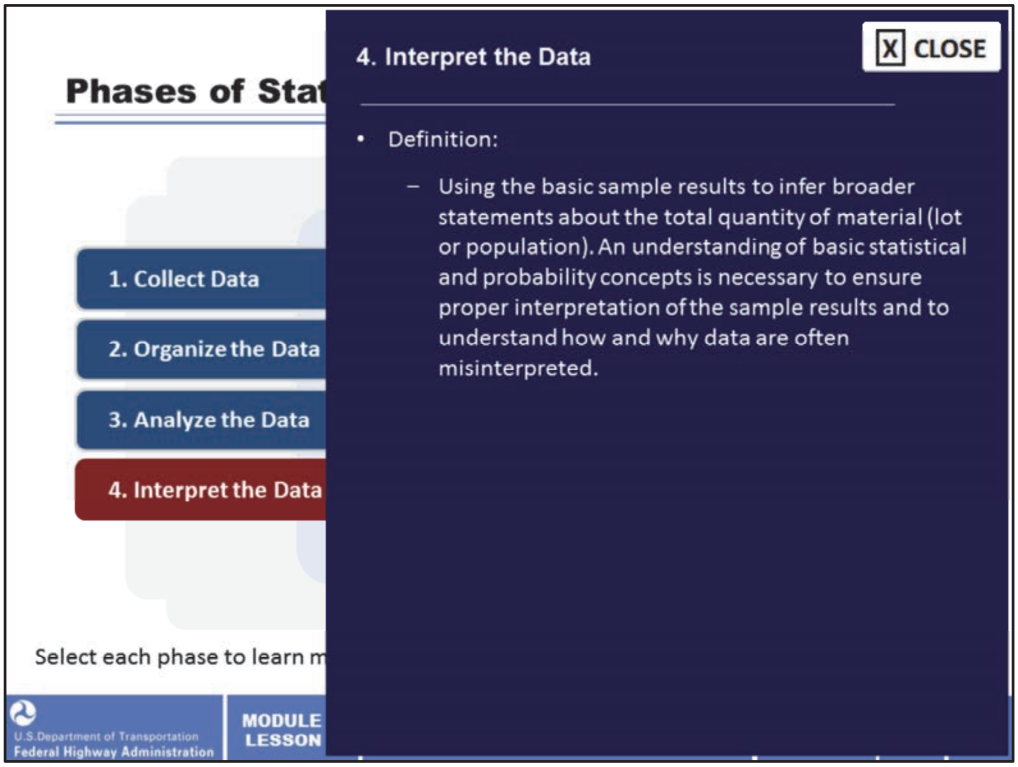
The first phase, to collect data, is the planned process of obtaining a relatively small number of measurements (sample data) from a fairly large quantity of material (lot or population). Proper sampling procedures are essential for the collection of valid, meaningful data.



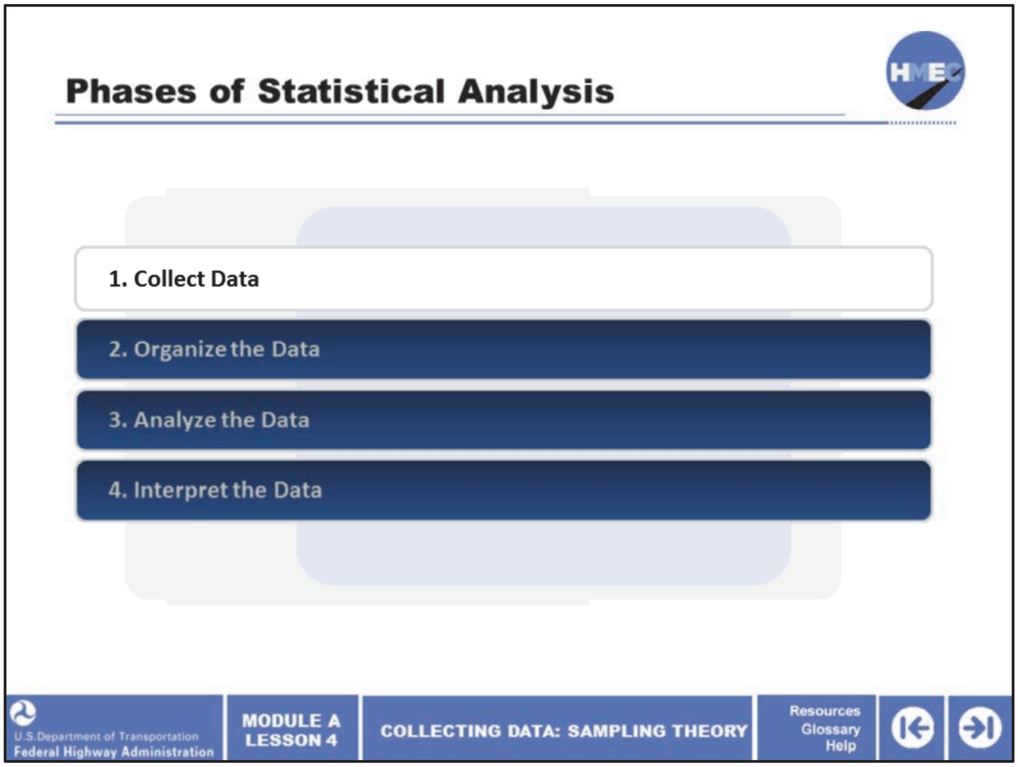
The second phase is to organize the data, which requires assembling of data into systematic groups or classifications from which logical conclusions can be drawn.



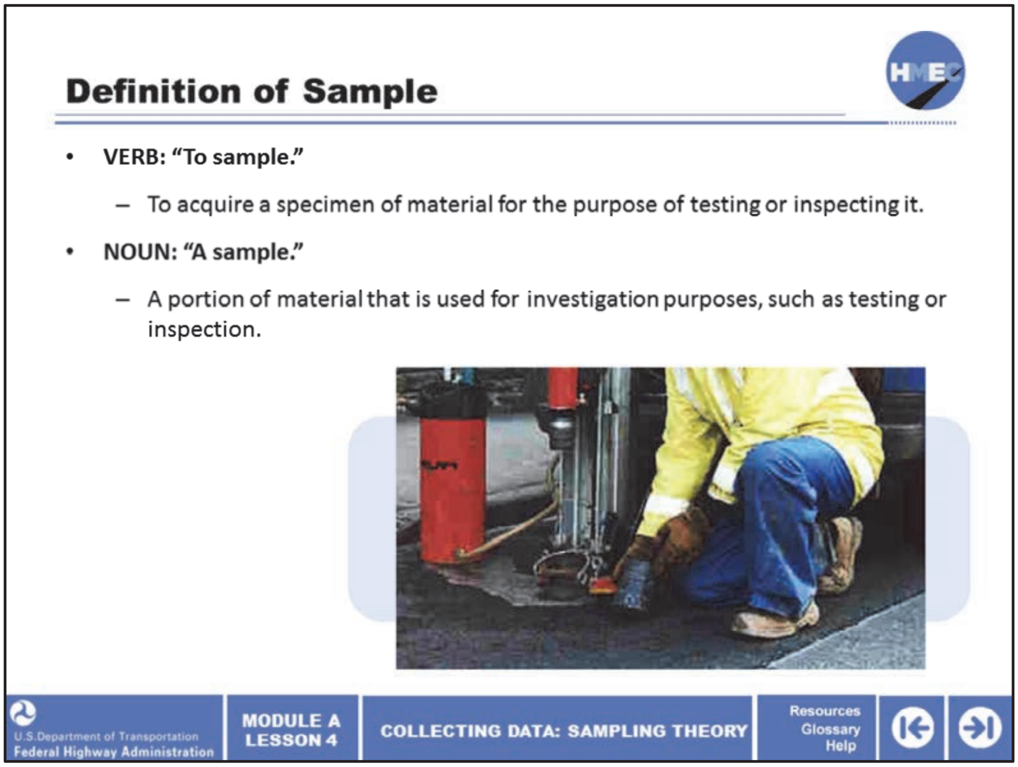
The third phase of statistical analysis is to analyze the data, which is a numerical determination of statistical measures that describe the important characteristics of the data.



The fourth and last phase is to interpret the data, which means to use the basic sample results to infer broader statements about the total quantity of material (lot or population). An understanding of basic statistical and probability concepts is necessary to ensure proper interpretation of the sample results and to understand how and why data are often misinterpreted.



In this lesson, we'll only take a look at the first phase: collect data. Specifically, how sampling theory is applied to collect data for organization, analysis, and interpretation. Data are of little value if they are collected in a haphazard, unplanned manner. We'll cover the other phases in the following lessons.



Let's define what a sample is.

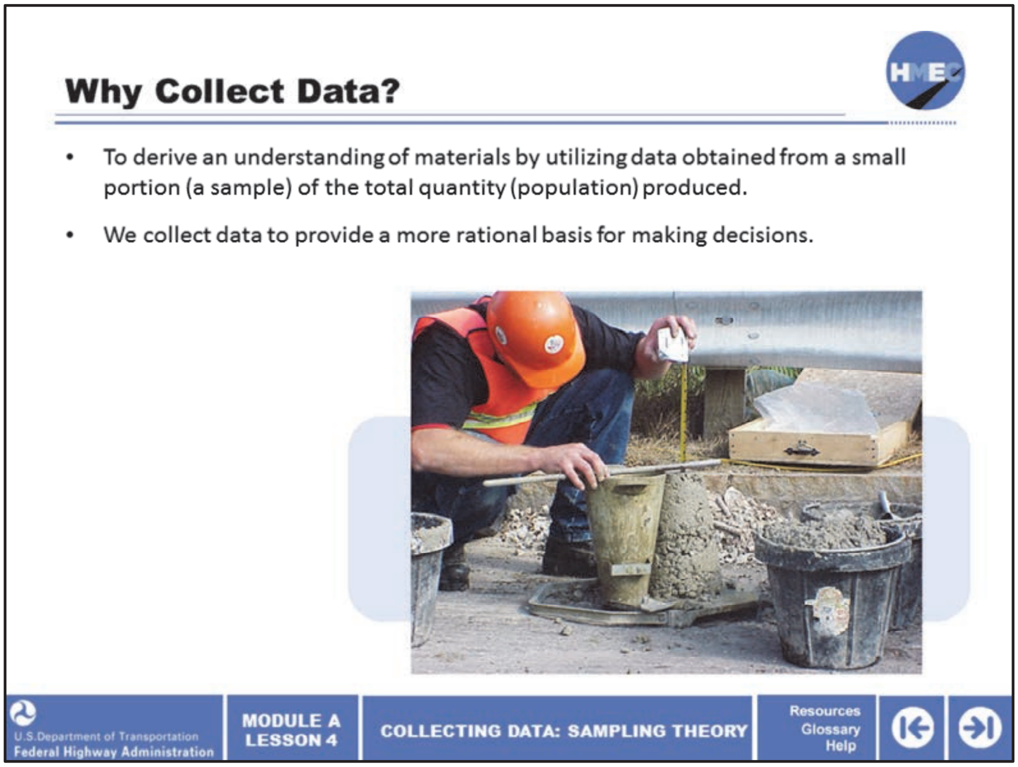
- "To sample," the verb, means to acquire a specimen of material for the purpose of testing or inspecting it.
- "A sample," the noun, is a portion of material that is used for investigation purposes, such as testing or inspection.

When we go out to the job site and take a portion of material, that is "sampling" the material. That portion we took during sampling is called "the sample." It is this latter use of the term sample that is used throughout this module.

A sample represents part of the whole project and it takes multiple samples to establish a method to evaluate or analyze the acceptability of material.

When we collect data, we take samples. Remember the risk scenarios with marbles? We were sampling the bag's contents, and the set of three marbles was the sample of what was in the bag and we call that "a sample size of three," not three samples.

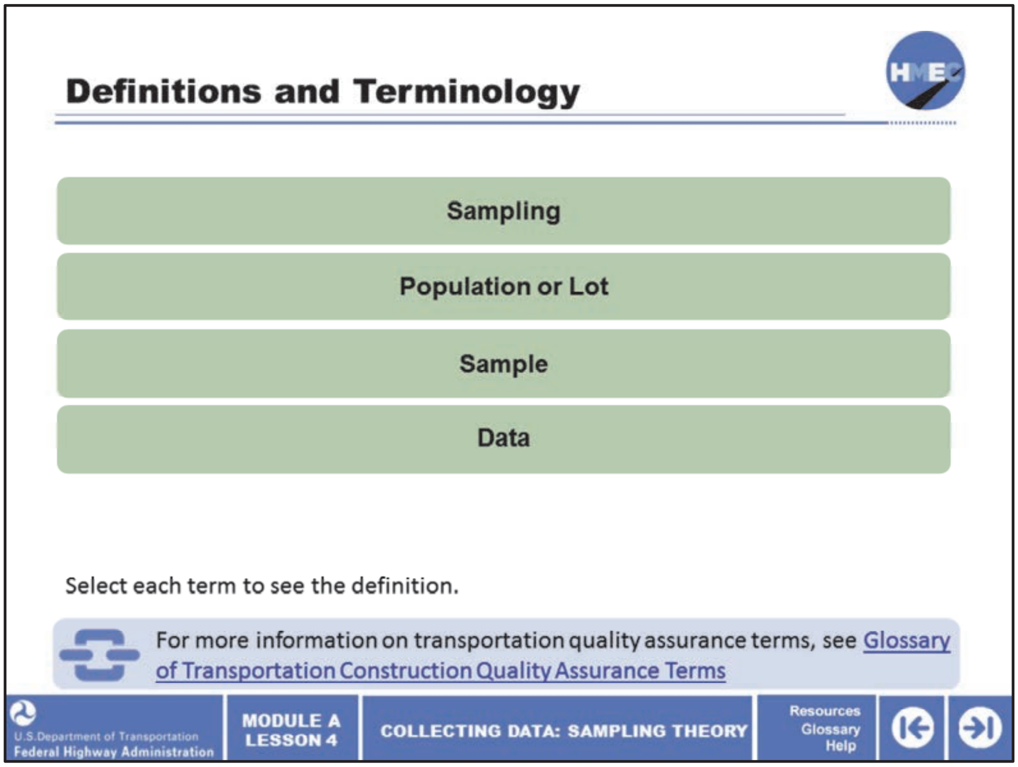
Image description: Worker gathering material sample.



Why do we collect data? The objective of statistical analysis, as it relates to construction materials, is to derive an understanding of these materials by utilizing data obtained from a small portion (a sample) of the total quantity (population) produced. This provides a more rational basis for making decisions.

The gentleman you see here is collecting concrete slump data using a slump cone.

Image description: Worker gathering material sample.



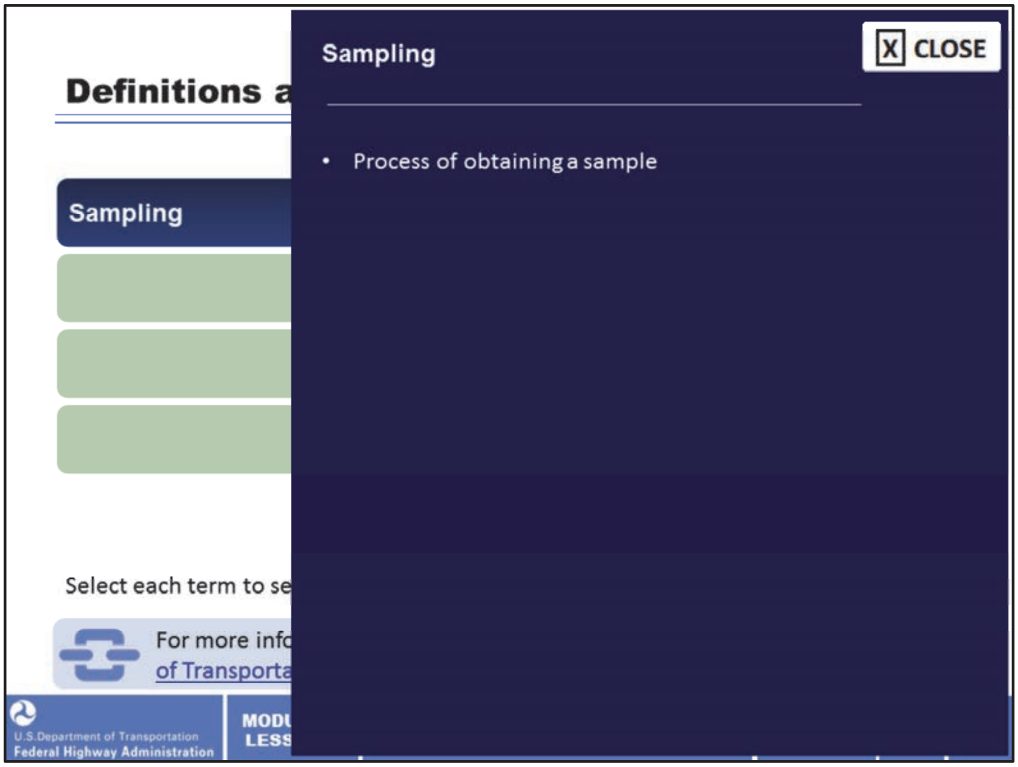
Definitions and terms related to collecting data are used throughout the module. They include:

- Sampling;
- Population or Lot (terms are used interchangeably);
- Sample; and
- Data.

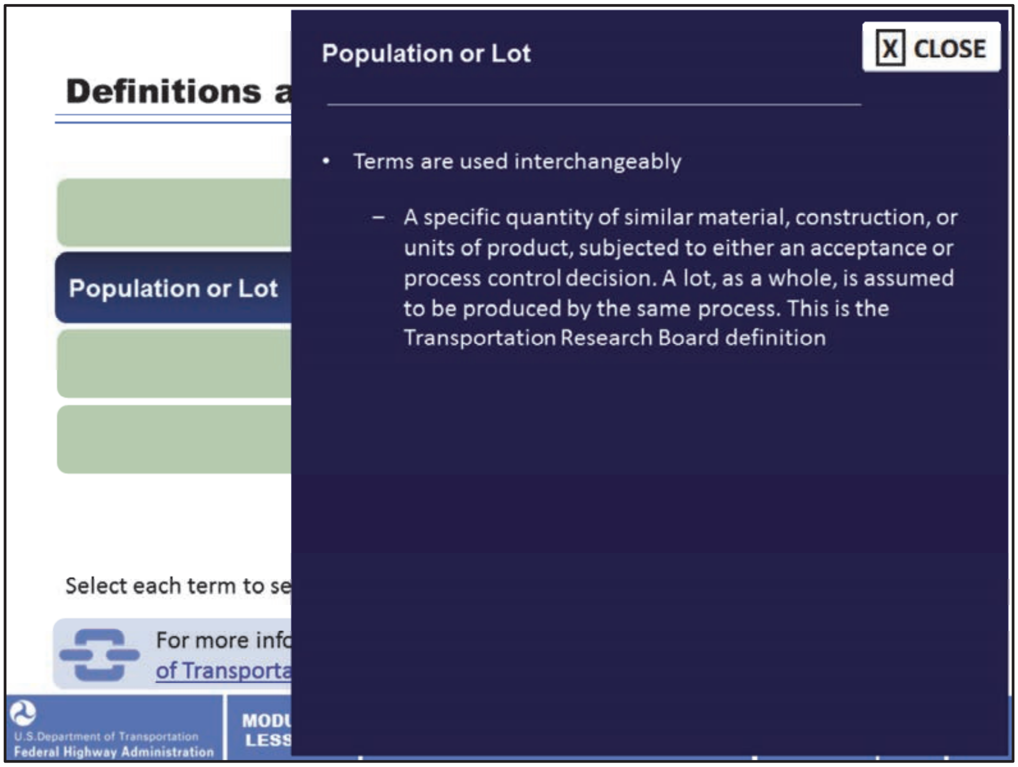
Select each term to see the definition.

For more information on transportation quality assurance terms, see Glossary of Transportation Construction Quality Assurance Terms.

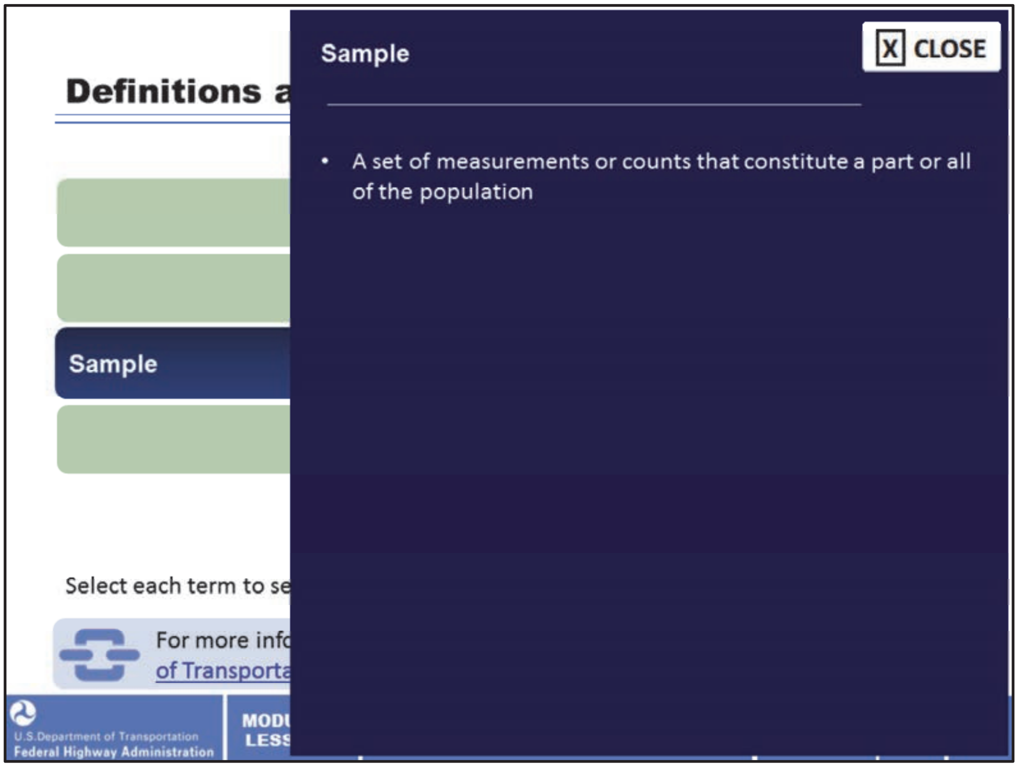
Hyperlink description: http://onlinepubs.trb.org/onlinepubs/circulars/ec173



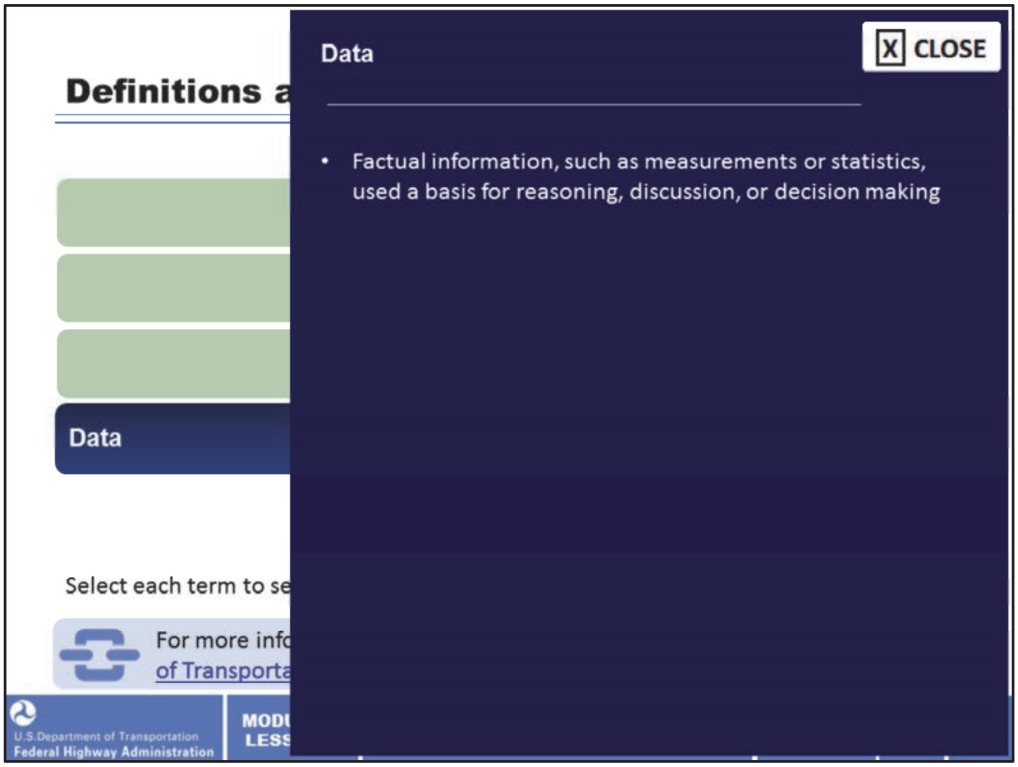
Sampling is the process of obtaining a sample.



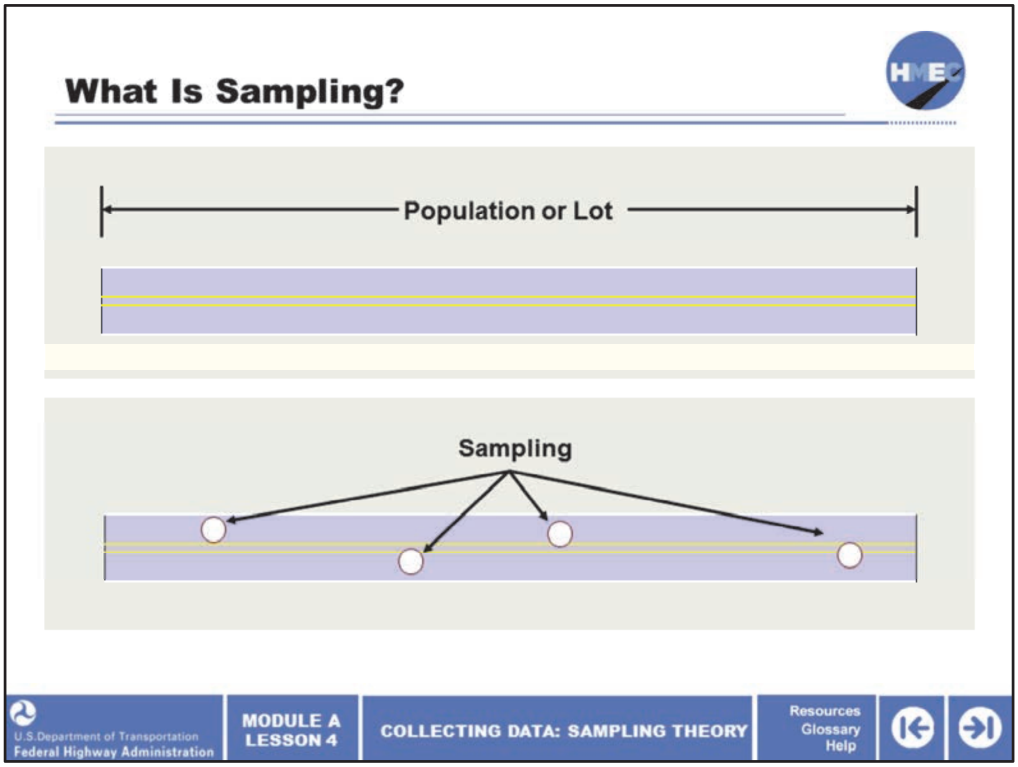
Population or Lot are terms that are used interchangeably. A population or lot is a specific quantity of similar material, construction, or units of product, subjected to either an acceptance or process control decision. A lot, as a whole, is assumed to be produced by the same process. This is the Transportation Research Board definition.



A sample is a set of measurements or counts that constitute a part or all of the population.



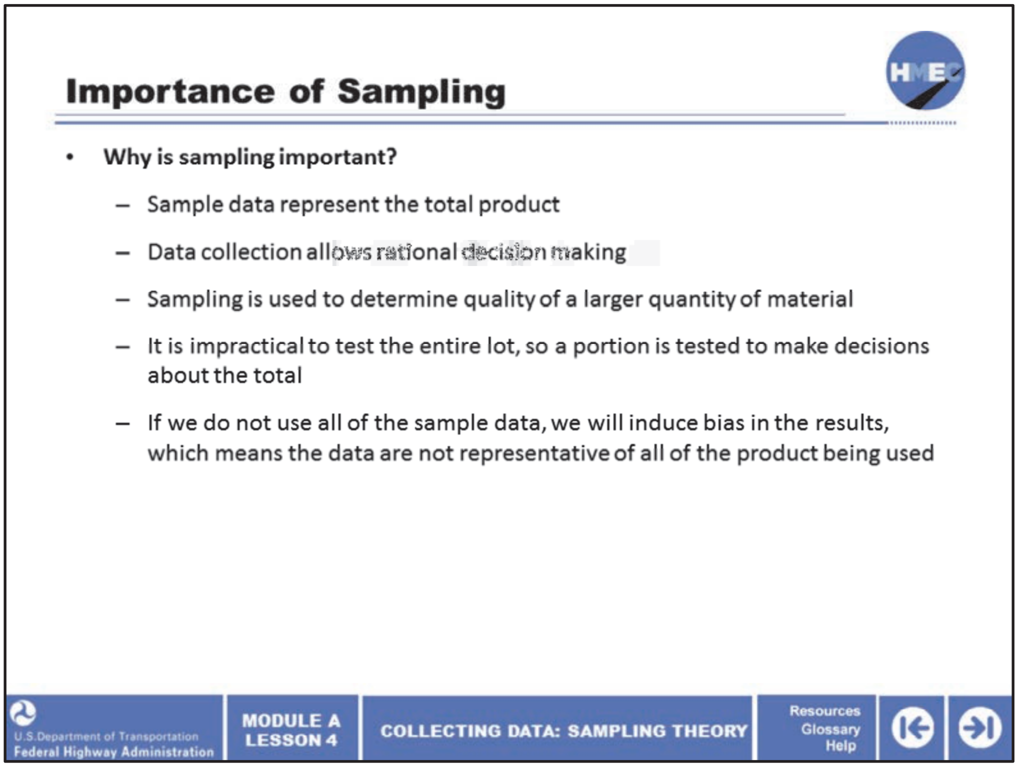
Data is factual information, such as measurements or statistics, used as a basis for reasoning, discussion, or decision making.



Typically, sampling is used to determine the characteristics (slump or asphalt content, for example) of a larger quantity of material. The lot is usually too large to test or evaluate in its entirety, so we must evaluate a portion of the lot, which we call a sample, to make decisions about the total lot.

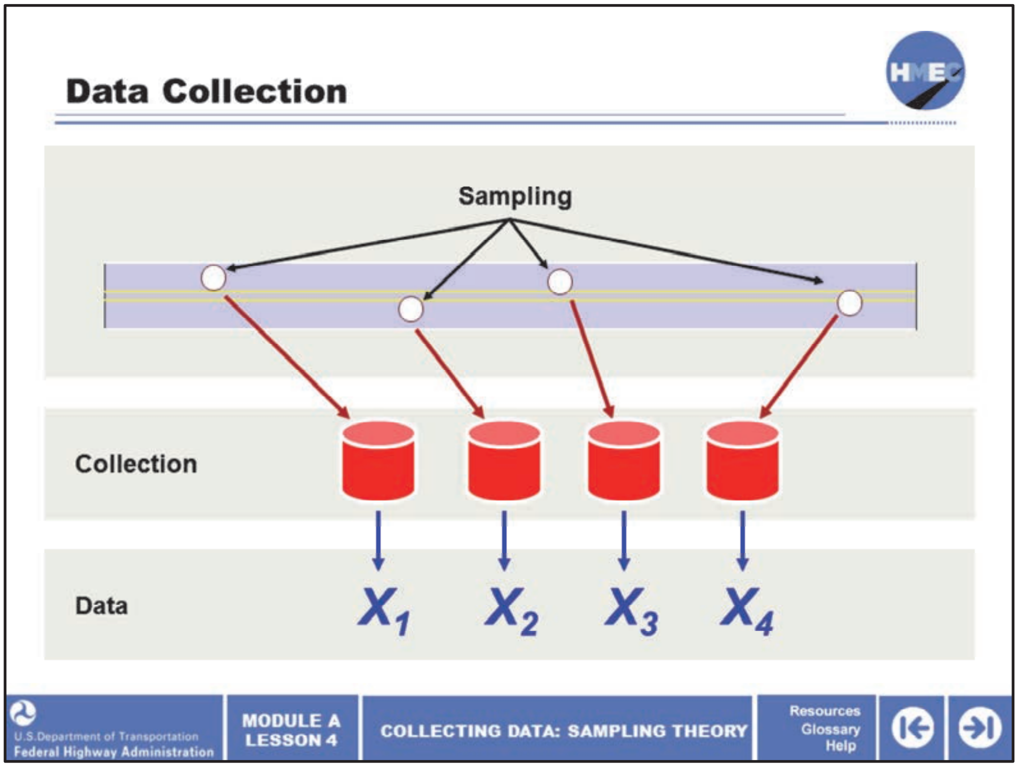
Image description: Example of Population or Lot.

Image description: Example of Sampling.



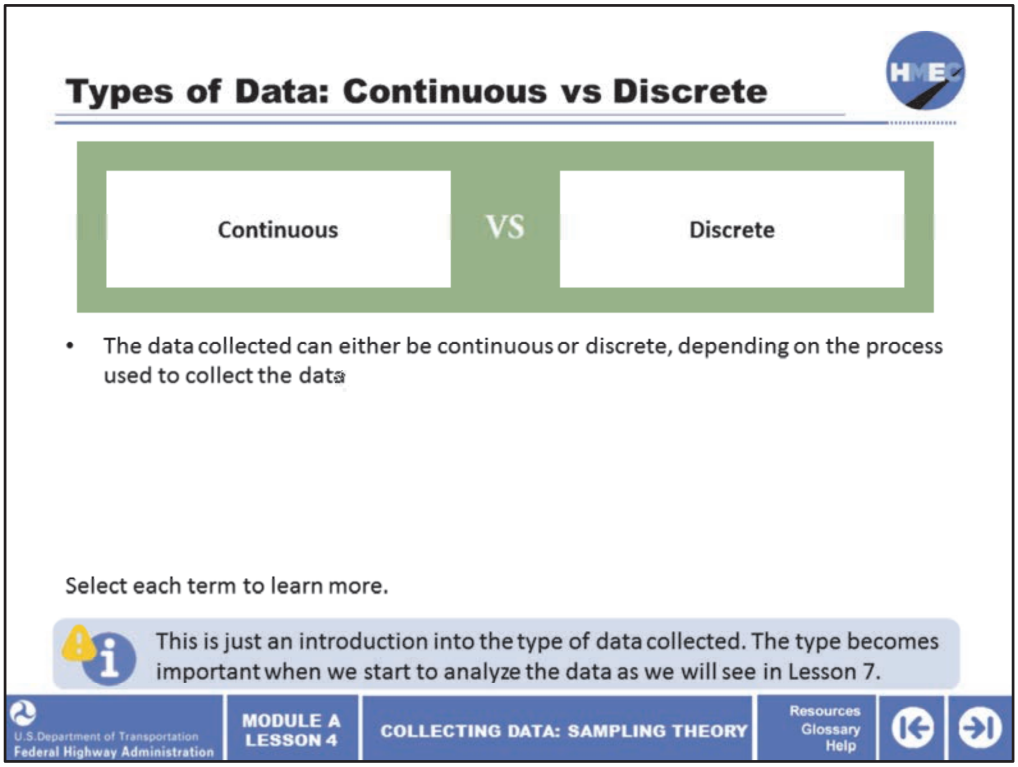
Why is sampling important?

- Sampling provides data that represent the total product. Statistical analyses use sampling data to estimate the quality of that product.
- Data collection allows rational decision making.
- Sampling is used to determine the quality characteristics, for example some property of interest, of a larger quantity of material, or population.
- It is impractical to test the entire lot, so a portion is tested to make decisions about the total lot using all available data.
- If we do not use all of the data provided by the sampling, we will induce bias in the results, which means the data are not representative of all of the product being used.



Data collection is done by obtaining a sample from a population. Here, the population is a strip of roadway from which the sample is taken. The sample provides information about a particular portion of the population. Each one of the samples obtained from the stretch of road provides a separate piece of data. All of the data collected during the process of sampling feed into the overall statistical analysis done on the product. In this example, cores are taken from a pavement to estimate some quality characteristic such as thickness or density. So the process of obtaining the data is from the top down, that is, the population provides the sample, which then provides the data. But the analysis is from the bottom up, that is, the data provides information about the sample, which then provides information about the population.

Image description: Example of Sampling, Collection and Data.

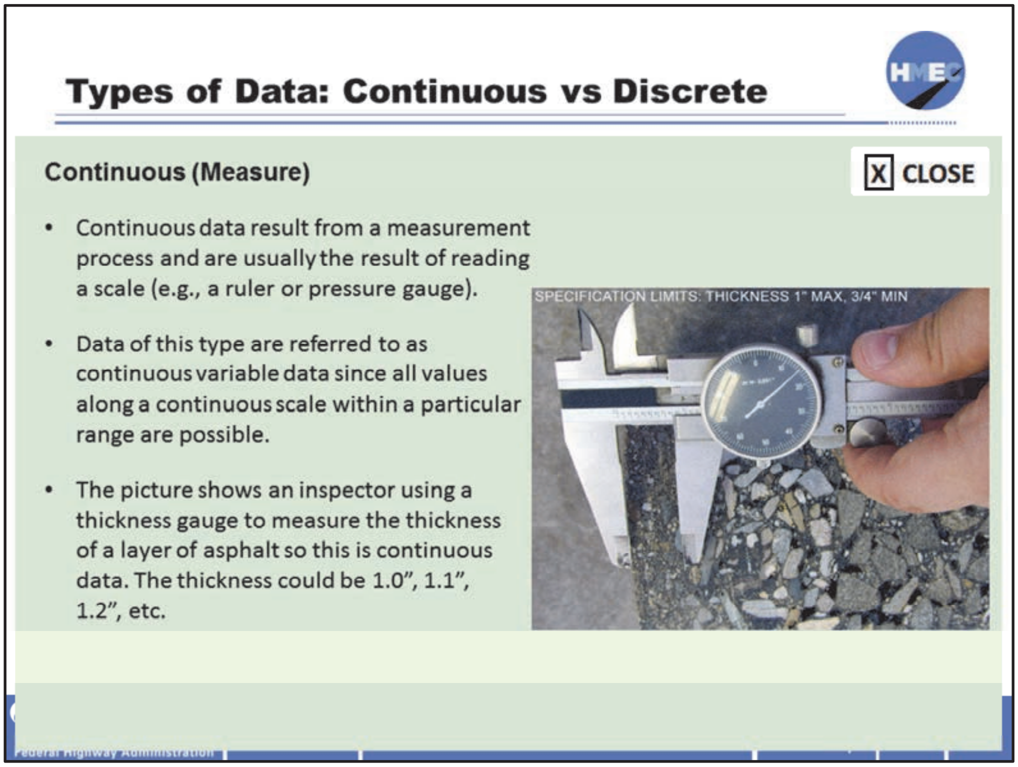


The data collected can either be continuous or discrete, depending on the process used to collect the data**.** The two types of data are:

- Continuous; and
- Discrete.

Select each term to learn more.

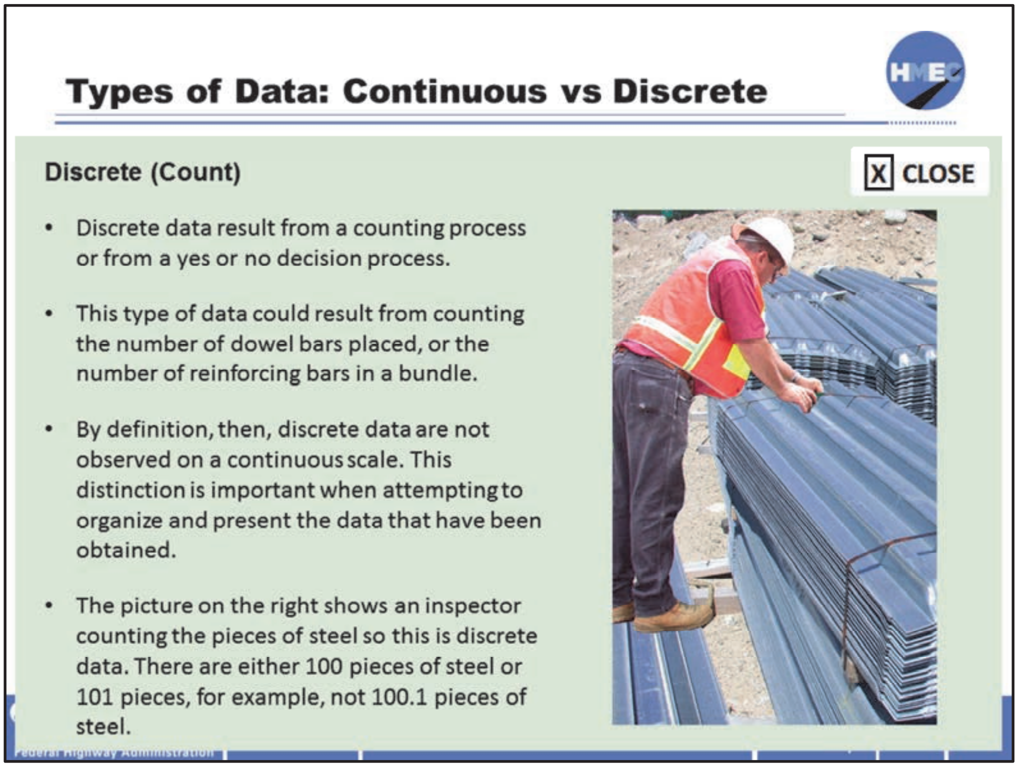
This is just an introduction into the type of data that is collected. The type of data becomes important when we start to analyze the data as we will see in Lesson 7.



Continuous data result from a measurement process and are usually the result of reading a scale (e.g., a ruler or pressure gauge). Data of this type are referred to as continuous variable data since all values along a continuous scale within a particular range are possible.

The picture on the left shows an inspector using a thickness gauge to measure the thickness of a layer of asphalt, so this is continuous data. The thickness could be 1.0”, 1.1”, 1.2”, etc.

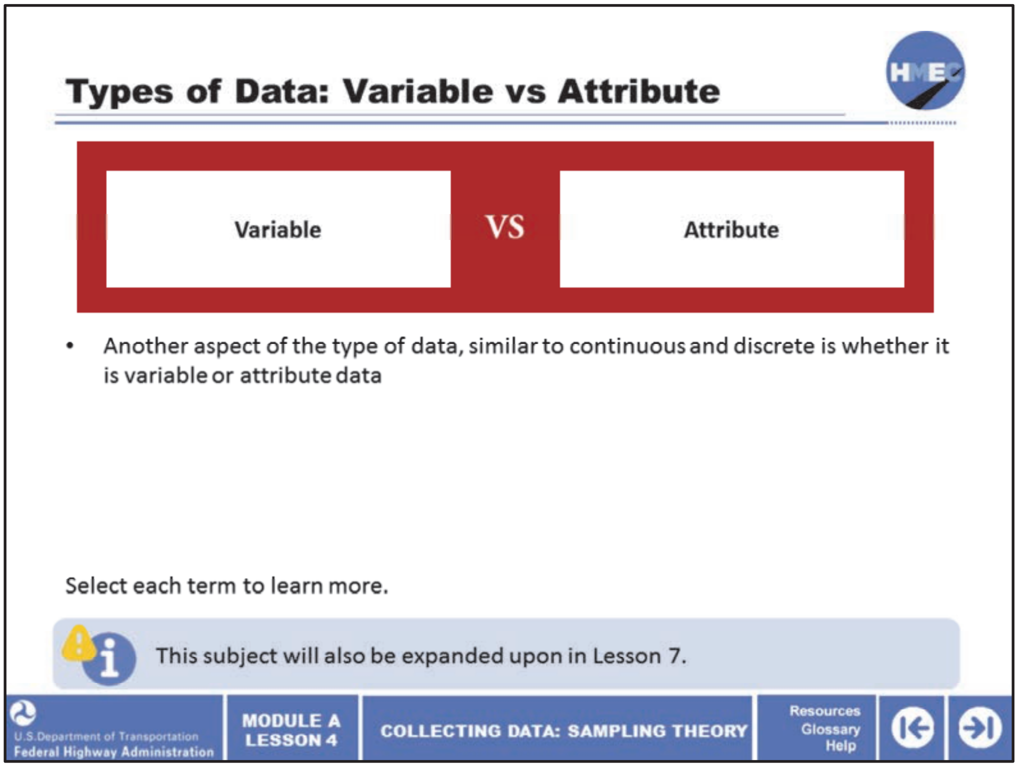
Image description: An inspector using a thickness gauge to measure the thickness of a layer of asphalt.



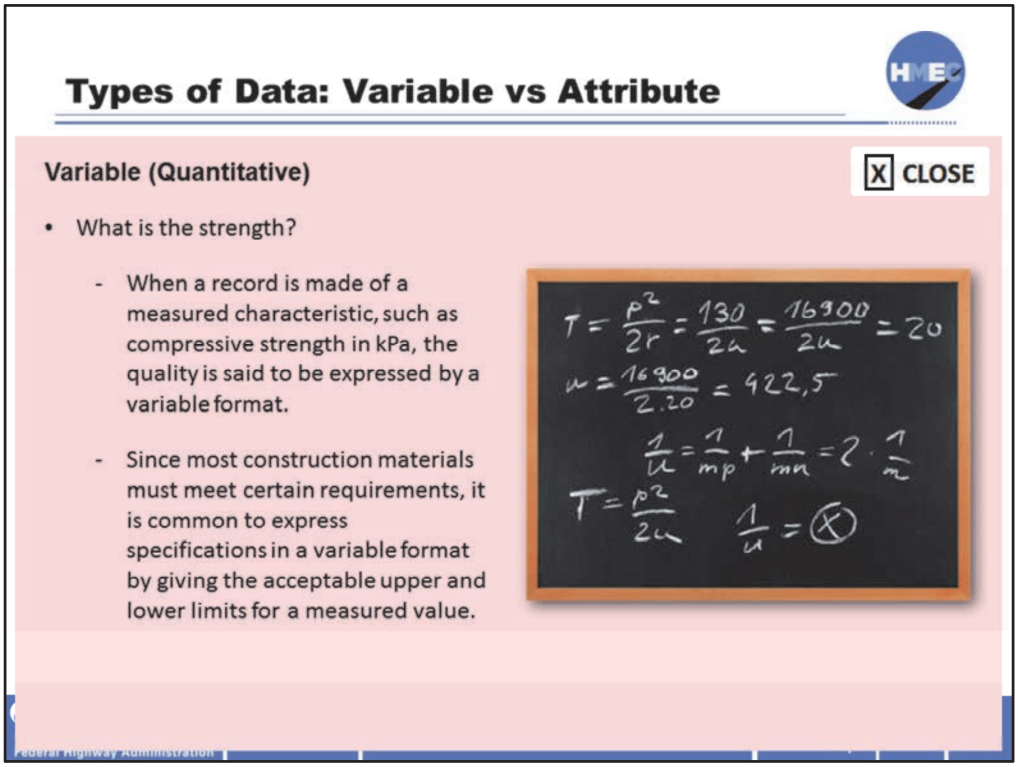
Discrete data result from a counting process or from a yes or no decision process. This type of data could result from counting the number of dowel bars placed, or the number of reinforcing bars in a bundle. By definition, then, discrete data are not observed on a continuous scale. This distinction is important when attempting to organize and present the data that have been obtained.

In the slide, the picture on the right shows an inspector counting the pieces of steel so this is discrete data. There are either 100 pieces of steel or 101 pieces, for example, not 100.1 pieces of steel.

Image description: An inspector counting the pieces of steel.

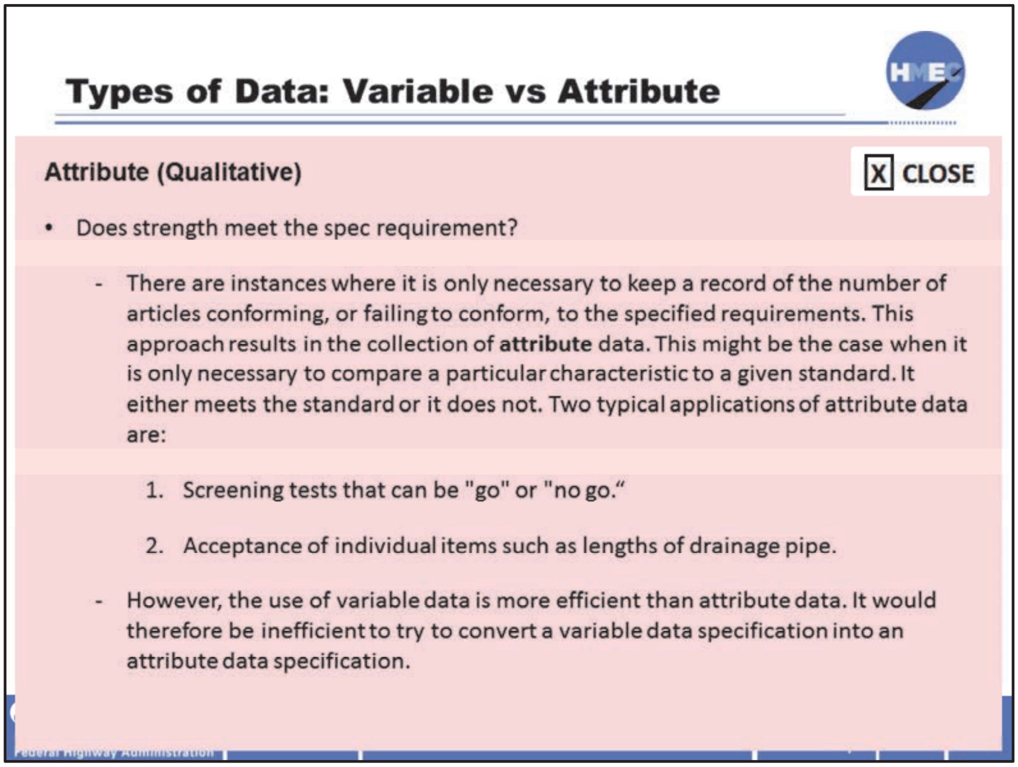


Another aspect of the type of data, similar to continuous and discrete, is whether it is variable data or attribute data. Select each term to learn more.



When a record is made of a measured characteristic, such as compressive strength in kPa, the quality is said to be expressed by a variable format. Since most construction materials must meet certain requirements, it is common to express specifications in a variable format by giving the acceptable upper and lower limits for a measured value.

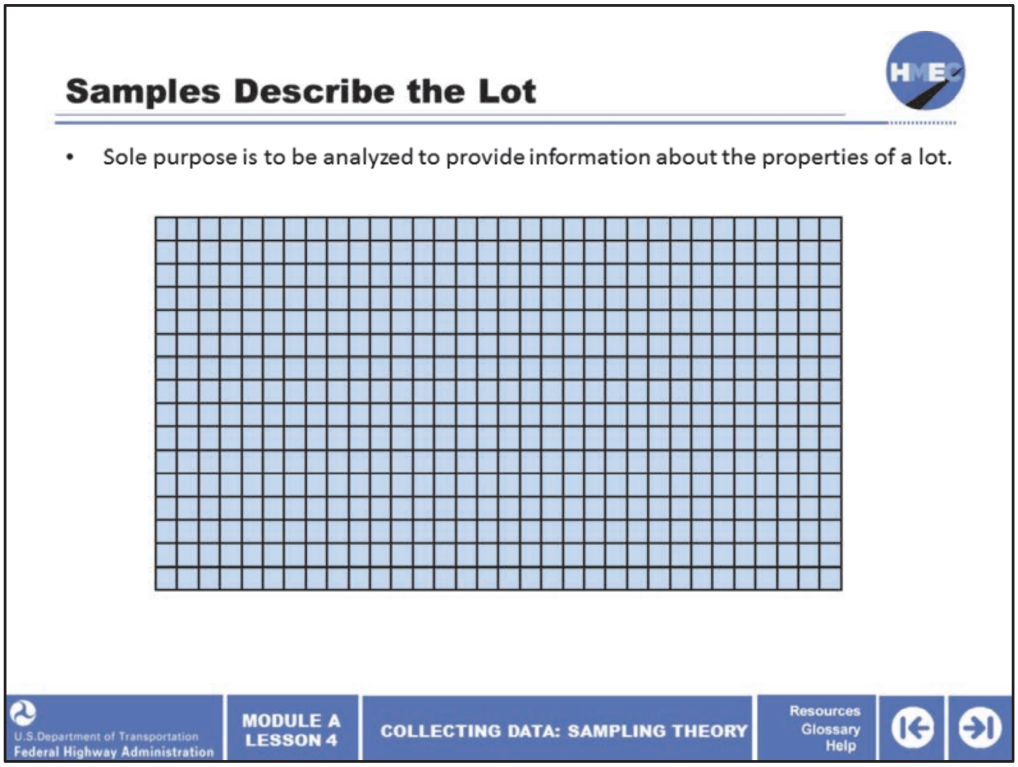
Image description: Blackboard.



There are instances where it is only necessary to keep a record of the number of articles conforming, or failing to conform, to the specified requirements. This approach results in the collection of attribute data. This might be the case when it is only necessary to compare a particular characteristic to a given standard. It either meets the standard or it does not. Two typical applications of attribute data are:

1. Screening tests that can be "go" or "no go.“
2. Acceptance of individual items such as lengths of drainage pipe.

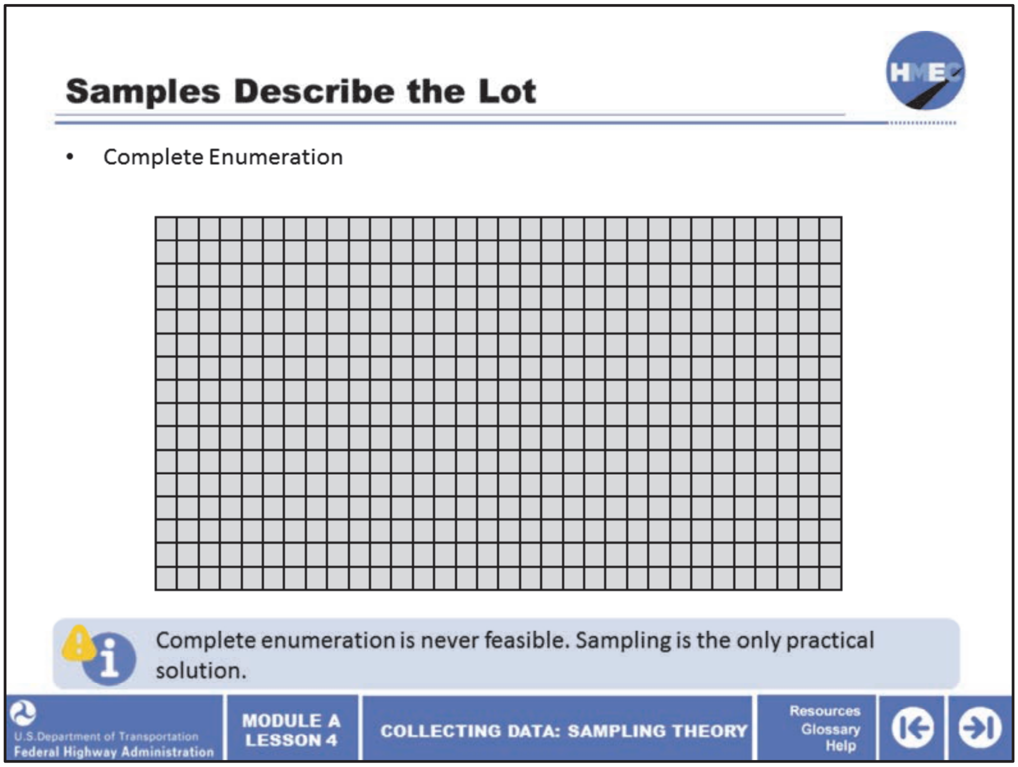
However, the use of variable data is more efficient than attribute data. It would therefore be inefficient to try to convert a variable data specification into an attribute data specification.



Now returning to the discussion of samples. Their sole purpose is to be analyzed to provide information about the properties of a lot.

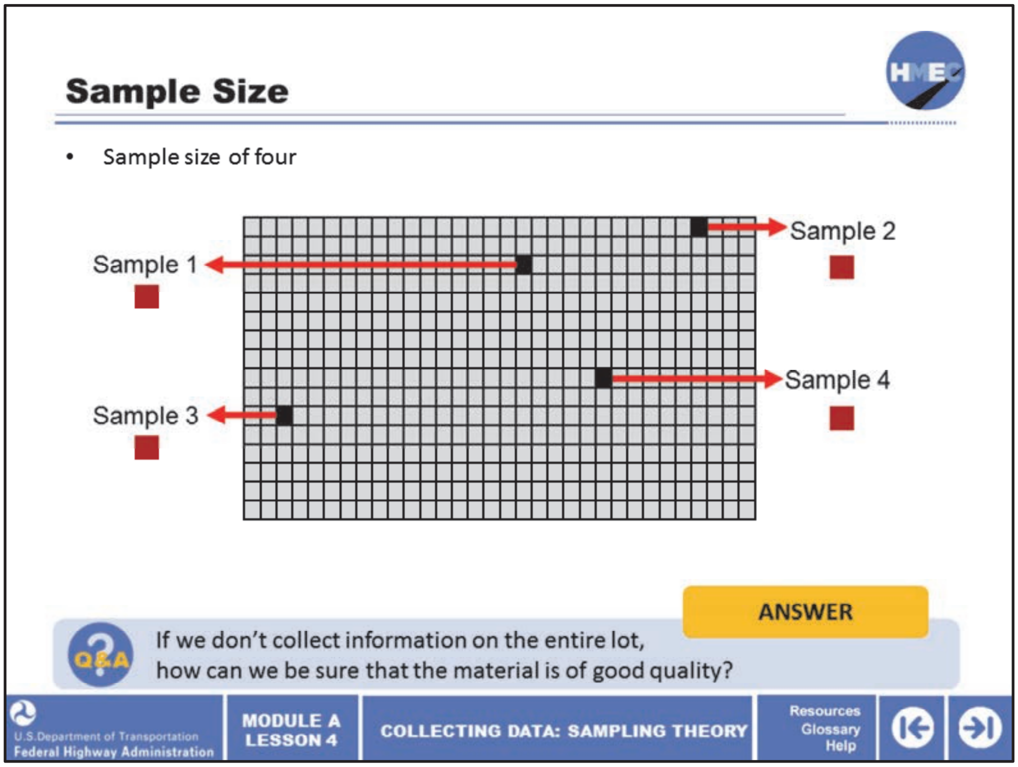
The large rectangle can be thought of as a lot consisting of the smaller squares that represent potential samples. You should keep in mind that it is always the properties of the lot that we wish to identify but that can be done only through a sample. And although this rectangle is finite, that is, it contains a certain number of potential samples that can be counted, in reality a population contains an infinite number of potential samples.

Image description: Example of a lot.



Obviously, to determine the "best" estimate of some property, such as thickness, every bit of pavement in the lot should be tested. This is called “complete enumeration.” However, complete enumeration is never feasible. For instance, in destructive testing, such as taking cores, the entire lot would be destroyed. For nondestructive testing, such as nuclear density tests, there is neither time nor labor available for complete enumeration. Therefore, sampling is the only practical solution.

Image description: Example of a lot.



The way the sample locations are chosen is very important.

Although the samples are numbered in the figure, we call this a sample size of four. The samples collected are used to provide an estimate of the overall quality of the lot. A sample can be selected from the lot of material, made up of several parts, and the data from the sample can be used to estimate some property in the lot in order to make a decision regarding its acceptability.

Select the box to answer the question, If we don't collect information on the entire lot, how can we be sure that the material is of good quality?

Image description: Example of a lot.

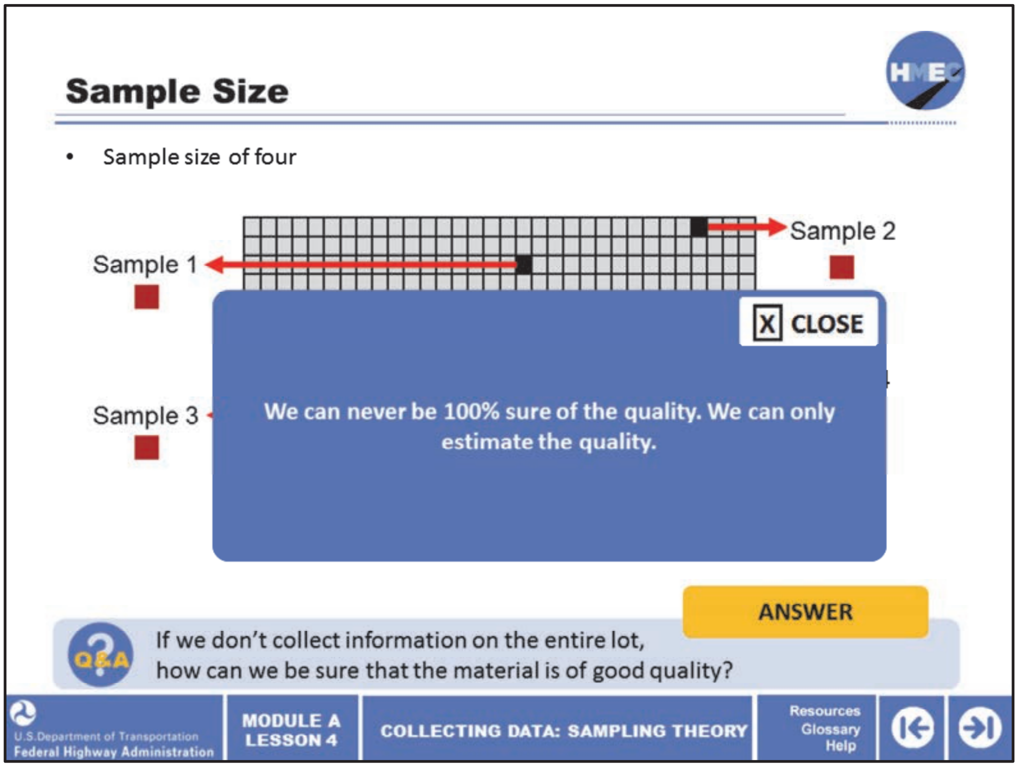
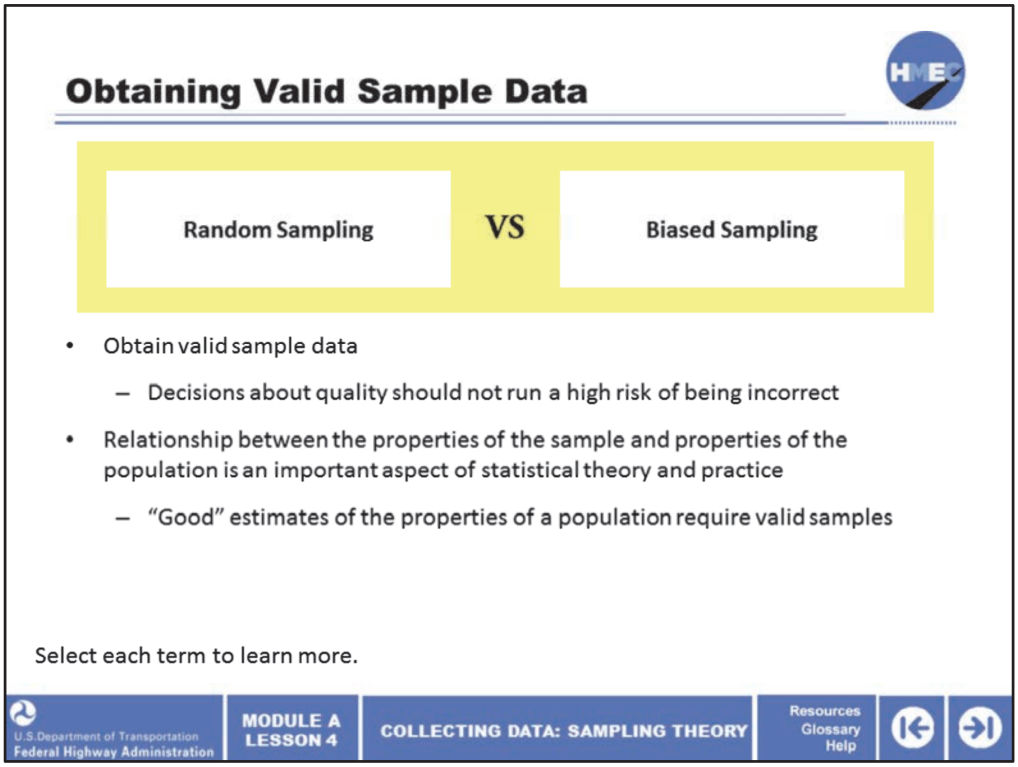


Image description: Example of a lot.



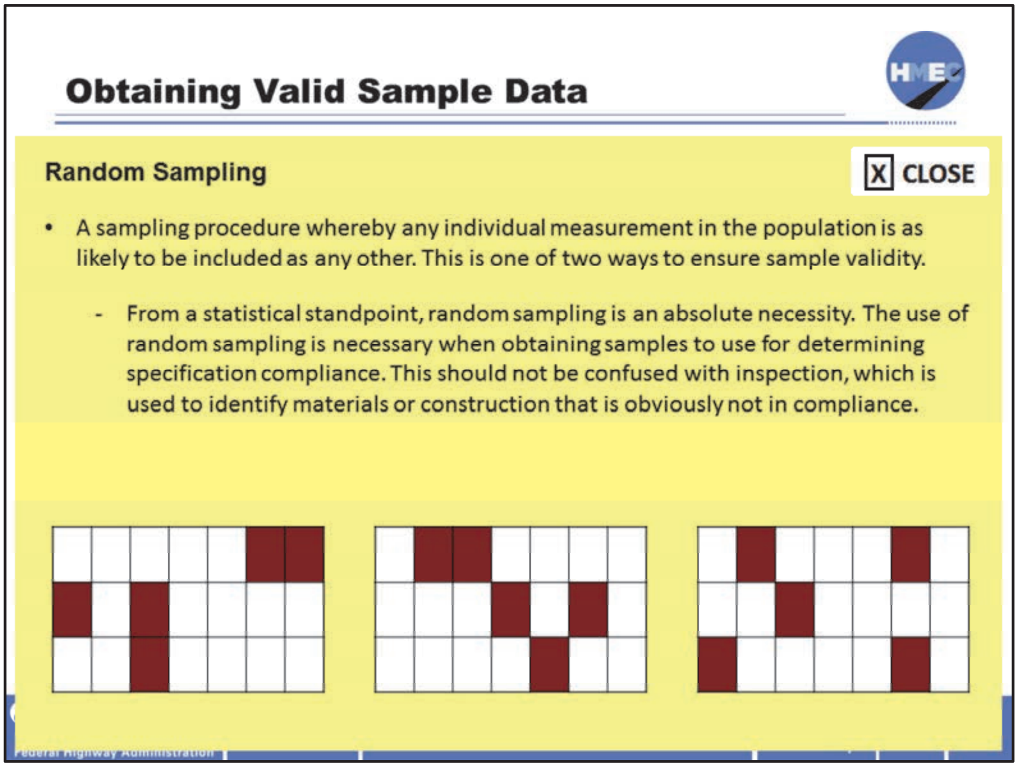
It is important to obtain valid sample data. If samples of a lot are going to be used to make determinations about quality, we don't want the decisions made about the quality of the material to run a high risk of being incorrect.

The relationship between the properties of the sample and the properties of the population is an important aspect of statistical theory and practice since "good" estimates of the properties of a population require valid samples.

Obtaining valid samples is not automatic. The following are two possible procedures for obtaining samples:

1. Random Sampling; and
2. Biased Sampling.

Select each term to learn more.



Random sampling is a sampling procedure whereby any individual measurement in the population is as likely to be included as any other. This is one of two ways to ensure sample validity.

Image description: Three examples of a lot.

Obtaining Valid Sample Data

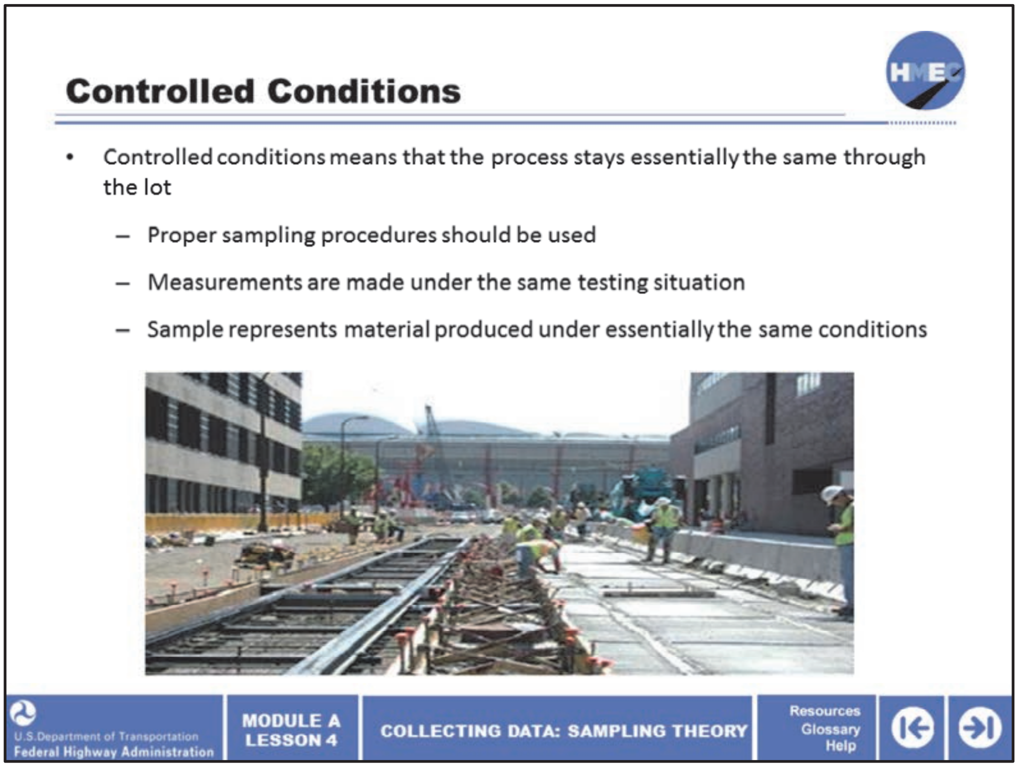
Biased Sampling

[X] CLOSE

- A sampling procedure whereby certain individual measurements have a greater chance of being included than others.
  - Biased sampling occurs when the person taking the sample uses "judgment“ regarding where or when to take the sample.
  - For example, if the people in the photo instructed workers to take a sample two feet from the shoulder of the roadway because the material looked good in that section, that would be a biased sample.
  - Another example would be to a take a sample at 10 o’clock each day.

Biased sampling is a sampling procedure whereby certain individual measurements have a greater chance of being included than others.

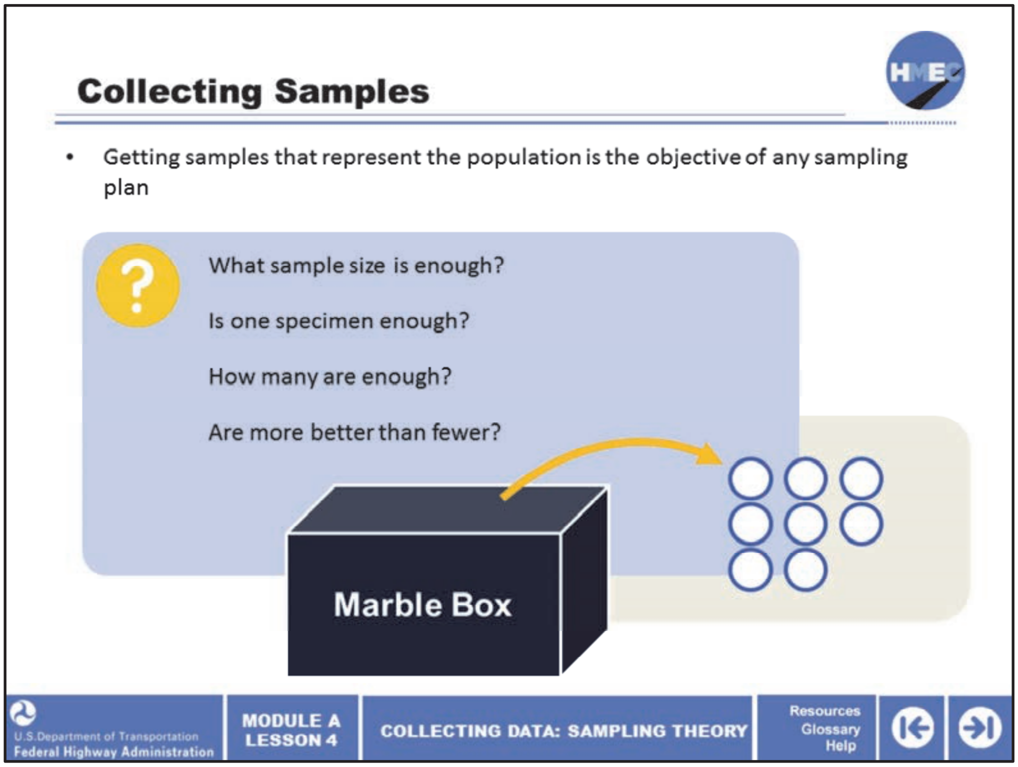
Image description: Workers gathering samples.



The second way to ensure sample validity must also occur under controlled conditions. This is in order to minimize the risk of making a wrong decision about the quality of a material. Controlled conditions means that the process stays essentially the same through the lot. If a concrete bridge deck is being placed and the ingredients of the concrete are changed during the placement, controlled conditions no longer exist. This means the lot is no longer the same as it was previously.

It is required that the set of data to be treated as a single group that represent homogeneous data. For example, measurements or counts made under the same testing situation and that represent construction materials that are produced under essentially the same conditions. Because sample data support so many important decisions, it is essential that proper sampling procedures are used; that is, the agency sampling protocols are followed as closely as possible. It is important to take measurements and perform tests following the agency testing protocols and under the same conditions as closely as possible to ensure the consistency of results is not affected. It is equally important to sample in a way that ensures the materials represent the entire population and are produced under essentially the same conditions.

Image description: Crews lay embedded track for the Hiawatha Light Rail in downtown Minneapolis.



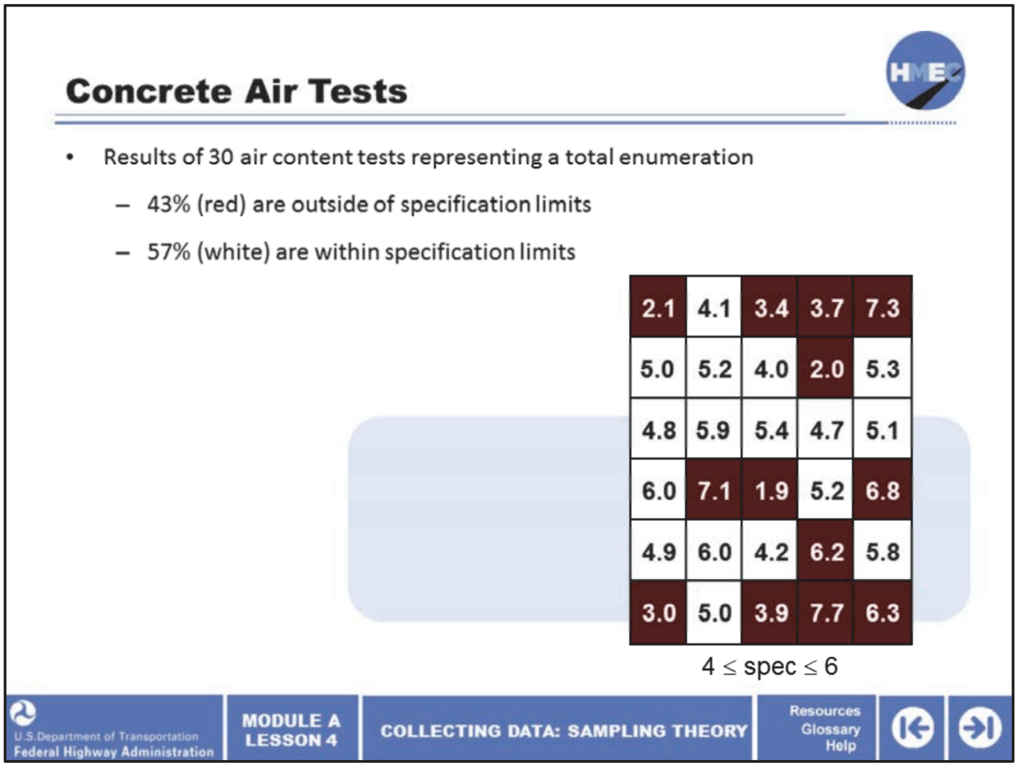
Getting samples that represent the population is the objective of any sampling plan. But what sample size is enough? Is one specimen enough? How many are enough? Are more better than fewer?

It is always the properties of the lot or population that are needed. Sampling is the only effective means for estimating the acceptability of a lot or population. However, how many specimens (1, 2, 3) should comprise the sample to arrive at a sound estimate?

The idea persists that a test on a single sample shows the "true" quality of the material, and that if any test result is not within some limit, there is something wrong with the material, construction, sampling, or testing. Thus, terms such as investigational, check, and referee samples have been in common use to either confirm or document these "failures." Nature dislikes identities; variation is the rule. Therefore, any acceptance or process control sampling must account for variability of materials or construction. Multiple sampling accomplishes this objective.

Recall the example of the marbles used in Lesson 3 where the property you were determining was the color of the material (recall that the specification was 8 out of the 10 marbles needed to be white). It was discussed that the original three samples taken raised questions about the level of risk of determining whether 8 of the 10 marbles were white. What if only one marble had been drawn? What do you think the risks would be?

Image description: A marble box with an arrow pointing to eight marbles sitting outside of the box.

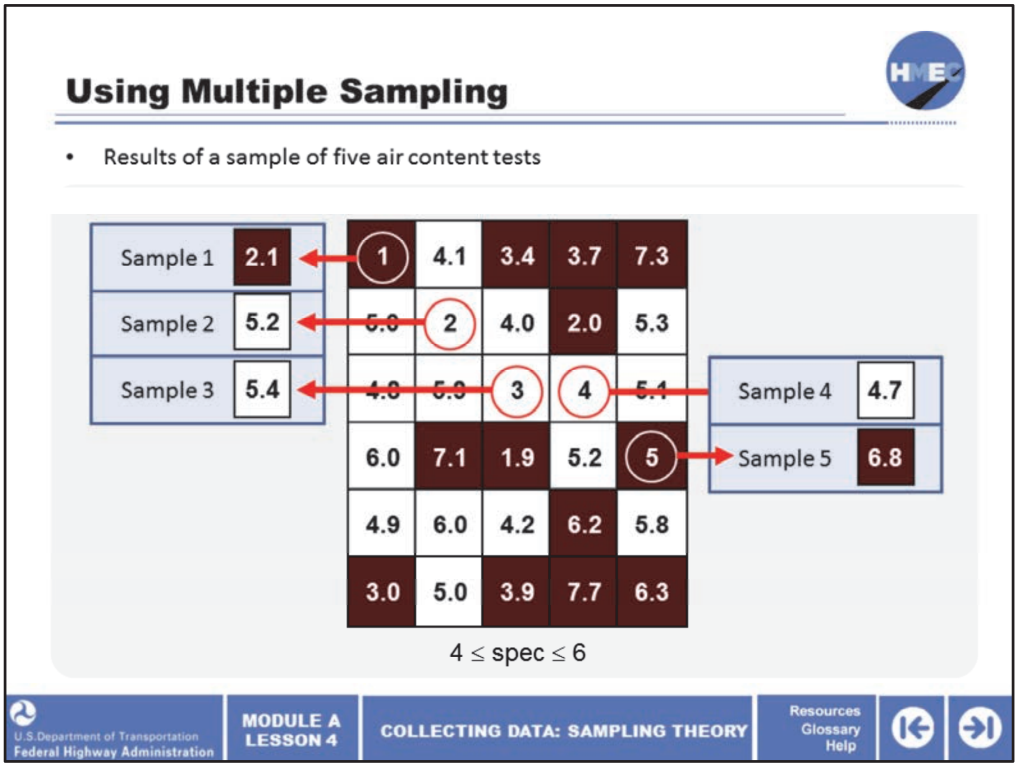


Let's look at an example of sampling a population of concrete air tests. In this example, it is clear that a sample size of one would not give enough data to make an accurate decision about the quality of this product.

This figure illustrates the results of 30 concrete air content tests that we will assume represent a total enumeration (entire population) of all the possible air content results for a lot of structural concrete. If the specification limits for air content for this mix are 5% ± 1%, then any air content result between 4% and 6% is within the specification requirements. A review of the figure indicates that 13 of the tests (43% those in red) are outside specification limits. Conversely, 57% (white) are within specification limits.

Obviously, if a sample size of one made up the entire sample (often referred to as a single sample) you would not be able to determine that 43% of the material was outside the specification limit.

Image description: A figure that illustrates the results of 30 concrete air content tests that we will assume represent a total enumeration (entire population) of all the possible air content results for a lot of structural concrete.



The more specimens that make up a sample, the greater the likelihood that the sample reflects the true properties of the lot.

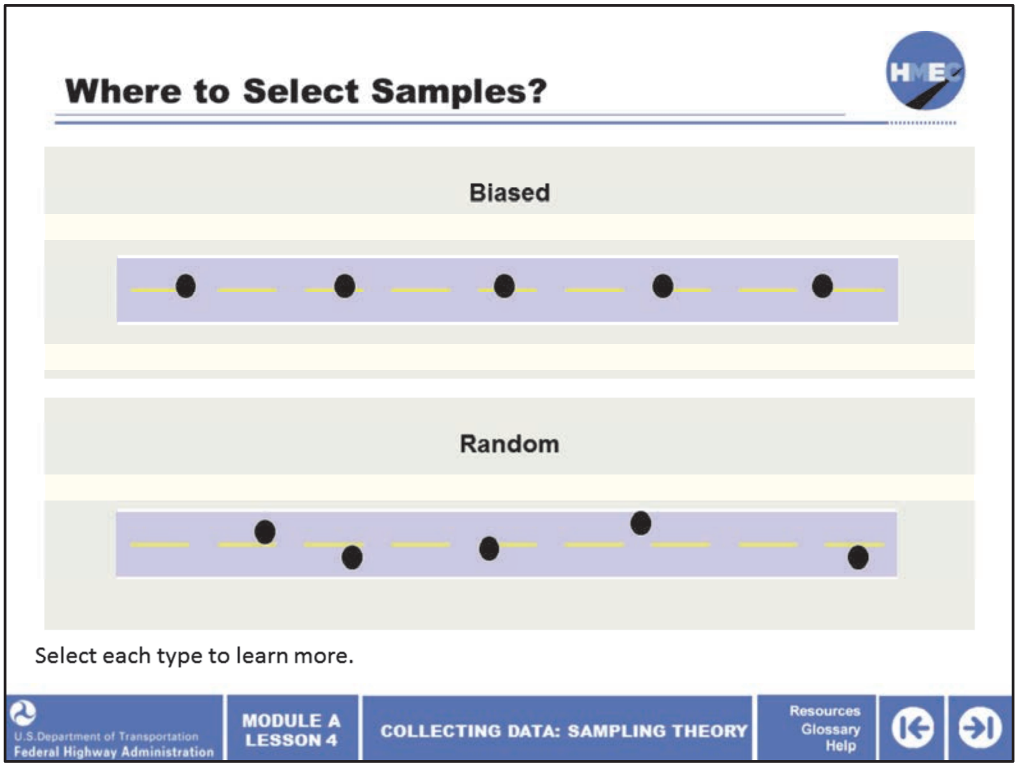
The results of a sample of five air content tests taken from the lot are shown here. If a decision were to be based on only the first test result (2.1%), then the lot would be rejected. If, however, we take a second test and the result is 5.2%, which is within the specification limits, some indecision is introduced since one test meets the requirement while the other does not. Increasing the sample size presents a clearer picture regarding how the air content is varying within the lot. This variation cannot be shown on the basis of a sample of size one, but can be estimated as the number of tests (specimens) increases (i.e., multiple sampling). We consider the question of how many specimens per sample should be taken in subsequent chapters.

Conceptually, 43% of the population was outside the specification limits (4–6). In this example, our sample size of five shows two out of five to be outside the specification limits (40%). That is a very close estimate. Obviously, a sample will not always estimate a population property this well, but it illustrates the point that multiple sampling improves our chances of making a correct decision. Note: don't confuse this with PWL, we're not there yet.

Image description: A figure showing the results of a sample of five air content tests taken from the lot.

Image description: Example highlighting results.

Image description: Example highlighting results.



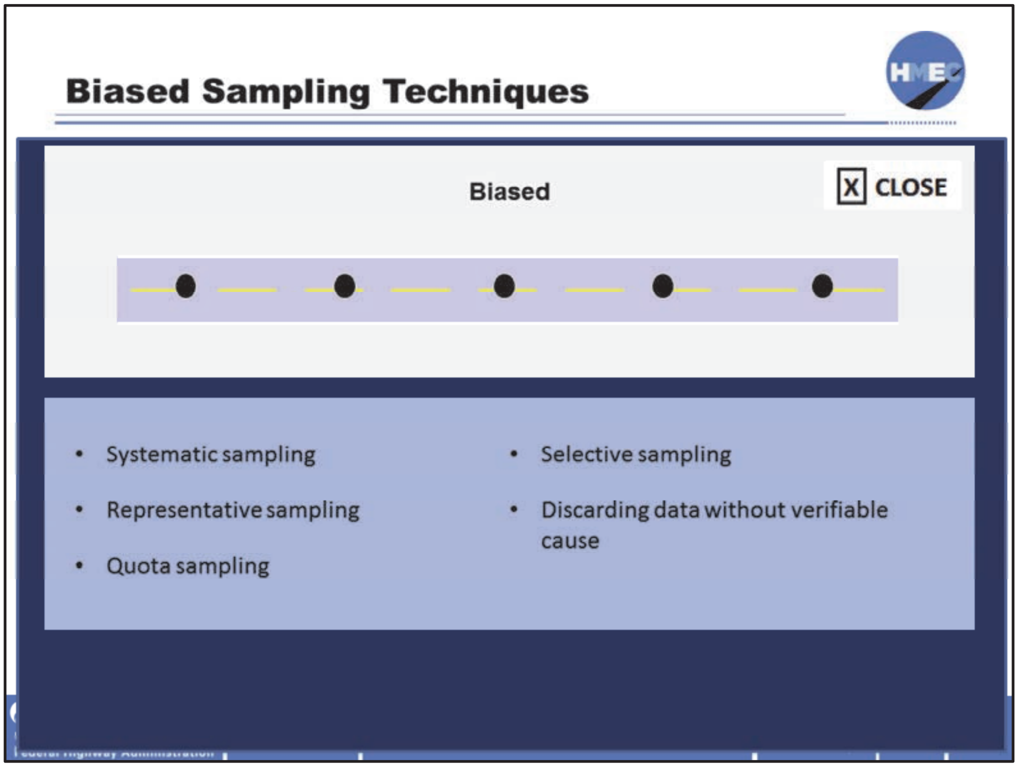
So where do we select samples? The selection of the sampling locations within the lot must be entirely random. "Random" does not mean "haphazard"— it means the sample is selected without bias.

In practice it may be difficult to train technicians who have been accustomed to inspection to select samples without regard to quality. Their tendency is to make sure that defective materials are represented in the sample, or possibly that only acceptable materials are included in the sample, thus subconsciously biasing the sample.

Select each type to learn more.

Image description: Example of a biased lot.

Image description: Example of a random lot.

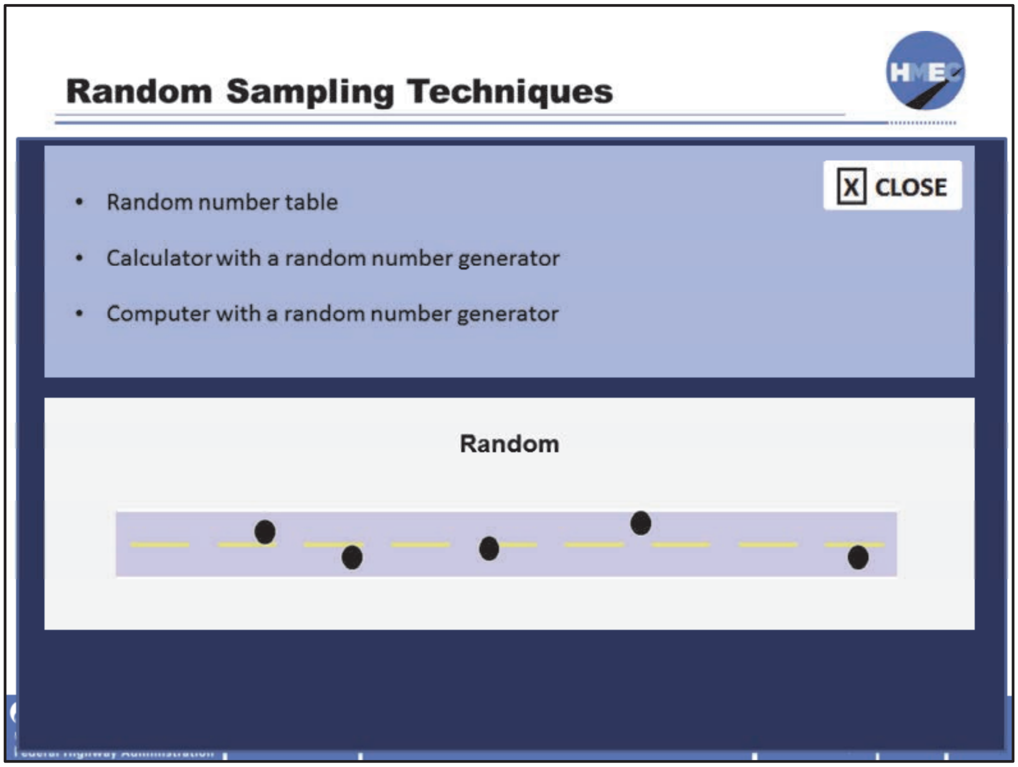


Biased sampling is any sampling that is not random.

Biased sampling techniques include:

- Systematic sampling such as sampling only down the center line;
- Representative sampling;
- Quota sampling;
- Selective sampling; and
- Discarding data without verifiable cause.

Image description: Example of a biased lot.



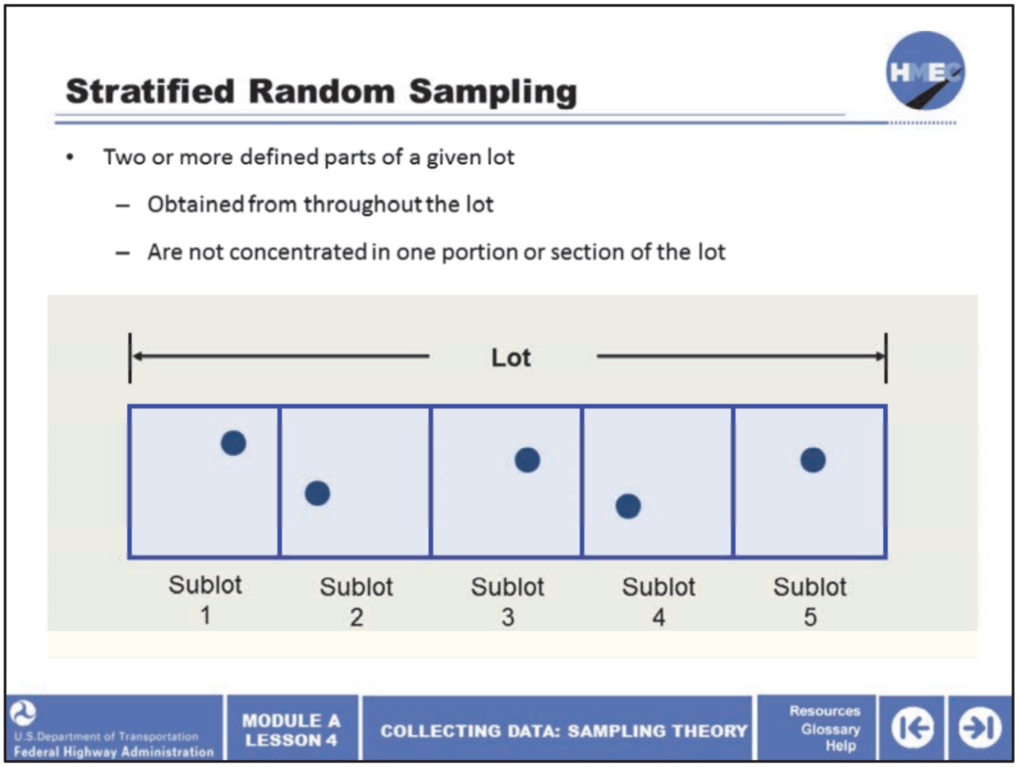
Random sampling can be generated by several methods.

The three most common methods for generating random numbers are:

- Random number table;
- Calculator with a random number generator. There are many types of calculators with random number generators. (You should be familiar with the calculator you use to assure it is, in fact, generating random numbers); and
- Computer with a random number generator.

In all cases, you should know the size of lot and number of specimens. This information comes from specifications and contract documents.

Image description:

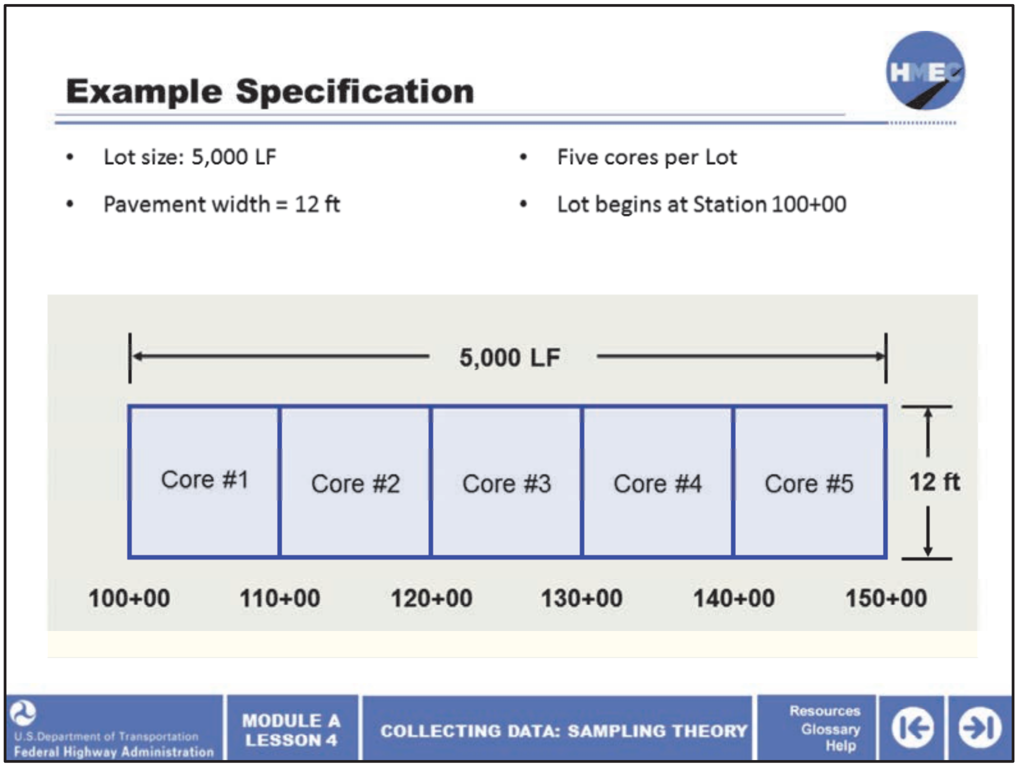


Random sampling ensures that each portion of a lot has the same chance of being selected for the sample.

Stratified random sampling additionally involves the selection of two or more defined parts of a given lot. Stratified sampling is used to ensure that the specimens for the sample are obtained from throughout the lot, and are not concentrated in one portion or section of the lot. For instance, if random sampling is used it is possible, although not likely, that all of the samples, or cores, could be selected within the first half of the lot. That would be random and would tell you a great deal about the first half, but nothing about the second half. Statistically this would be correct but engineering-wise would not provide a comfortable feeling about the quality of the entire lot.

The lot can be stratified into a number of sublots equal to the sample size to be selected from the lot. One core is then randomly selected from within each sublot. By using this procedure, each portion of the lot has the same chance of being selected while, at the same time, ensuring that the sampling is spread out over the entire lot instead of being bunched in one area. This is the typical procedure used.

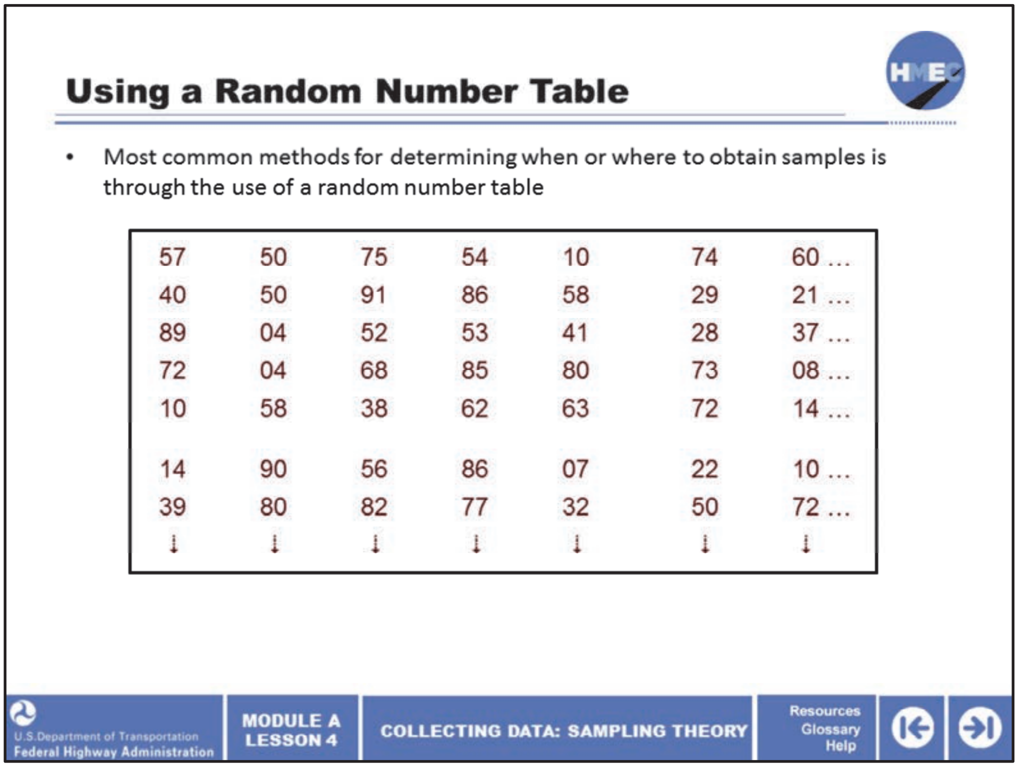
Image description: Example of a lot, stratified into a number of sublots.



Suppose one is to sample an asphalt mixture from the roadway to obtain cores for density determination. The specification states that the lot size shall be 5,000 linear feet (LF) of pavement, and that the sample size is five cores per lot, and that stratified random sampling will be done. If we assume that the pavement width is 12 feet and the lot begins at station 100+00, then we can use a random number table to select the sampling locations of the sublots.

The example in the next slide illustrates how to determine the locations for sampling using a random numbers table for five sublots.

Image description: Example of a random number table.



57	50	75	54	10	74	60 ...
40	50	91	86	58	29	21 ...
89	04	52	53	41	28	37 ...
72	04	68	85	80	73	08 ...
10	58	38	62	63	72	14 ...
14	90	56	86	07	22	10 ...
39	80	82	77	32	50	72 ...
↓	↓	↓	↓	↓	↓	↓

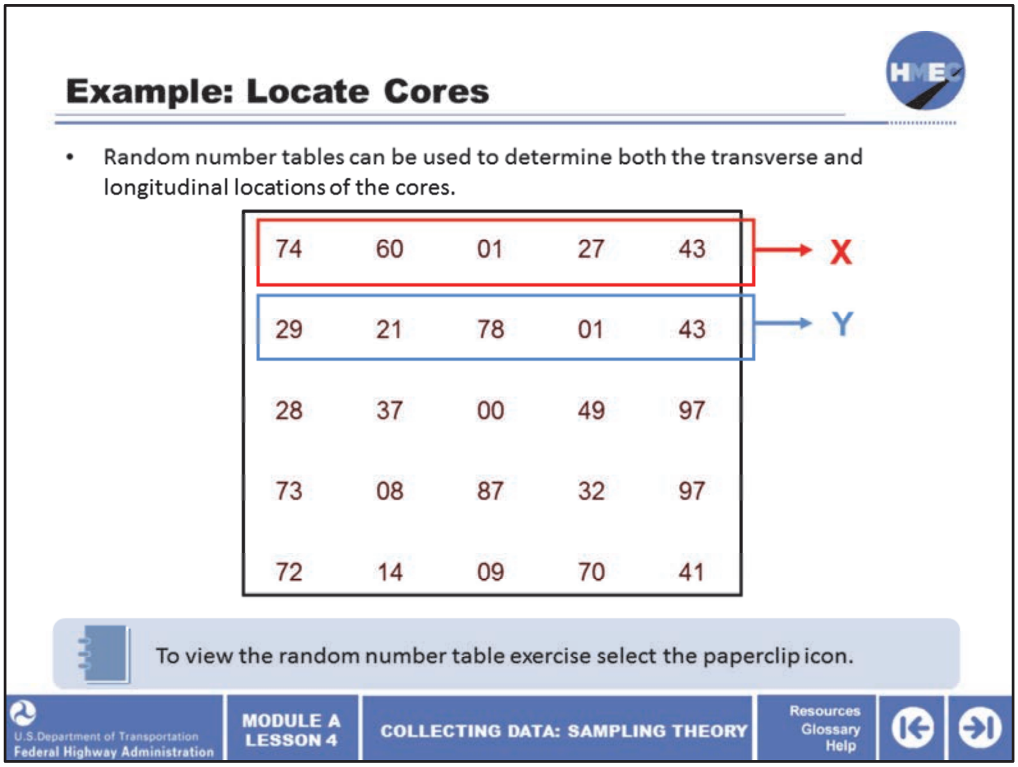
One of the most common methods for determining when or where to obtain samples is through the use of a random number table. A random number table is a collection of random digits. Random number tables come in many forms—some are short, some are long, some grouped by pairs of digits, some with as many as five digits per group. When using a random number table, the key is that bias must be avoided.

A brief example of a random number table is shown. This table is used in this lesson because it is simple. A more through procedure is the use of ASTM D 3665. These random numbers in the table in this lesson are presented in pairs of digits and, for the methods that we will consider, can be thought of as two-place decimal fractions. For example, the random number 57 in the table would be used as 0.57. Sampling locations can be determined on the basis of time, tonnage, volume, distance, area, etc. The only other necessary information is the size of the lot to be sampled and the number of samples, or the sample size.

When selecting a group of random numbers, one can enter the table at any point (but never at the same point twice) and select the required amount of numbers. The numbers can be selected by columns or rows, by going left or right, up or down, selecting alternate numbers, or any other pattern desired, but this decision should be made before starting the process of selecting the numbers. Otherwise the numbers may be biased.

To open the random number table exercise select the paperclip icon.

Image description: Example of a random number table.

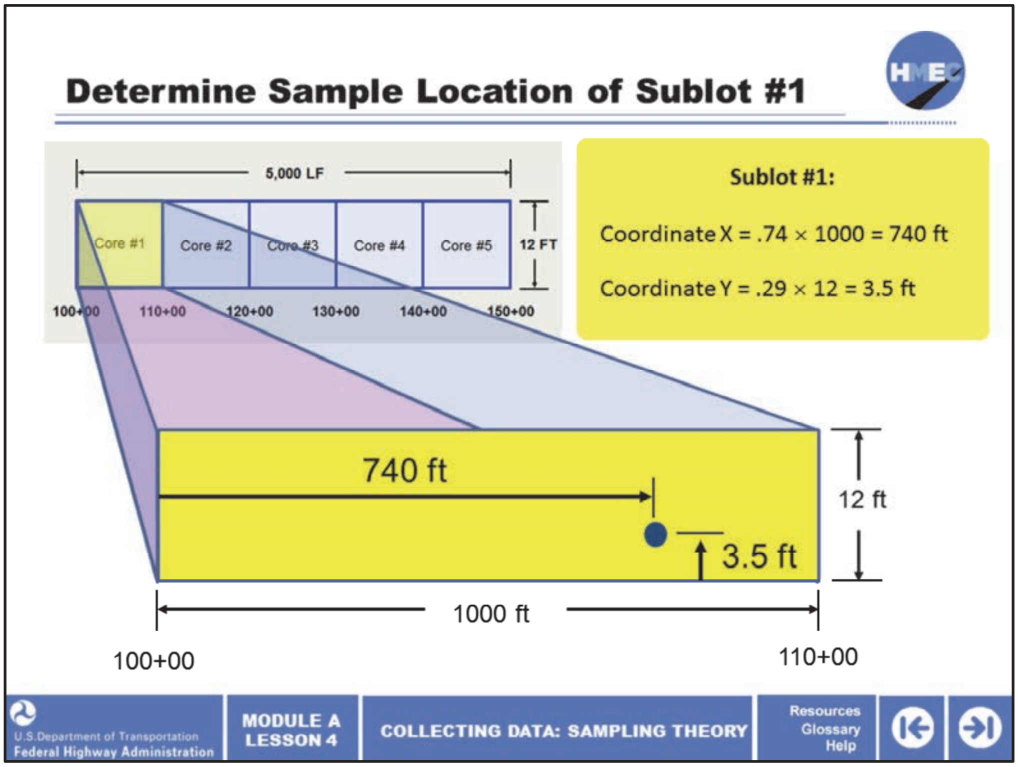


The random number table can be used to determine both the transverse and longitudinal locations for the cores. Two sets (columns, rows, etc.) of random numbers are selected: one for the transverse position, the other for the longitudinal position. The specification states that a set of five random numbers (the number of cores) for the longitudinal (X) position and five random numbers for the transverse (Y) position of the sample be chosen.

This abbreviated table uses the second block of numbers from the table in the attached exercise and in itself was chosen randomly. (The block chosen, where to start, and what direction to use, that is, horizontally or vertically should all be selected randomly.)

The way quality control and acceptance activities use of the random number table is explained next. To view the random number table exercise select the paperclip icon. We will complete this exercise later in the lesson.

Image description: Example of a random number table highlighting the X and Y positions.



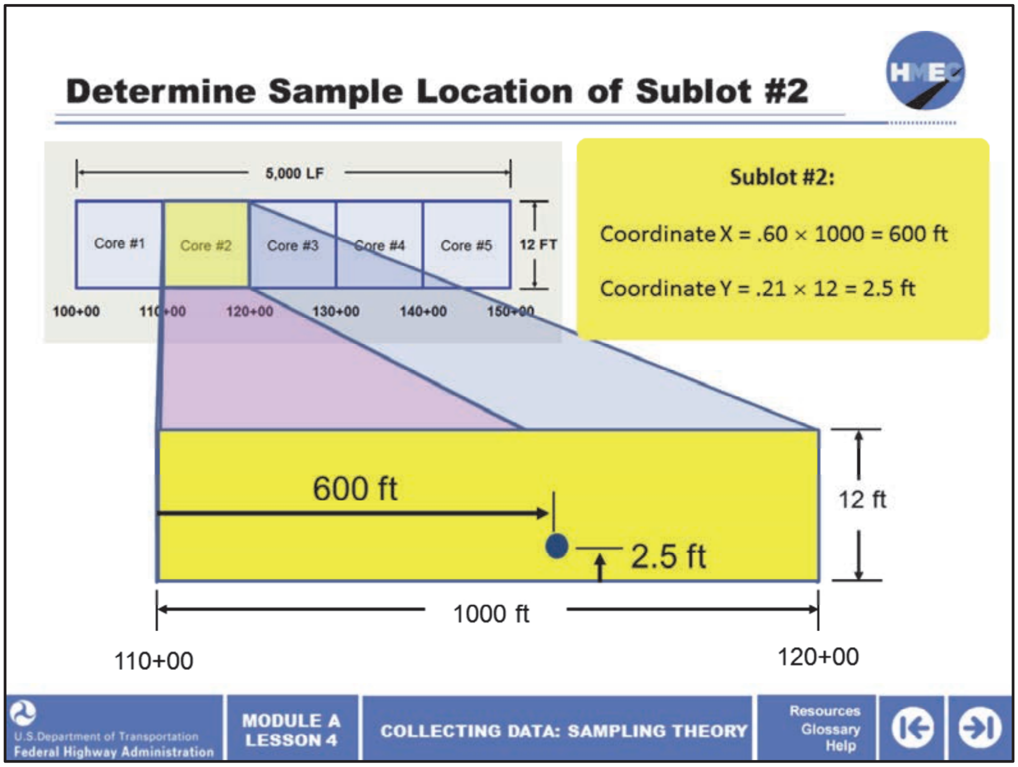
Once you have selected the sets of random numbers, apply basic multiplication to derive longitudinal and transverse coordinates for each core in the sublot. These X and Y random numbers are multiplied by the sublot length and width, respectively, as shown in the example:

Sublot #1 (start at station 100+00 for the longitudinal reference point and the bottom edge of pavement for the vertical reference point. The edge of pavement to use as the reference point is arbitrary, but once selected has to be used throughout the lot.)

- The first number in the first row, which was randomly selected as the X coordinate is used for the horizontal coordinate in the first sublot and is multiplied by the sublot length, 1,000 feet. Thus, the longitudinal coordinate for sublot #1 is X = 0.74 x 1,000 = 740 ft.
- Likewise, the first number in the second row, which was randomly selected as the Y coordinate is used for the vertical coordinate in the first sublot and is multiplied by the sublot width, 12 feet. Thus, the vertical coordinate for sublot #1 is Y = 0.29 x 12 = 3.5 ft measured from the bottom edge of pavement.

This same procedure if followed for sublot #2 shown on the next slide.

Image description: Example of core samples, highlighting sublot 1.

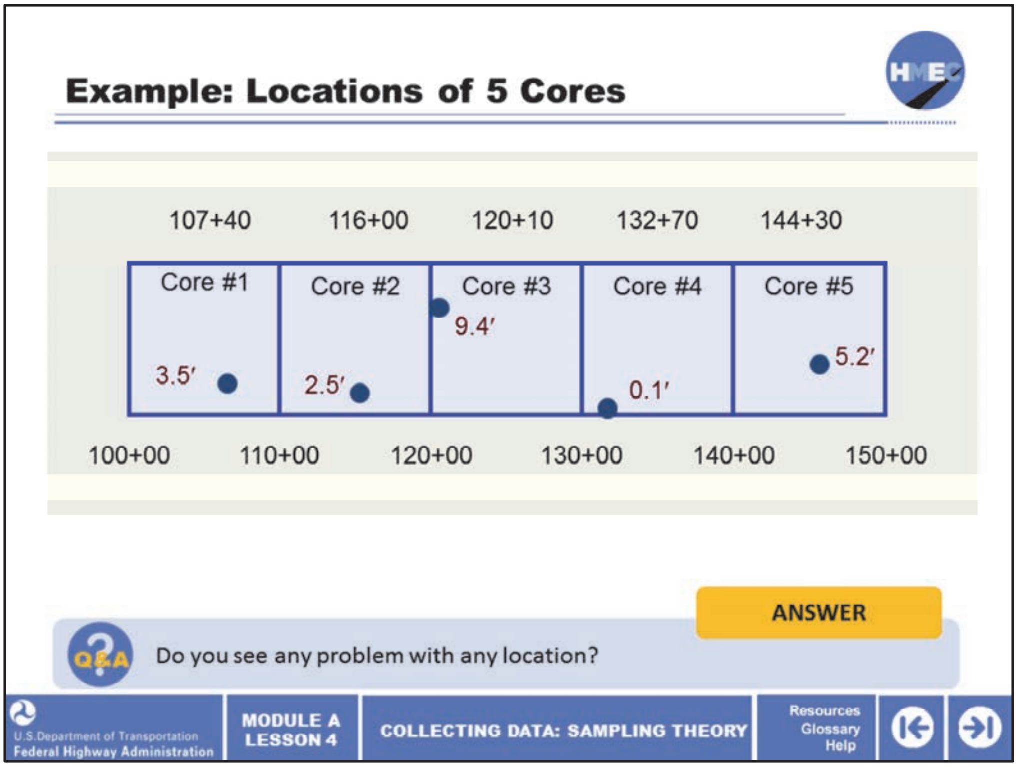


Sublot #2 (start at station 110+00)

- The second number in the first row, which was randomly selected as the X coordinate is used for the horizontal coordinate in the second sublot and is multiplied by the sublot length, 1,000 feet. Thus, the longitudinal coordinate for sublot #2 is X = 0.60 x 1,000 = 600 ft and is added to Station 110+00 for the longitudinal reference point.
- Likewise, the second number in the second row, which was randomly selected as the Y coordinate is used for the vertical coordinate in the second sublot and is multiplied by the sublot width, 12 feet. Thus, the vertical coordinate for sublot #2 is Y = 0.21 x 12 = 2.5 ft from the bottom edge of the pavement.

This same procedure is followed for sublot #3, sublot #4, and sublot #5 and the results of all five locations are shown on the next slide.

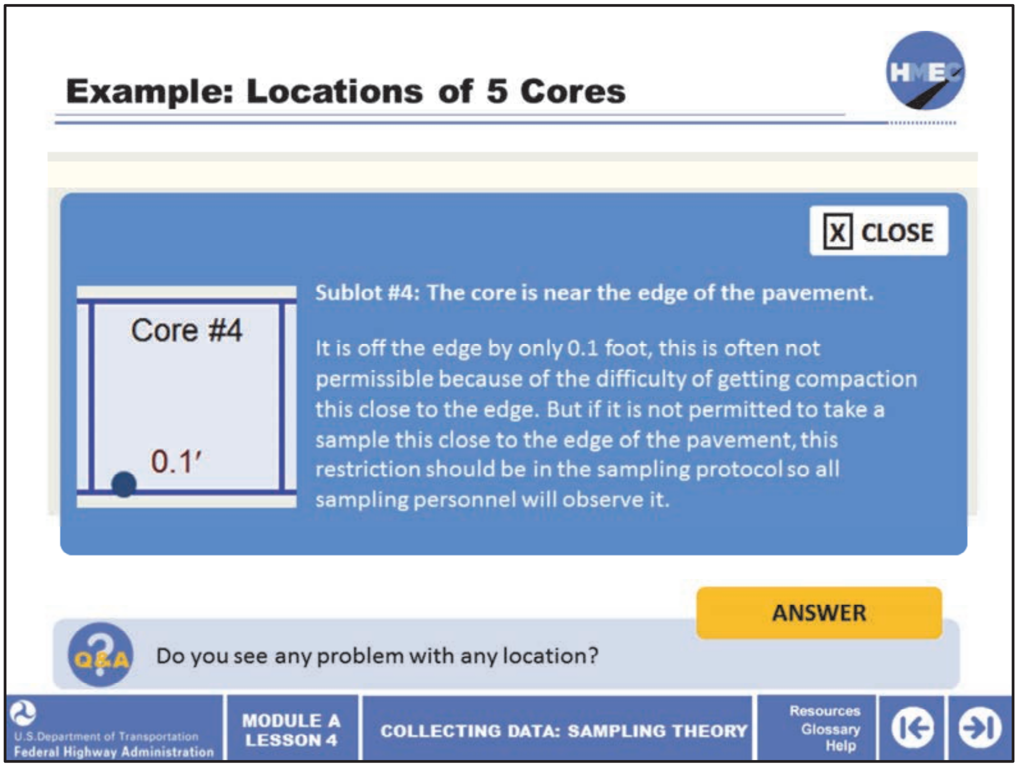
Image description: Example of core samples, highlighting sublot 2.



Here are the locations of the five cores.

Select the box to answer the question, Do you see any problem with any location?

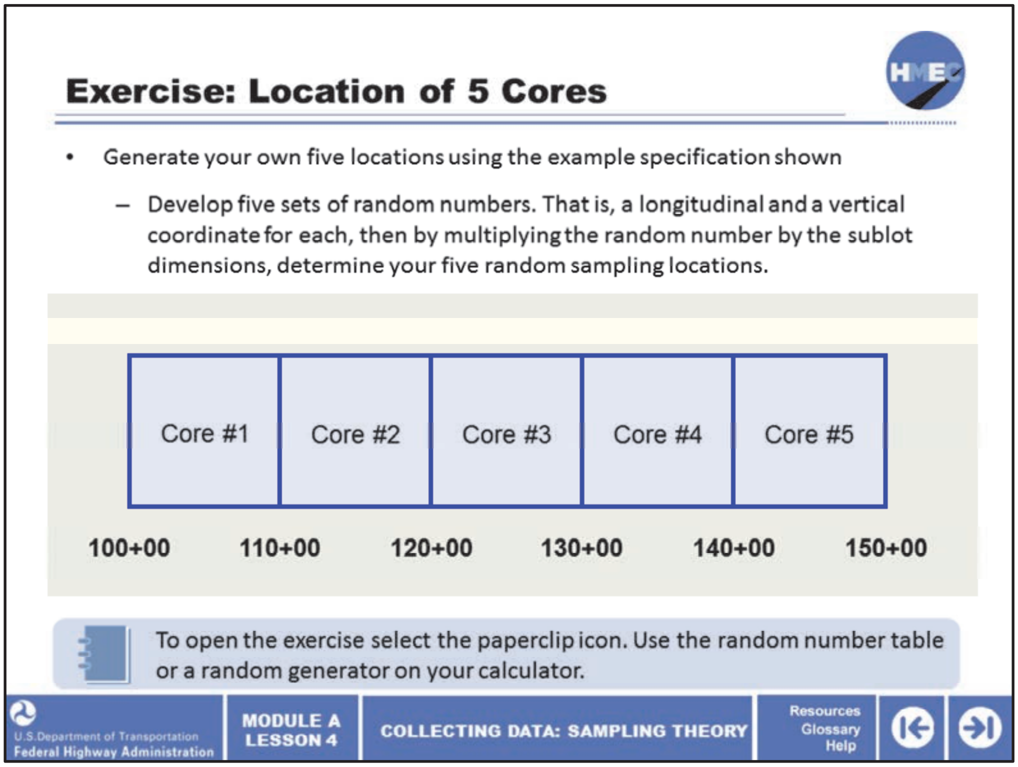
Image description: Example of 5 core samples.



Look at sublot #4. The core is near the edge of the pavement. It is off the edge by only 0.1 foot. This is often not permissible because of the difficulty of getting compaction this close to the edge. But if it is not permitted to take a sample this close to the edge of the pavement, this restriction should be in the sampling protocol so all sampling personnel will observe it.

If this situation occurs, it is permissible statistically to take the next set, that is X and Y values to determine the sample location. That would require the next set of random numbers to be used for sublot #5.

Image description: Core 4 example.

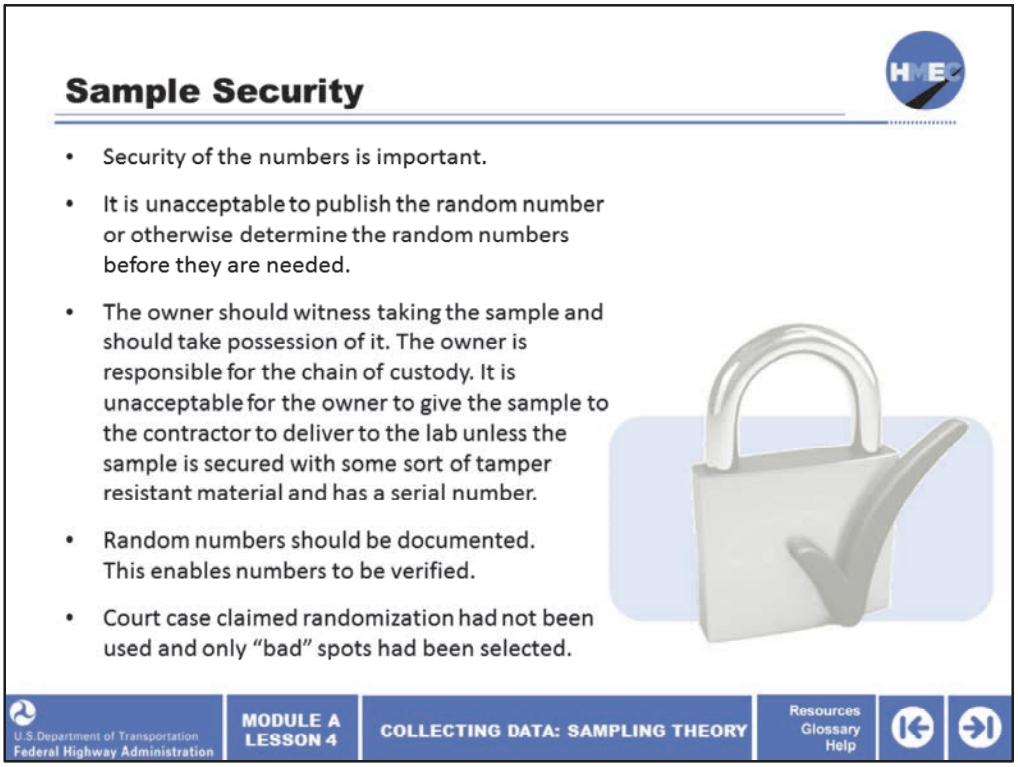


Now we'll give you a chance to generate your own five locations using the example specification shown. Using the random number table in the attached exercise, or a random generator on your calculator, iPhone, or computer, develop five sets of random numbers. That is, a longitudinal and a vertical coordinate for each. Then by multiplying the random number by the sublot dimensions, determine your five random sampling locations.

Take 15 minutes to determine five random sampling locations before moving on with this lesson. Write the answers in your exercise. You will be asked to share your results later on.

To open the exercise select the paperclip icon. Use the random number table or a random generator on your calculator.

Image description: Example of 5 core samples.

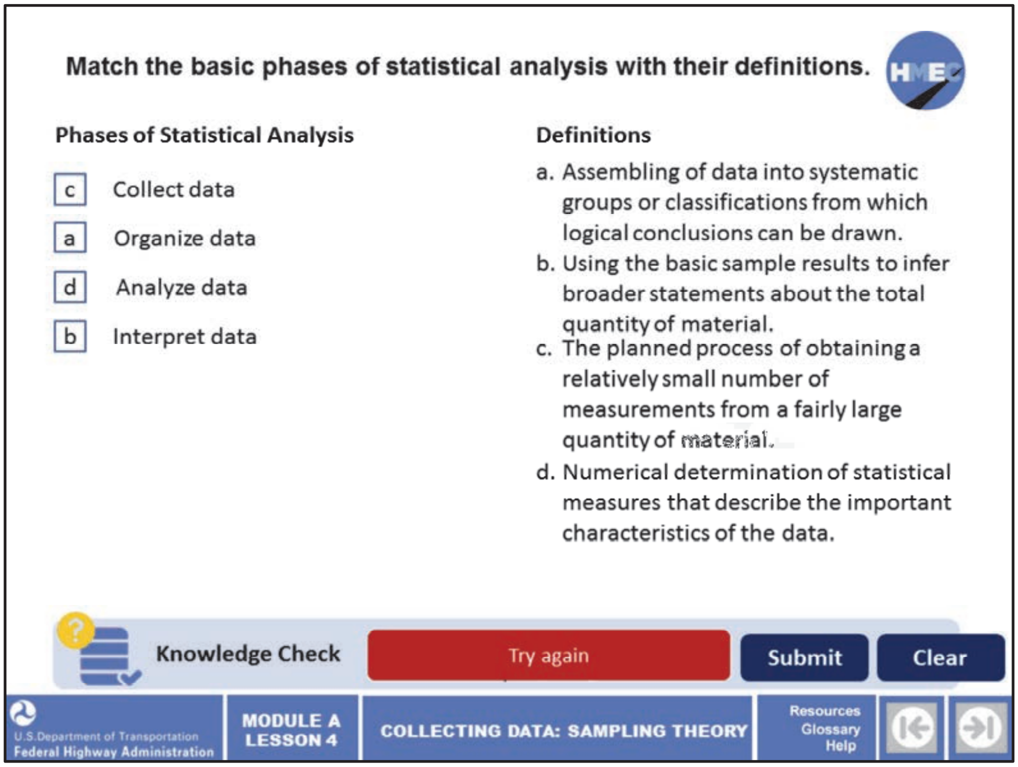


The random numbers can be compromised if they are known prior to being needed, so the security of the numbers once they are selected is important. It is unacceptable to publish the random number or otherwise determine the random numbers before they are needed.

The owner should witness taking the sample and should take possession of it. The owner is responsible for the chain of custody. It is unacceptable for the owner to give the sample to the contractor to deliver to the lab unless the sample is secured with some sort of tamper resistant material and has a serial number.

The tonnage, or location and time, for example, determined by the random numbers need to be documented. Likewise, when the random numbers and resulting tonnage, or location and time that are used, should be recorded so that verification of the information is possible. This may seem to be a trivial matter, but in the case of a dispute or claim, it can be one of the deciding points. In fact, a court case (contractor claim) in one State hinged on the agency proving that the samples were selected randomly. The contractor argued that the sample locations were biased toward "bad" spots in the pavement. The inspector's log book had to be produced showing the determination of the random locations to refute this claim.

Image description: lock and check mark.



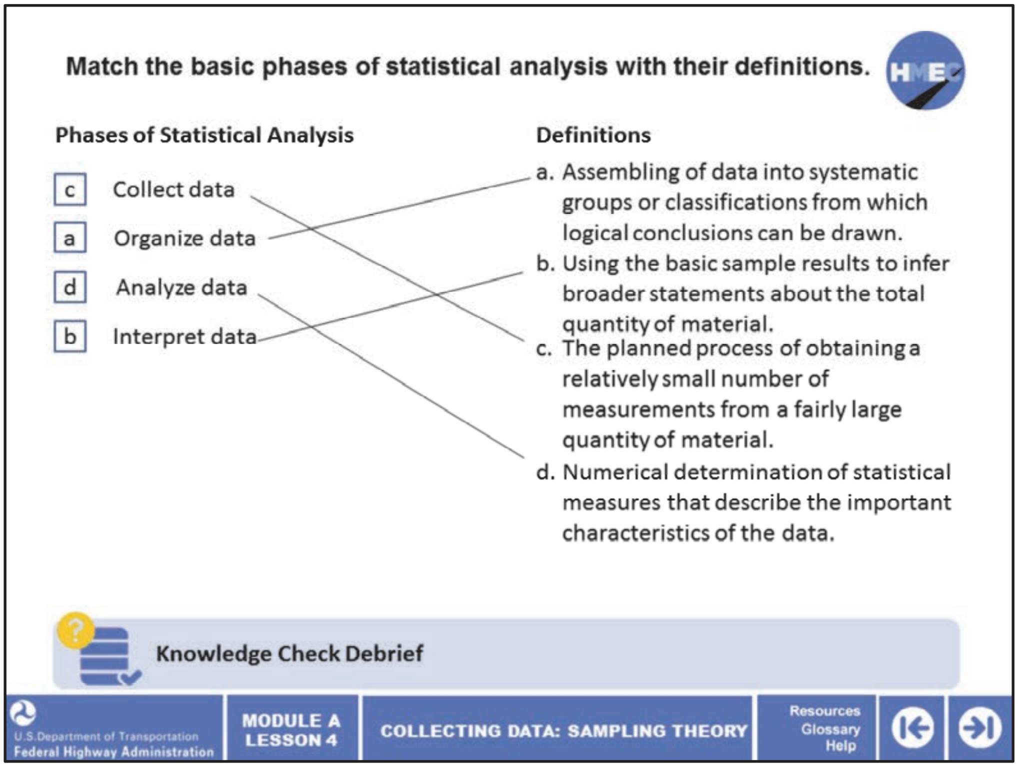
Match the basic phases of statistical analysis with their definitions.

Phases of statistical analysis:

- Collect data;
- Organize data;
- Analyze data; and
- Interpret data.

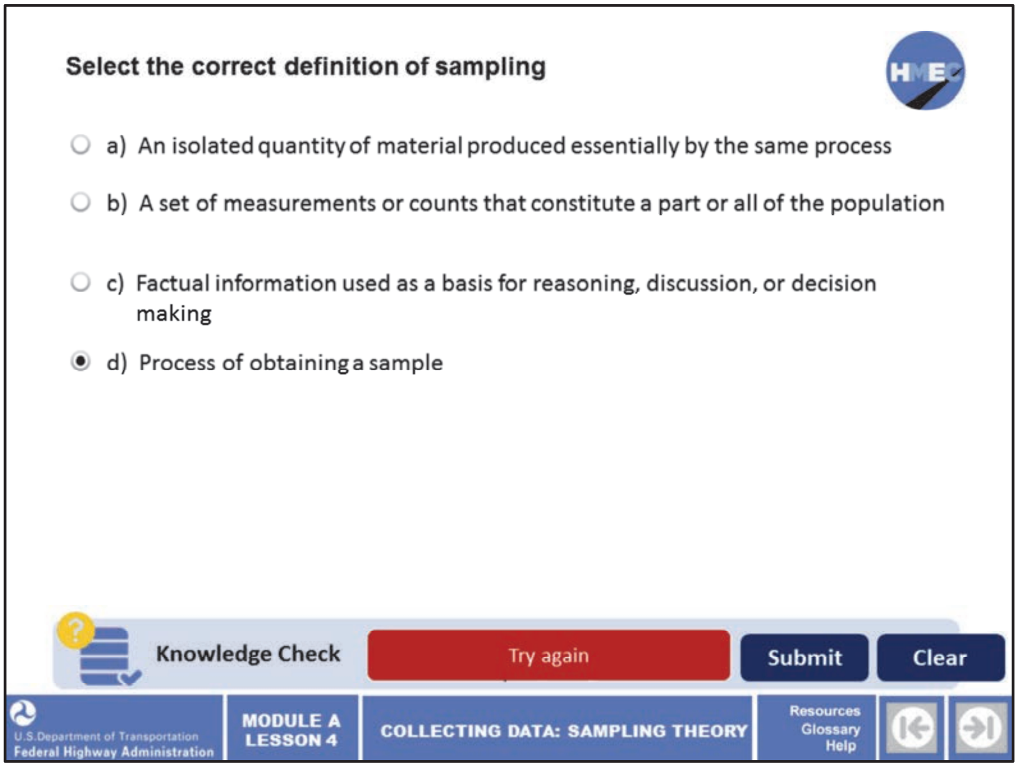
Definitions:

a) Assembling of data into systematic groups or classifications from which logical conclusions can be drawn;
b) Using the basic sample results to infer broader statements about the total quantity of material;
c) The planned process of obtaining a relatively small number of measurements from a fairly large quantity of material; and
d) Numerical determination of statistical measures that describe the important characteristics of the data.



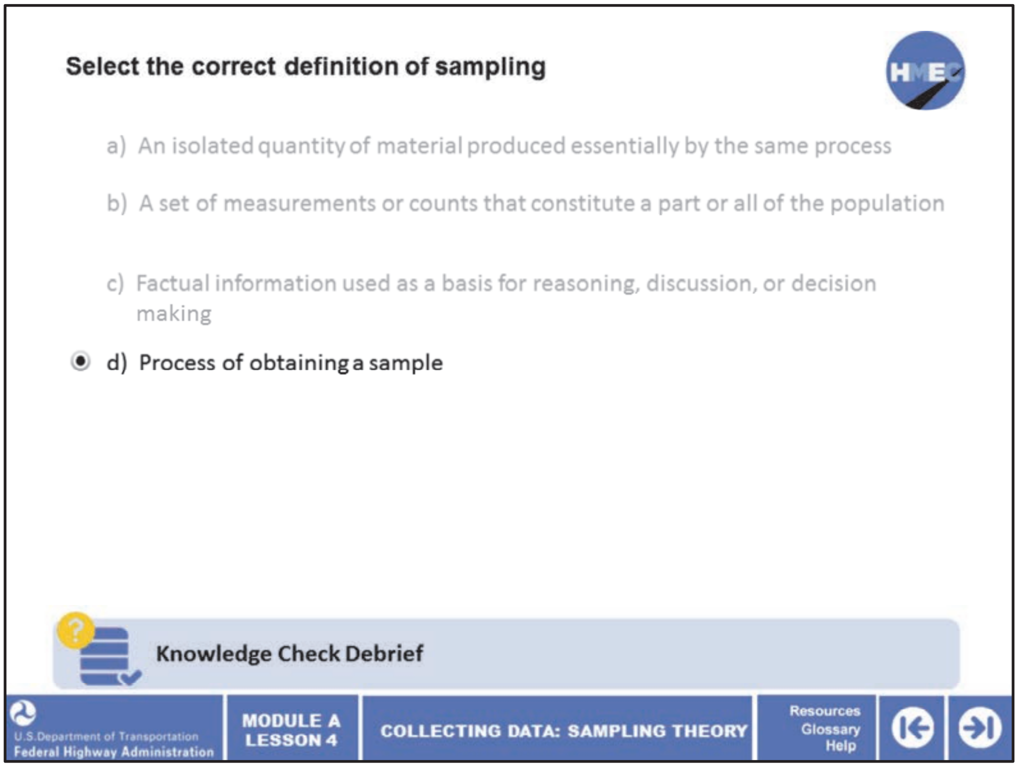
The correct answers are:

• Collect Data is c) The planned process of obtaining a relatively small number of measurements from a fairly large quantity of material;
• Organize the Data is a) Assembling of data into systematic groups or classifications from which logical conclusions can be drawn;
• Analyze the Data is d) Numerical determination of statistical measures that describe the important characteristics of the data; and
Interpret the Data is b) Using the basic sample results to infer broader statements about the total quantity of material.

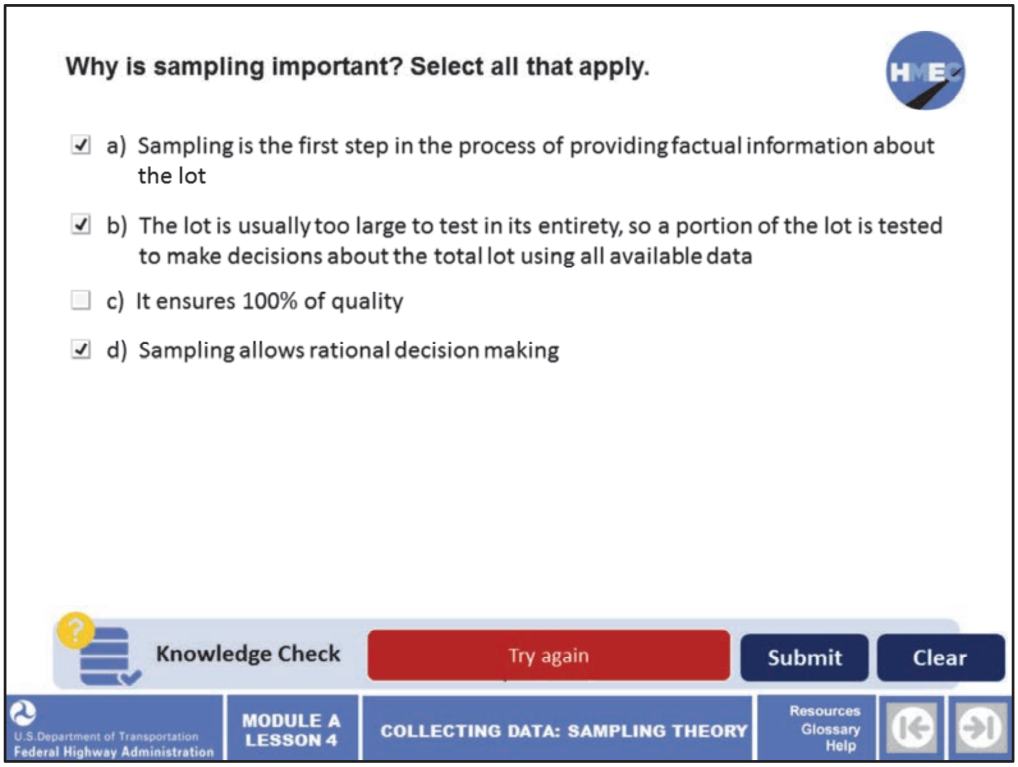


Select the correct definition of sampling:

a) An isolated quantity of material produced essentially by the same process;
b) A set of measurements or counts that constitute a part or all of the population;
c) Factual information used as a basis for reasoning, discussion, or decision making; or
d) Process of obtaining a sample.

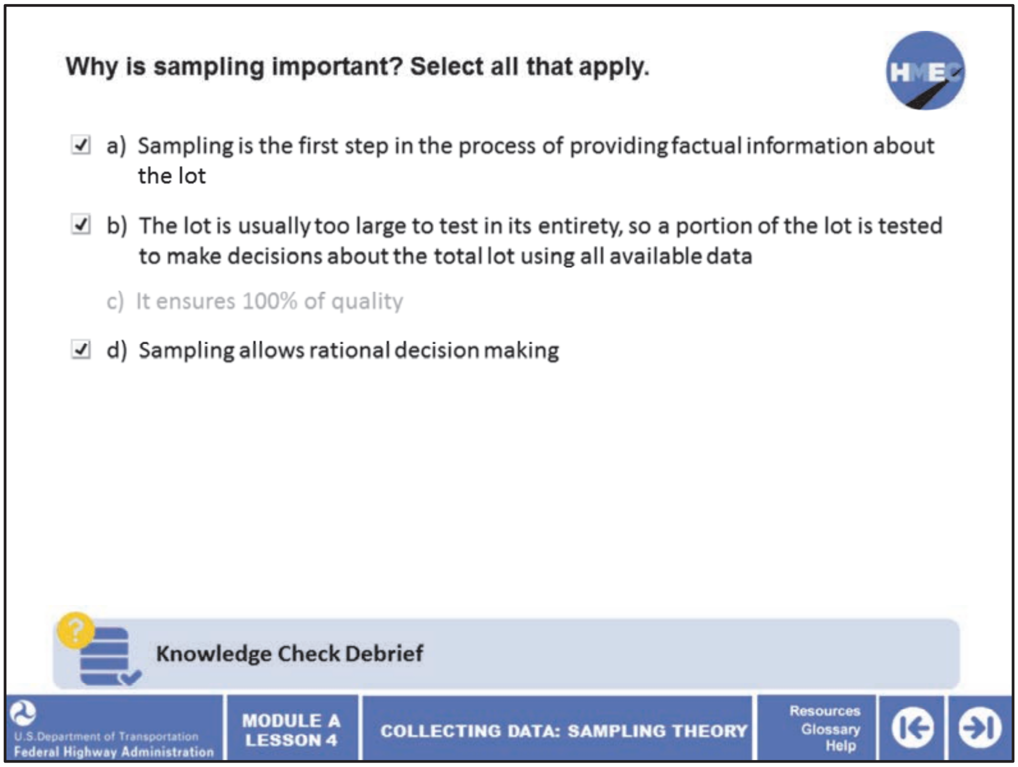


The correct answer is d) Process of obtaining a sample. Sampling is a process. It should not be confused with the "sample," which become a set on measurements or counts and provide factual information for decision making.



Why is sampling important? Select all that apply.

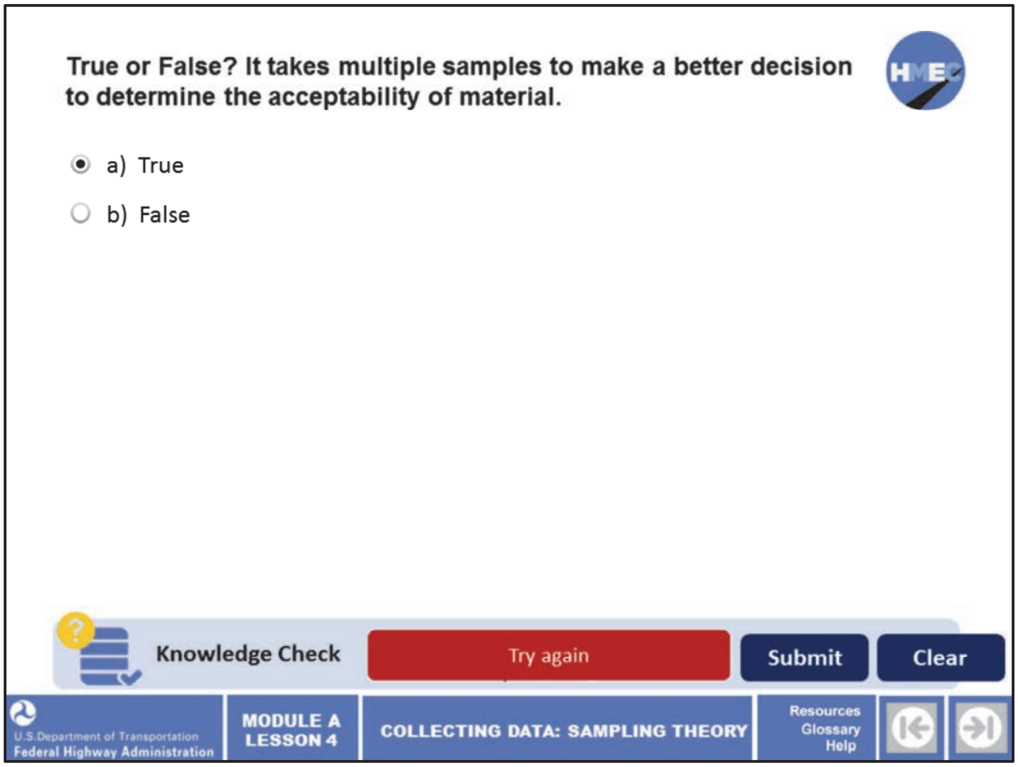
a) Sampling is the first step in the process of providing factual information about the lot;
b) The lot is usually too large to test in its entirety, so a portion of the lot is tested to make decisions about the total lot using all available data;
c) It ensures 100% of quality; and
d) Sampling allows rational decision making.



The correct answers are:

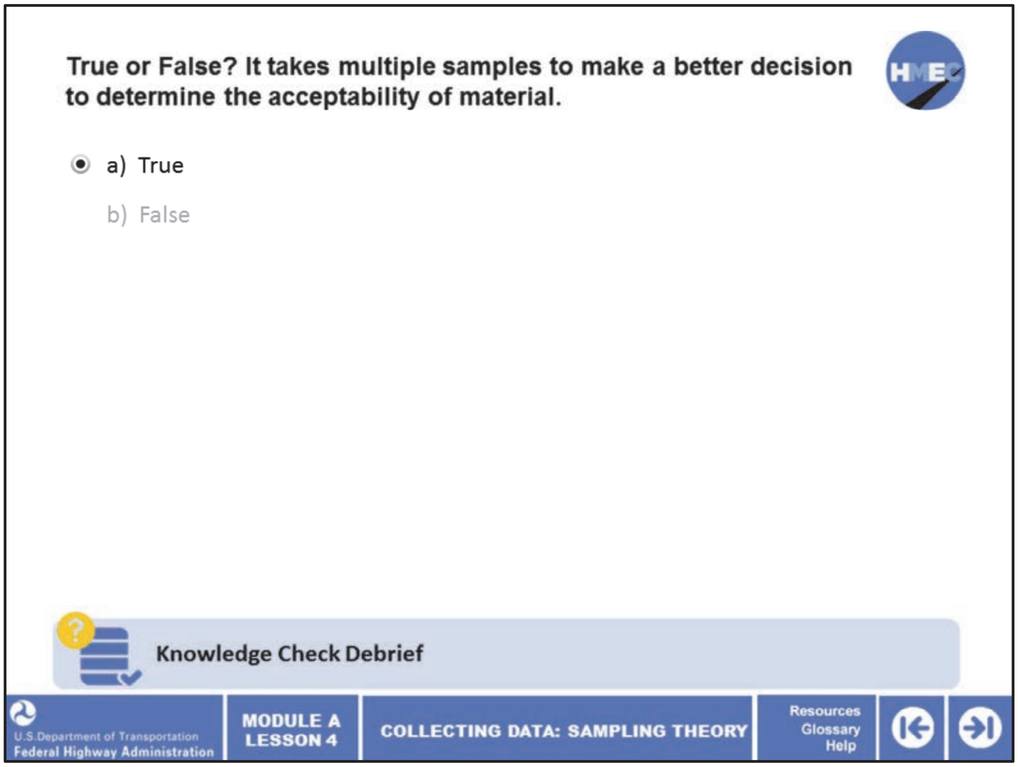
a) Sampling is the first step in the process of providing factual information about the lot;
b) The lot is usually too large to test in its entirety, so a portion of the lot is tested to make decisions about the total lot using all available data; and
d) Sampling allows rational decision making.

Sampling is the first step in the process of providing factual information about the lot. Sampling the entire lot, that is, “complete enumeration” is not feasible. Thus, it is the sampling that provides the information necessary for rational decision making. If the sampling is done incorrectly, it can provide information that can lead to incorrect decisions.

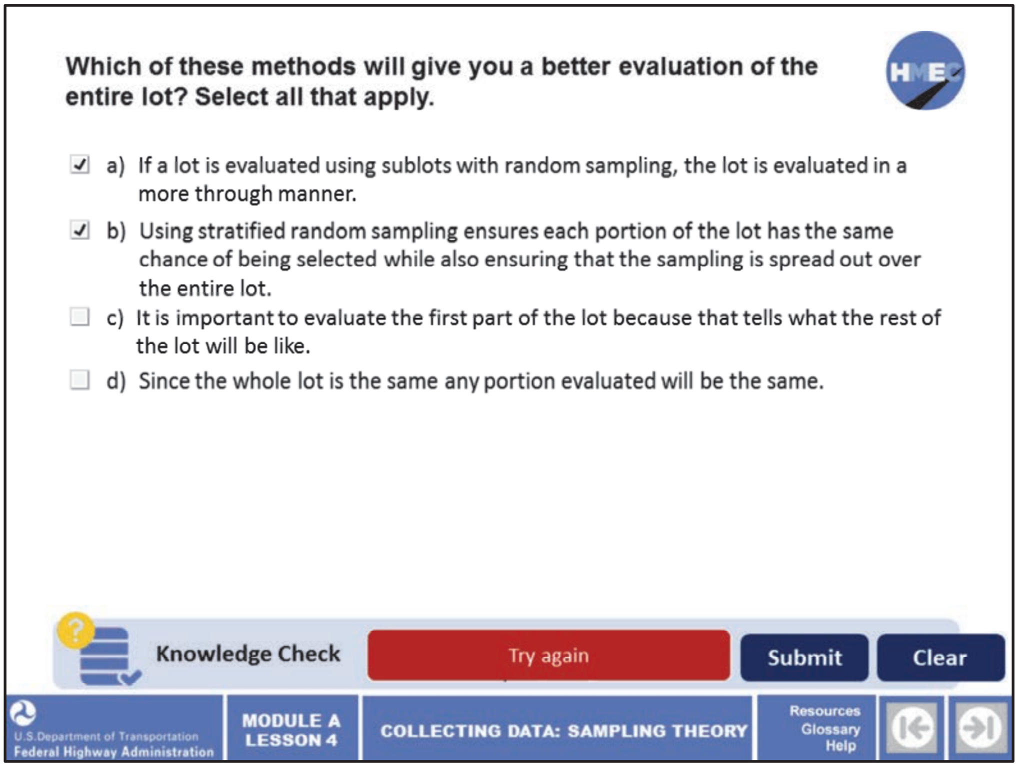


True or false? It takes multiple samples to make a better decision to determine the acceptability of material.

a) True; or
b) False.

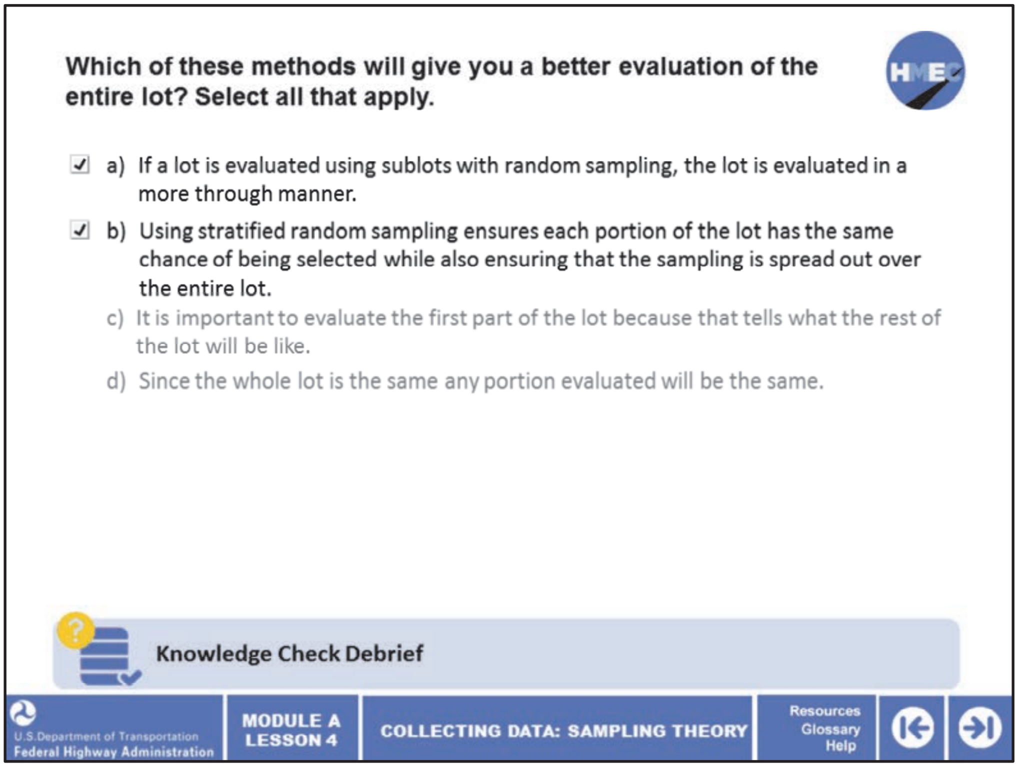


The correct answer is a) True. As observed in the marble example in Lesson 3, taking only a sample size of three can lead to high risks to the agency. It takes multiple samples, that is a sample size greater than one, to provide a measure of variability and make a better decision as the acceptability of material.



Which of these methods will give you a better evaluation of the entire lot? Select all that apply.

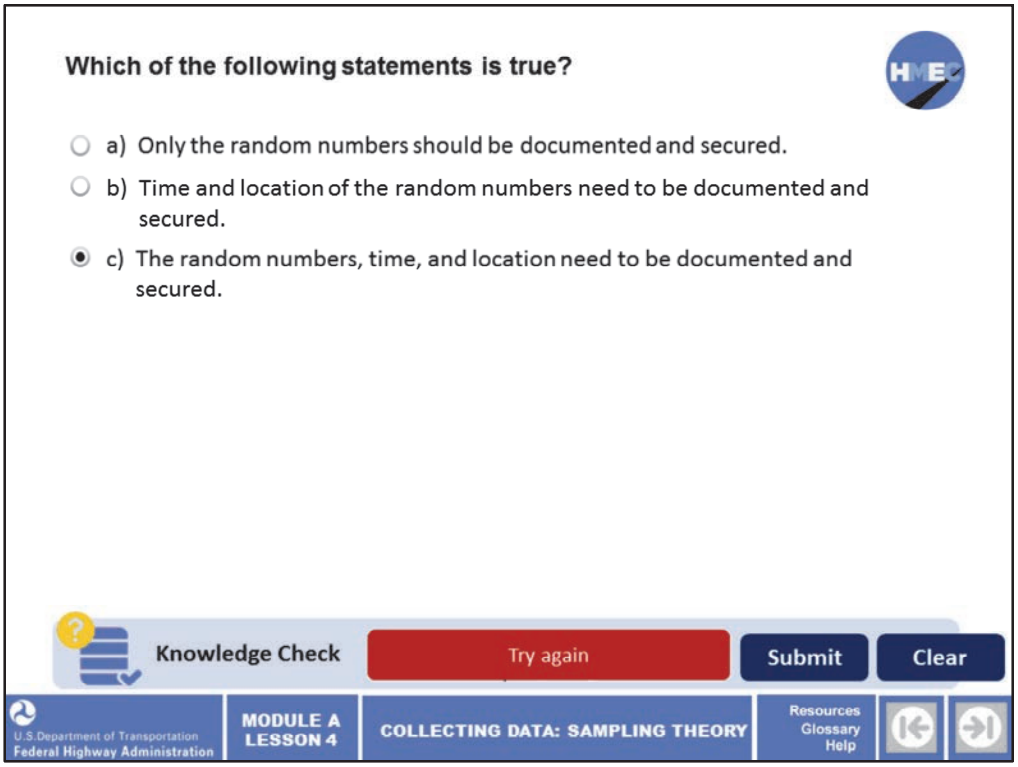
a) If a lot is evaluated using sublots with random sampling, the lot is evaluated in a more through manner;
b) Using stratified random sampling ensures each portion of the lot has the same chance of being selected while also ensuring that the sampling is spread out over the entire lot;
c) It is important to evaluate the first part of the lot because that tells what the rest of the lot will be like; and
d) Since the whole lot is the same any portion evaluated will be the same.



The correct answers are:

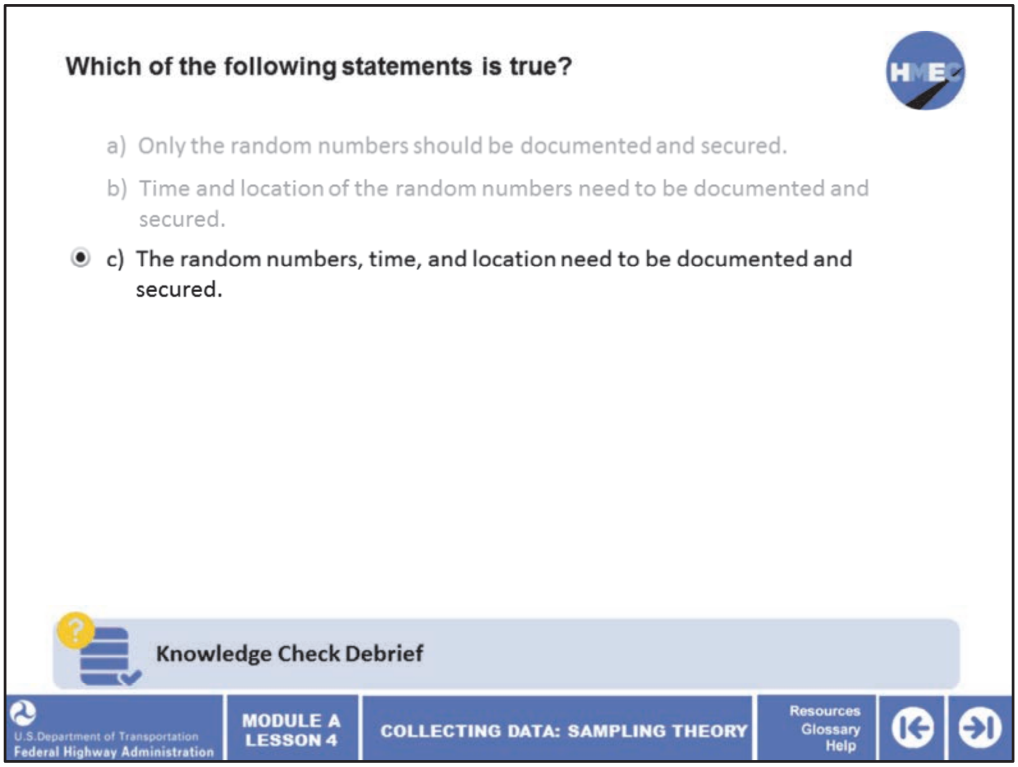
a) If a lot is evaluated using sublots with random sampling, the lot is evaluated in a more through manner; and
b) Using stratified random sampling ensures each portion of the lot has the same chance of being selected while also ensuring that the sampling is spread out over the entire lot.

The use of sublots with random sampling ensures that the sampling process is valid and will be spread over the entire lot. It also ensures that each portion of the lot has the same chance of being accepted. Sampling the first part of a lot does not provide any indication of what the rest of the lot will be like. Assuming the whole lot is the same is an erroneous assumption.

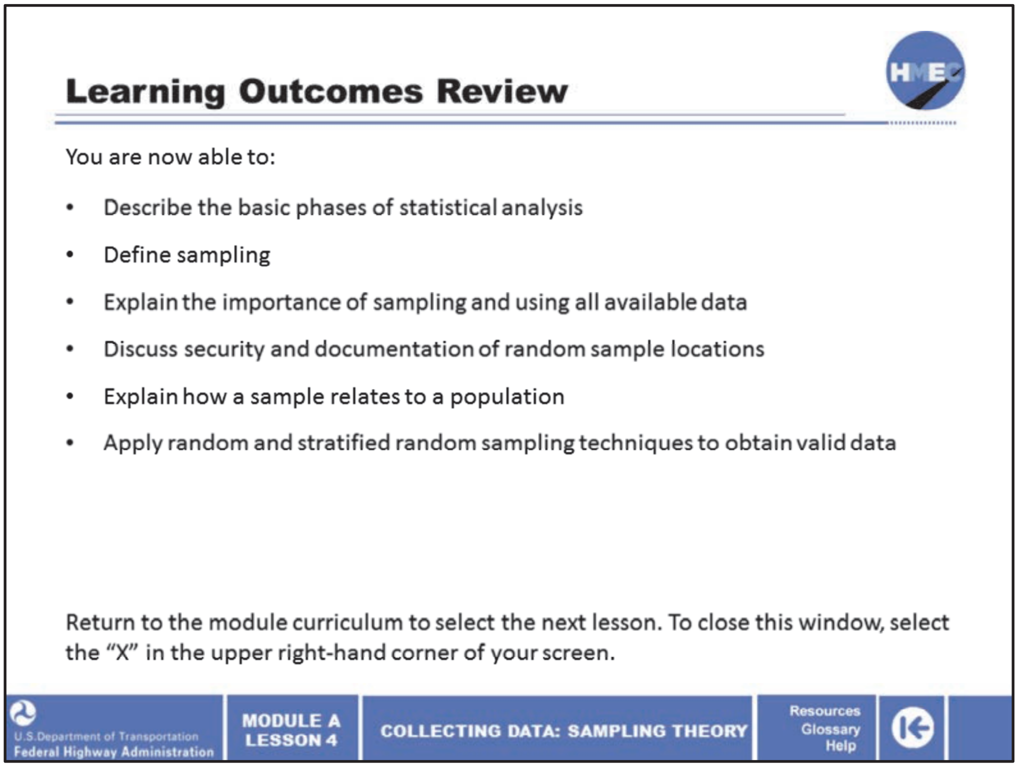


Which of the following statements is true?

a) Only the random numbers should be documented and secured;
b) Time and location of the random numbers need to be documented and secured; or
c) The random numbers, time, and location need to be documented and secured.



The correct answer is c) The random numbers, time, and location need to be documented and secured. The random numbers can be compromised if they are known prior to being needed, so the security of the numbers once they are selected is important. It is also important so that verification of the information is possible.



You have completed Module A, Lesson 4: Collecting Data: Sampling Theory. You are now able to:

- Describe the basic phases of statistical analysis;
- Define sampling;
- Explain the importance of sampling and using all available data;
- Discuss security and documentation of random sample locations;
- Explain how a sample relates to a population; and
- Apply random and stratified random sampling techniques to obtain valid data.

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.