



Welcome to the Highway Materials Engineering Course Module A, Lesson 7: Interpreting Data. This lesson provides an understanding of the basic elements of a statistically based quality assurance program and includes an introduction to quality assurance as well as techniques for collecting, organizing, and analyzing 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.

Learning Outcomes



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

- Define statistical inference, probability, and probability distributions
- Identify various probability distributions
- Explain the importance of the normal distribution
- Relate the concept of probability to the calculation of risks

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



This lesson will take approximately 145 minutes to complete.



U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



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

- Define statistical inference, probability, and probability distributions;
- Identify various probability distributions;
- Explain the importance of the normal distribution; and
- Relate the concept of probability to the calculation of risks.

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

This lesson will take approximately 145 minutes to complete.

Document Needed For This Lesson



Take a moment to download and print the Module A Lesson 7 Exercises PDF document by selecting the paperclip icon.

For this lesson, you will need the following exercises:

- Module A, Lesson 7: Exercise 1
- Module A, Lesson 7: Exercise 2
- Module A, Lesson 7: Exercise 3
- Module A, Lesson 7: Exercise 4
- Module A, Lesson 7: Exercise 5




To open the exercise document, select the paperclip icon.



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.

Phases of Statistical Analysis




1. Collect Data
2. Organize the Data
3. Analyze the Data
4. Interpret the Data

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA


Resources
Glossary
Help



Now let's get started. To review, the four phases of statistical analysis are:

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

Phases of Statistical Analysis




1. Collect Data
2. Organize the Data
3. Analyze the Data
4. Interpret the Data

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



This lesson will cover the fourth phase of statistical analysis: interpreting the data. Up to this point we have covered the first three phases of statistical analysis. These have concentrated on collecting, organizing, and analyzing sample data. While all three previous phases are important, it is even more important to understand how the information provided by the sample data is used to generalize about the population or lot from which the sample data was obtained. This last phase is crucial to writing and interpreting comprehensible and defensible specifications.

We 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 is often misinterpreted.

Definitions and Terminology



- Terms:
 - [Statistical Inference](#)
 - [Probability](#)
 - [Probability Distribution](#)
 - [Normal Distribution](#)



Select each term to see the definition



U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



The terms statistical inference, probability, probability distribution, and normal distribution are used throughout the course. Their definitions are discussed to make sure we share a common language when discussing quality assurance (QA).

Select each term to see the definition.

Image Description: Stack of books.


Definitions and Terminology

- Terms:
 - Statistical Inference ←
 - Probability
 - Probability Distribution
 - Normal Distribution

Select each term to see the definition

Statistical Inference [X] CLOSE

- An estimate or decision



U.S. Department of Transportation
Federal Highway Administration

**MODULE A
LESSON 7**

Statistical inference is an estimate or decision.

Image Description: Photo of a desktop with calculator, pencil, paperclips and a hand doing calculations.


Definitions and Terminology

- Terms:
 - [Statistical Inference](#)
 - [Probability](#) ←
 - [Probability Distribution](#)
 - [Normal Distribution](#)

Select each term to see the definition

Probability [X] CLOSE

- A measure of chance



U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

Probability can be defined as a measure of chance.

Image Description: Image of rolling dice.

Definitions and Terminology

- Terms:
 - [Statistical Inference](#)
 - [Probability](#)
 - [Probability Distribution](#) ←
 - [Normal Distribution](#)

Select each term to see the definition

Probability Distribution [X] CLOSE

- A set of measurements or counts that constitute a part or all of the population

More Unlikely | More Likely

Impossible | Even Chance | Certain

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

A probability distribution is a continuous distribution that may take many shapes.

Image description: A graph ranging from impossible to certain with mid point being an even chance.

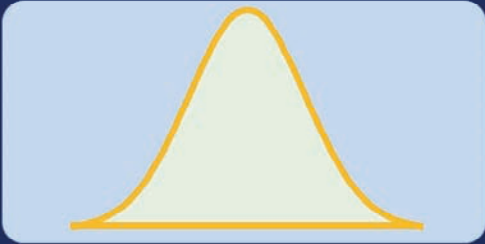
Definitions and Terminology

- **Terms:**
 - [Statistical Inference](#)
 - [Probability](#)
 - [Probability Distribution](#)
 - [Normal Distribution](#)

Select each term to see the definition

Normal Distribution X CLOSE

- A specific continuous distribution (curve) having a bell shape that is determined by values of mu (μ) and sigma (σ), and is often used to describe the distribution of individual measurements (such as construction materials properties)

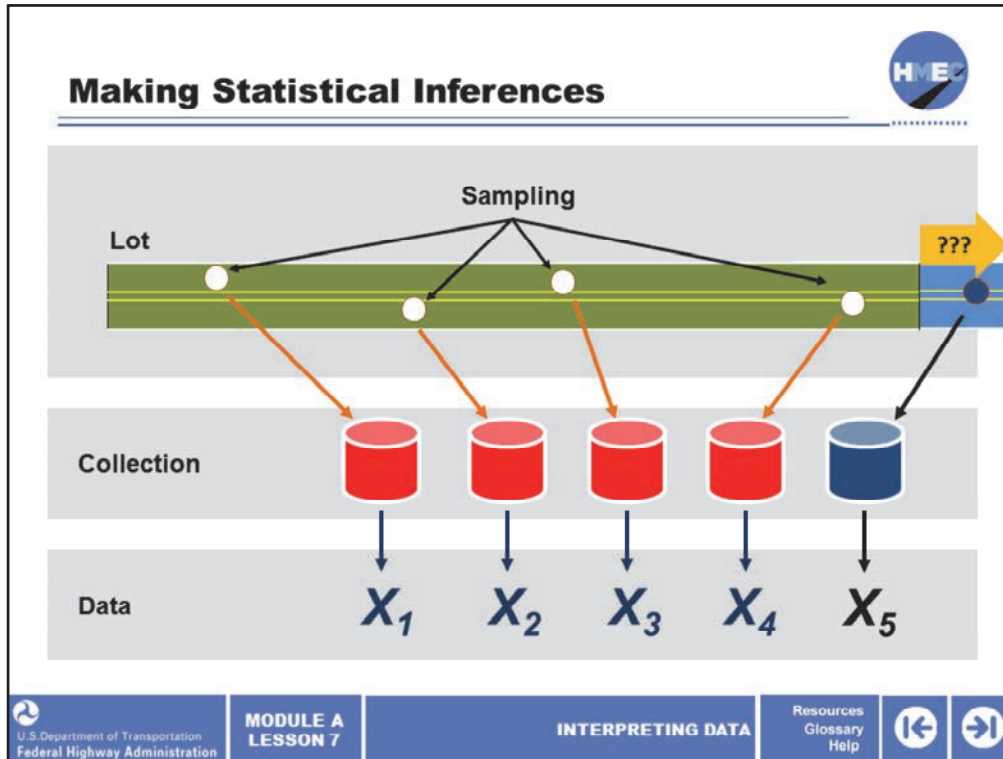


U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

The normal distribution is a specific continuous distribution (curve) having a bell shape that is determined by values of mu (μ) and sigma (σ), and is often used to describe the distribution of individual measurements (such as construction materials properties).

Image description: Distribution curve.



This screen is similar to the one shown previously where we saw the relationship among the lot, the sample, and the data obtained from the sample. In the present discussion, we want to ensure that the sample data are interpreted correctly to make proper statistical inferences about the lot (or population) of material that was sampled. In this case, an inference is an estimate or prediction about a population based on information from the sample. For example, we might be interested in drawing inferences about the acceptability of a lot of Portland cement concrete based on the compressive strength of a sample size of five cores. Think about your experience and consider how large a lot (length of highway, for example) would be sampled to obtain five cores.

Image description: The relationship among the lot, the sample, and the data obtained from the sample.

Calculating Probability



- Inferences rely on probability
- Probability can be calculated using histograms and areas under the normal curve



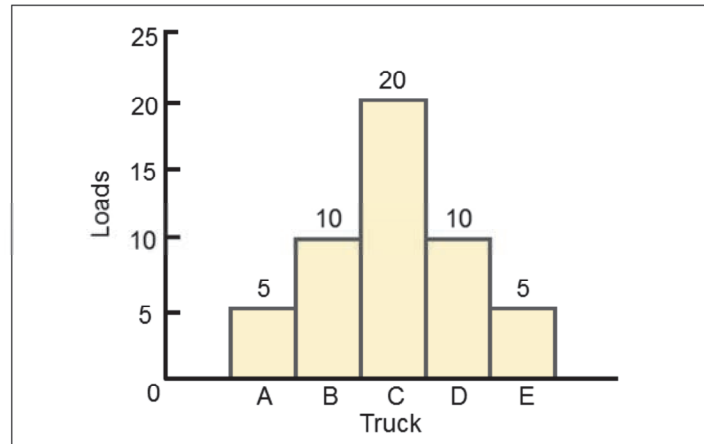
To be able to make inferences about lots or populations based on sample data, we must introduce and understand some basic points about probability. Inferences rely on probability. Probability can be calculated using histograms and areas under the normal curve. We will do this by investigating how histograms, relative frequency histograms, and areas under the normal curve can be used to calculate probabilities of certain outcomes.

Image Description: Image of rolling dice.

Discrete Probability: Frequency



- Sampling from a discrete population provides different information than sampling from a continuous population



Sampling from a discrete population provides different information than sampling from a continuous population. The following is a discrete data example: A paving contractor is using five trucks (A, B, C, D, E) to deliver asphalt concrete to a highway project. On any given day, the number of loads (or frequency) delivered by each truck will vary. This histogram illustrates the distribution of 50 truckloads that were delivered to the project on a certain day. Truck A delivered five loads to the project; truck B, 10 loads; truck C, 20 loads; truck D, 10 loads; and truck E, 5 loads.

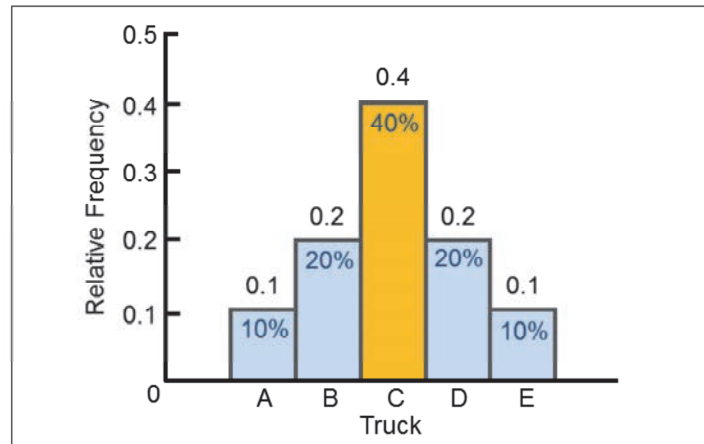
This is discrete data and probability because it is either truck A, B, C, D, or E. This distinguishes it from a continuous data situation.

Image description: Histogram.

Discrete Probability: Relative Frequency



- Converting loads (or frequency) into the relative frequency on the vertical axis and the use of the histogram for the truck distribution



If we convert loads (or frequency) into the relative frequency on the vertical axis and use this for the histogram for the distribution of trucks, we get the histogram shown on the screen. We can interpret the term "relative frequency" in two ways:

- 40% (20 divided by 50) of the truckloads were delivered by truck C. Note that the relative frequency can be expressed in either percent or decimal form; and
- If a truck were randomly selected for sampling on the day in question, there is a 40% chance (.4 or 40% probability) that truck C would have been selected.

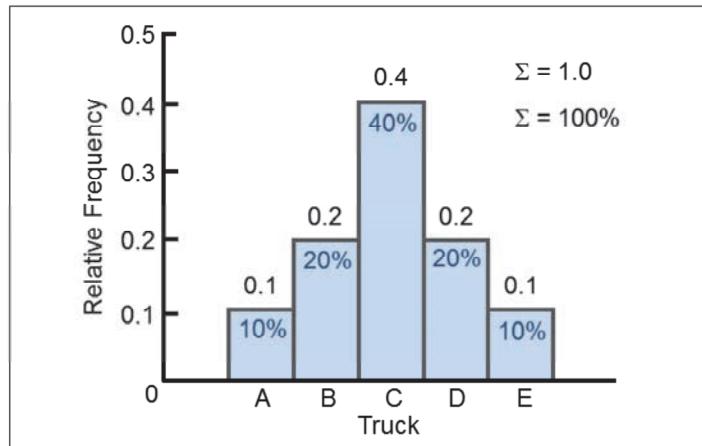
This is a simple concept but becomes important as we make decisions about populations and risks.

Image description: Histogram.

Relative Frequency = Probability



- Probability can be defined as a measure of chance, or the sum of the probabilities of all of the possible outcomes



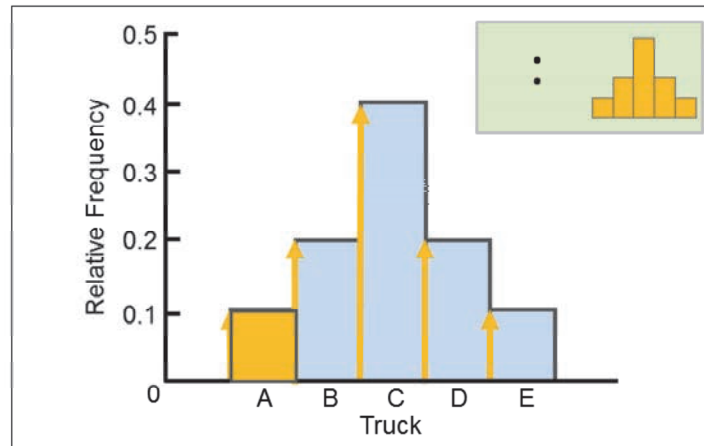
Probability can be defined as a measure of chance. A specific truck's relative frequency is therefore a measure of the probability that it will be the one that is selected if a random selection process is used and the sum of all of the relative frequencies is 1.00 (or 100%). In other words, the sum of the probabilities of all of the possible outcomes (in this case, the five different trucks) is 100% (or 1.0).

Image description: Histogram.

Probability with Discrete Data



- If the total area of the histogram is set equal to 100 units (or 1.0), then the area of each rectangle directly indicates the probability



To summarize, the probability of discrete data can be obtained from the relative frequency histogram in two ways:

- The height of each rectangle represents the probability; and
- The ratio of the area of each rectangle to the total area of all the rectangles indicates the probability.

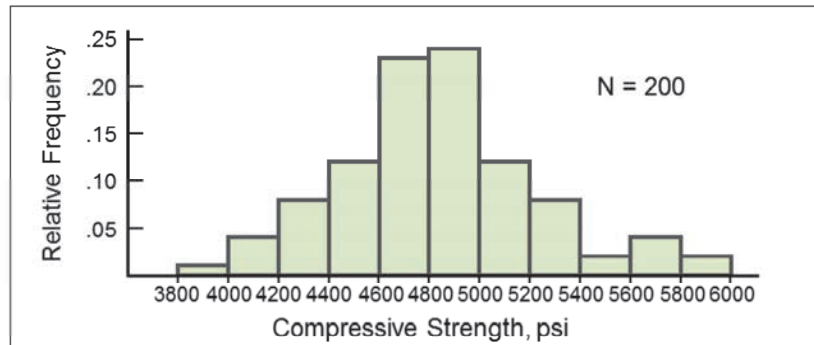
So, that if the total area of the histogram is set equal to 100 units (or 1.0), then the area of each rectangle directly indicates the probability.

Image description: Histogram.

Probability with Continuous Data



- Think about how the interpretation of the relative frequency histogram is different when we are dealing with measurements of construction materials



ANSWER



What makes this continuous data?



Now looking at a case of probability using continuous data, let's think about how the interpretation of the relative frequency histogram is different when we are dealing with measurements of construction materials, since measurements are usually continuous data. This is the case if we are analyzing pavement thickness, concrete strength, asphalt content, aggregate gradation, etc. The difference in interpretation is illustrated in this screen using the concrete compressive strength population of 200 concrete cylinders. The relative frequency histogram for this population is shown on the vertical axis of the screen and the compressive strength is on the horizontal axis. We see the compressive strengths range from a low of about 3,800 psi. to a high of about 6,000 psi. and the center of the distribution is somewhere around 4,800 to 5,000 psi. This is all important data about the population.

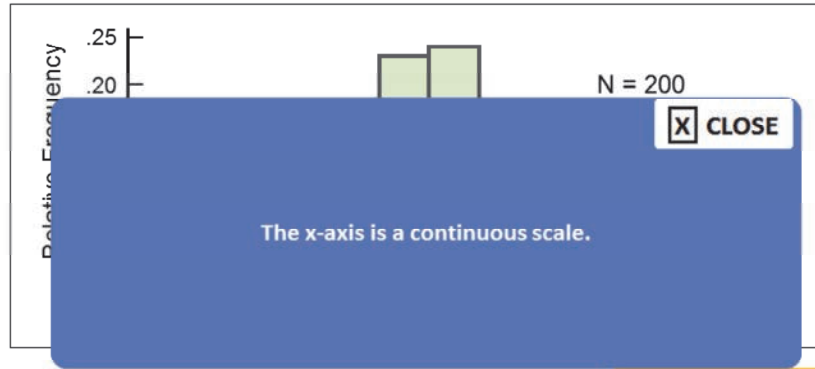
Select the box to answer the question, what makes this continuous data?

Image description: Histogram.

Probability with Continuous Data



- Think about how the interpretation of the relative frequency histogram is different when we are dealing with measurements of construction materials



What makes this continuous data?

ANSWER



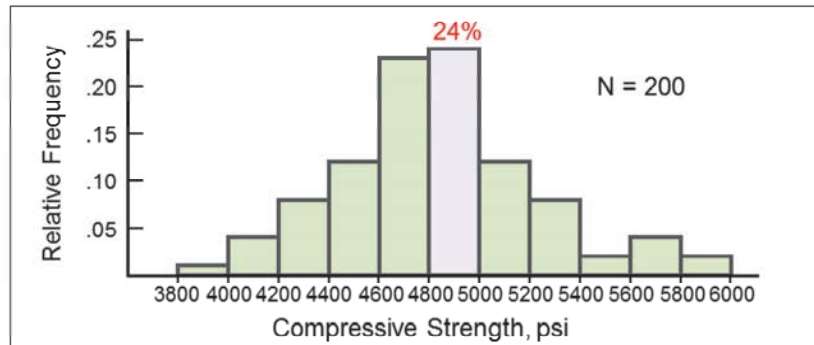
The x-axis is a continuous scale.

Image description: Histogram.

Probability with Continuous Data



- The probability must be related to an area not a single point



24% of the 200 test results are between 4800 psi and 5000 psi

If a test result were randomly selected from the population (or lot), there is a 24% chance (24% or 0.24 probability) that the test result will have a value between 4800 psi and 5000 psi.



In the case of discrete data, the probability of an occurrence could be read as a single point such as truck A or B. That is not the case for continuous data. The probability must be related to an area not a single point. In this case of concrete compressive strength values, if we set the total area under the histogram equal to 100 percent (or 1.0), we can interpret the 24% area (and height) of the blue shaded rectangle in two possible ways of relative frequency:

- 24% of the 200 test results are between 4,800 psi and 5,000 psi; and
- If a test result were randomly selected from the population (or lot), there is a 24% chance (24% or 0.24 probability) that the test result will have a value between 4,800 psi and 5,000 psi.

Image description: Histogram.

Discrete Vs. Continuous Data



Discrete

VS.

Continuous

- Here is a comparison between the interpretation of discrete and continuous data from a histogram

Select each term to learn more.



Here is a comparison between the interpretation of discrete and continuous data from a histogram.

Select each term to learn more.

Discrete Vs. Continuous Data



Discrete

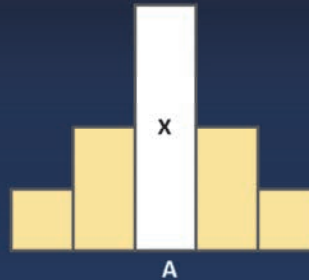
VS.

Continuous

Discrete

- Probability: $X = A$

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



X CLOSE

In the discrete data example, the areas of the rectangles represent the probability that each of the possible outcomes will occur; here it might be the white rectangle A, as in truck A.

This will be explained later in this lesson.

Image description: Histogram.

Discrete Vs. Continuous Data



Discrete

VS.

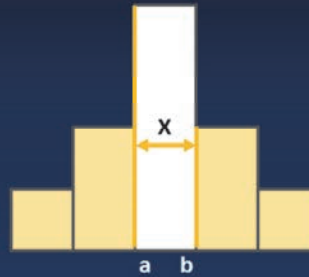
Continuous

Continuous

CLOSE

- Probability: $a \leq X \leq b$

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



In the continuous data example, the areas of the rectangles represent the probability that an outcome will be *between* two possible values as between the 4800 and 5000 psi in the concrete compressive strength example.

This will also be explained later in this lesson.

Image description: Histogram.

Continuous Data



- Probability of Exactly $X = \frac{1}{\infty} = 0$



In this example, how many outcomes can X have to be located on the scale between “a” and “b”?

ANSWER



U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



Here we see the case of continuous data and the likelihood of getting an exact number, which we call X . The explanation for the different interpretations for the two different types of data is due to the infinitely large number of possible outcomes in the continuous data case.

Select the box to answer the question: In this example, how many outcomes can X have to be located on the scale between “a” and “b”?

Image description: Continuous data line with a one end, b at the other end and x in the middle.

Continuous Data



- Probability of Exactly $X = \frac{1}{\infty} = 0$



CLOSE

It is an infinite number depending on the accuracy or the measurement scale.



In this example, how many outcomes can X have to be located on the scale between a and b?


ANSWER



It is an infinite number depending on the accuracy of the measurement scale. The probability that a continuous random variable will have a single specific value is, therefore, very, very small. It is actually one possible outcome from an infinite number of outcomes (that is, 1 divided by infinity, which is zero). Think of it this way: Which would you be willing to bet—that the next compressive strength test result you obtained would be exactly 3,612 psi or that the next test result would be between 3,500 psi and 3,800 psi? It is obvious that the chance that the next result would be 3,612 psi is very, very small, but if you knew something about the prior test results, there might be a good chance of getting a result between 3,500 and 3,800.

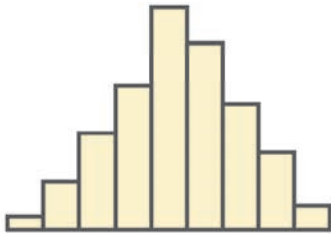
Image description: Continuous data line with a one end, b at the other end and x in the middle.

Relationship Between a Histogram and a Probability

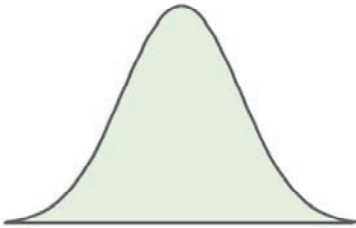



Relative Frequency Histogram

There are limitations
in using histograms to
determine probabilities.



Probability Density Function (Probability Distribution)






 U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

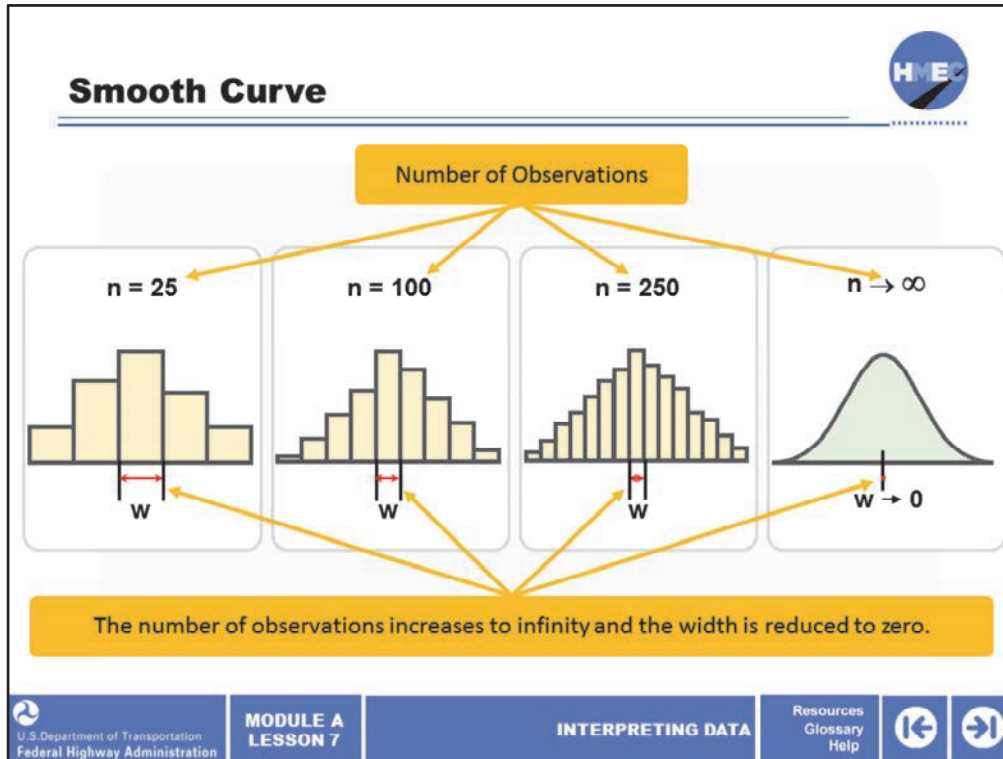
Resources
Glossary
Help

Here we see the relationship between a relative frequency histogram and a probability density function, which is often called a probability distribution. We have previously shown that the areas of the rectangles of a relative frequency histogram can be used to determine the probability that an outcome will be between two values. However, there are limitations in using histograms to determine probabilities. Histograms are limited to the results of the sample from which they were developed, and probabilities can only be determined for the class intervals that were used for developing the histogram. If we were to make inferences about the total lot or population from which the sample was taken, then we must develop a method for making more general probability statements. We will approximate the relative frequency histogram with a smooth curve that we will call a probability distribution. Statisticians call such a curve a probability density function (pdf). We will use both terms (probability and pdf) in this lesson.

Image 1 description: Histogram.

Image 2 description: Distribution curve.



Here we see a transformation from a histogram to a smooth curve probability distribution. Here we increase the number of observations, n , used in developing the histogram and this corresponds to reducing the width of the class interval, w , used in plotting the histogram. As the number of observations increases to infinity and the width is reduced to zero, then the histogram changes from $n = 25$ to $n = 100$, then with $n = 250$, and then where n becomes infinity and the width zero at which point it approaches a smooth curve.


Image 1 description: Histogram.

Image 2 description: Histogram.


Image 3 description: Histogram.




Image 4 description: Distribution curve.

Select the best answer. Why do we want a smooth curve?



- a) So the curve looks pretty
- b) So the mean and variability are more visible
- c) So we can make generalized decisions about the probability without being restricted by class widths, class limits, or sample size

 Knowledge Check Try again Submit Clear

 **MODULE A**
LESSON 7 **INTERPRETING DATA** Resources
Glossary
Help  

Select the best answer. Why do we want a smooth curve?

- a) So the curve looks pretty;
- b) So the mean and variability are more visible; or
- c) So we can make generalized decisions about the probability without being restricted by class widths, class limits, or sample size.

Select the best answer. Why do we want a smooth curve?



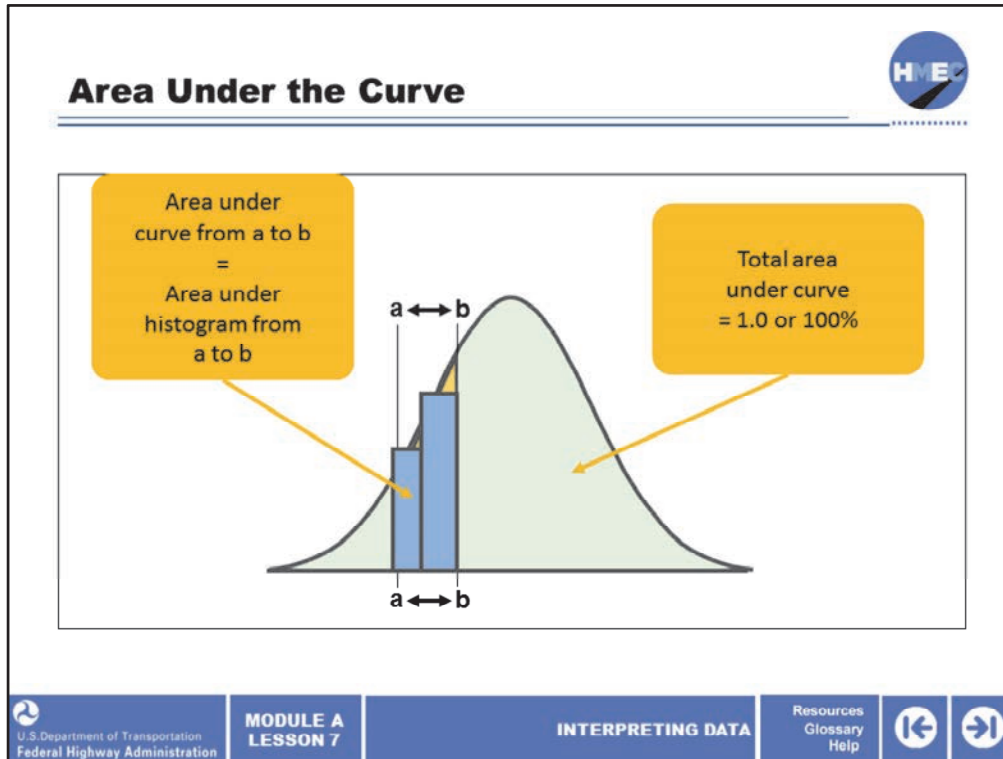
- a) So the curve looks pretty
- b) So the mean and variability are more visible
- c) So we can make generalized decisions about the probability without being restricted by class widths, class limits, or sample size



Knowledge Check Debrief



The correct answer is c) So we can make generalized decisions about the probability without being restricted by class widths, class limits, or sample size.



Here we see how the area under the probability distribution is interpreted. The smooth curve is drawn so that the area under the curve between two values corresponds to the area under the histogram between the same two values, and the total area under the curve is equal to 1.0 (i.e., 100%).

Conceptually, the triangular area of the rectangle outside the smooth curve fits into the space within the curve above the rectangle. The transformation from a histogram allows us to make probability decisions without the restrictions imposed by the decisions we made when we developed the histogram. So the smooth curve function becomes a more generalized representation of the population.

Image description: Distribution curve.

Calculating Probability

• A smooth curve and how probability relates to it

Not probability $f(x)$

Probability = Area between 2 values

The vertical axis of the probability distribution is NOT probability.

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7


INTERPRETING DATA

Resources
Glossary
Help


Here we see the smooth curve and how probability relates to it. It is important to note that although the axis of the relative frequency histogram can be expressed in terms of relative frequency, the vertical axis of the probability distribution is NOT probability; it is simply $f(x)$. Probability can only be expressed as the area under the curve between two values. Here the probability is the purple area of the smooth curve.

Image description: Distribution curve.

Typical Probability Distributions

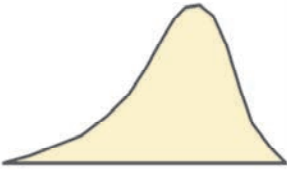


#1



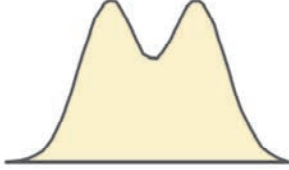
Most often represents materials and construction

#2




Represents an asymmetrical distribution


#3



Has two peaks




The last two indicate that something unusual happened while the data was being collected.



MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



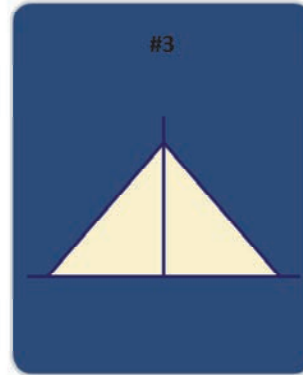
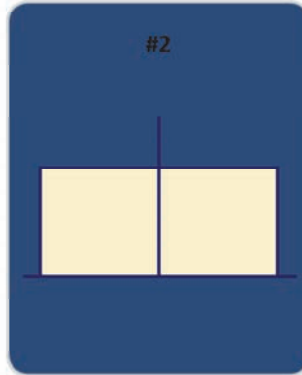
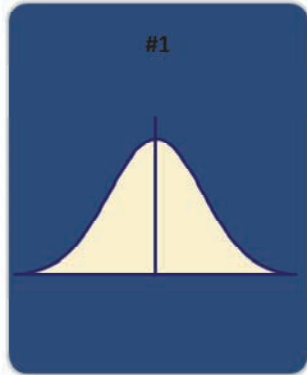
So far, we have only looked at a simple, and the most common shape, of a smooth curve. Although the shapes of smooth curve probability functions can vary greatly, experience has shown that particular types of data tend to exhibit certain shapes. For illustrative purposes, some of the more common shapes are discussed and related to concrete compressive strength. Curve #1 is normal and is one that most often represents materials and construction. Curve #2 represents an asymmetrical distribution and curve #3 has two peaks. The last two indicate that something unusual happened while the data was being collected.

Image 1 description: Distribution curve.

Image 2 description: Distribution curve.

Image 3 description: Distribution curve.

Typical Symmetric Distributions



Select each type to learn more.



There are several symmetrical distributions, meaning a distribution with a shape that is a mirror image on opposite sides of a line dividing the middle of the distribution.

Select each distribution to learn more.

Image 1 description: Distribution curve.

Image 2 description: Uniform distribution.

Image 3 description: Triangular shape.

Typical Symmetric Distributions



#1

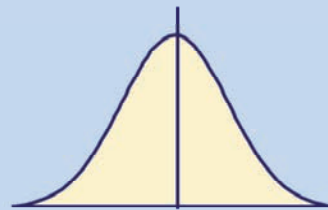
#2

#3

Normal or Gauss

CLOSE

- Most common and important symmetric distribution
- Unimodal (single peak)
 - Describes many populations that occur in:
 - Nature
 - Research
 - Industry



The most common and important symmetric distribution is the normal or Gauss probability distribution shown in curve #1. The normal distribution is a unimodal (a single peak) distribution that has been found to adequately describe many populations that occur in nature, research, and industry (including highway materials and construction).

Image description: Distribution curve.

Typical Symmetric Distributions



#1

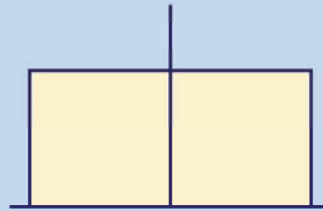
#2

#3

Uniform Distribution

CLOSE

- Do you think a distribution such as this can occur?
- If you plotted the population data from a random number table, what do you think it would look like?
- Remember:
 - A random number table in theory has the same number of numbers (1s, 2s, 3s, etc.)



Look at shape #2, which is called a uniform distribution. Do you think a distribution such as this can occur? If you plotted the population data from a random number table, what do you think it would look like? Remember a random number table in theory has the same number of numbers (1s, 2s, 3s, etc.) throughout. So it should plot as a uniform distribution.

Image description: Uniform distribution.

Typical Symmetric Distributions



#1

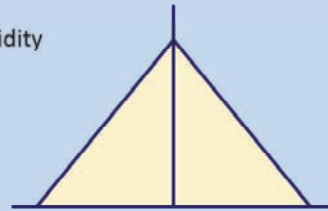
#2

#3

Triangular Shape

CLOSE

- Indicate biased data
 - Numbers at the center occur much more frequently
 - Numbers decrease uniformly as you get further from the center
- Distribution should raise some questions about the validity of the data



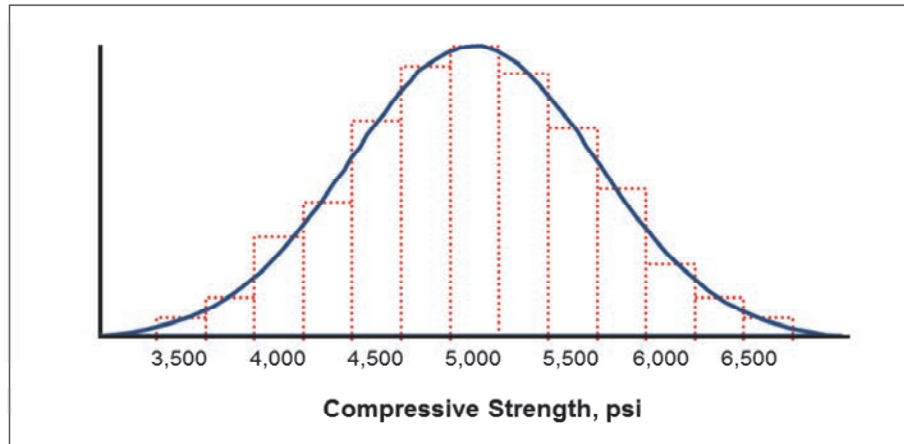
The triangular shape #3 would indicate some biased data where the numbers at the center occur much more frequently and decrease uniformly as you get further from the center. A distribution such as this should raise questions about the validity of the data.

Image description: Triangular shape.

Normal Distribution



- A normal distribution superimposed over the relative frequency histogram for a population of concrete compressive strengths



This screen indicates a normal distribution superimposed over the relative frequency histogram for a population of concrete compressive strengths. There are several important features of the normal distribution. One of the most important is that the normal probability distribution is completely described if the mean and standard deviation are known. Inspection of this distribution shows the mean (center) of the distribution is about 5,000 psi. We cannot determine the standard deviation exactly, but we see the spread, which we have seen is measured by the standard deviation is about 3,000 psi (that is 6,500 - 3,500 psi). The normal distribution is discussed in great detail in this lesson because it has been found to apply to many highway material and construction situations. Most properties are approximately normal; few are perfectly normal.

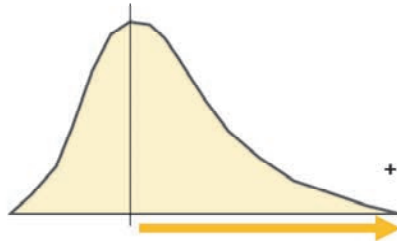
Image description: Distribution curve over an histogram.

Skewed Distributions

A probability distribution that is not symmetric.

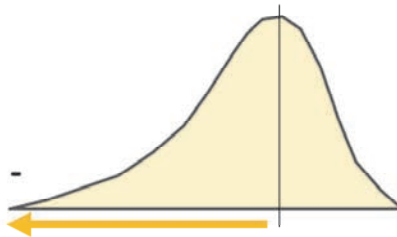
Positive

- Longer tail extending toward the higher values on the right side



Negative

- Longer tail extending toward the lower values on the left side




Here we see skewed distributions. This is a probability distribution that is not symmetric; although it may be unimodal, it is called a skewed distribution. Positive skewness indicates that the distribution has a longer tail extending toward the higher values on the right side. Negative skewness indicates the distribution has a longer tail extending toward the lower values on the left side.

Image 1 description: Distribution curve.

Image 2 description: Distribution curve.

How Skewed Results Can Occur




- Few materials have skewed distributions unless something is wrong with the process

PCC

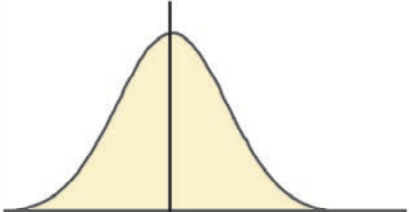

Few material properties, of the 21 measured on 3 projects, were significantly skewed

HMA Projects

Out of 52 material properties measured on 3 projects, 6 (11.5%) had values that indicated a significant skewness



100%






U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help

Here we see an indication of how skewed results can occur. Few materials have skewed distributions unless something is wrong with the process.

A study done for the FHWA on both hot mix asphalt (HMA) and Portland cement concrete (PCC) projects examined the occasions where skewed results occurred. For PCC, few material properties, of the 21 measured on 3 projects, were significantly skewed. For the HMA projects, out of 52 material properties measured on 3 projects, 6 (11.5%) had values that indicated a significant skewness. Five were from gradation results. One potential source of skewness is the presence of a physical barrier, such as what occurs with the top size sieve.

Since a gradation cannot exceed 100%, when the mean approaches this limiting value, the distribution will typically appear to be skewed. So if you have a specification on percent passing on the top size sieve that allows say, 90 to 100% passing, and the job mix formula (JMF) is set at 98% passing, it is almost a certainty that the results will be skewed.

Image 1 description: Distribution curve.

Image 2 description: Distribution curve.

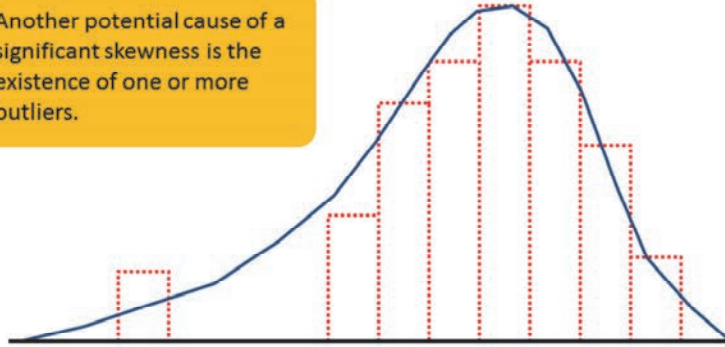
Image 3 description: Gradation example.

Outliers and Skewness



- If the sample size is relatively small, a single outlier can pull the tail of the distribution in the direction of the outlier


Another potential cause of a significant skewness is the existence of one or more outliers.



Another potential cause of a significant skewness is the existence of one or more outliers. If the sample size is relatively small, a single outlier can pull the tail of the distribution in the direction of the outlier. We previously discussed the effect of outliers on the standard deviation. In an earlier screen, we saw the effect of changing one value from 50 to 58. Although in this case, it was said the 58 might not be an outlier (we were presenting the concept and did not perform a statistical test to determine if it was or was not); it did almost double the standard deviation. This increase in standard deviation caused by a single value is an indication that significant skewness exists.

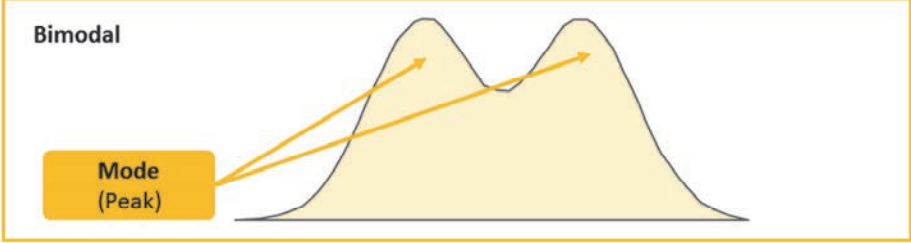
Image description: Distribution curve over an histogram.

Multimodal Distributions

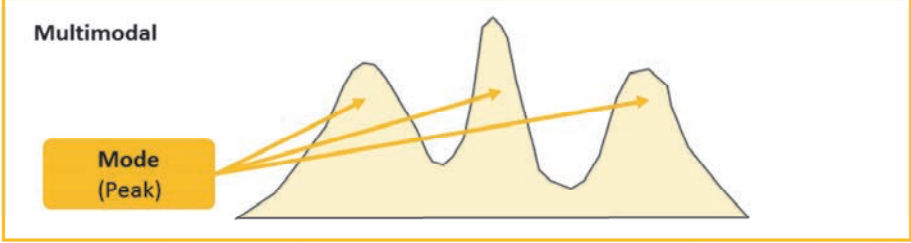


- Multimodal probability distributions exhibit more than one peak or mode

Bimodal



Multimodal




U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



Here we see the concept of multimodal distributions. Multimodal probability distributions exhibit more than one peak or mode. A distribution with two peaks, for instance, would be called bimodal. Thus, multimodal distributions show several peaks or modes. We will see in the next few screens how this can occur. But the important point is that in these cases, something unusual has occurred. Materials and construction do not ordinarily follow a multimodal distribution.

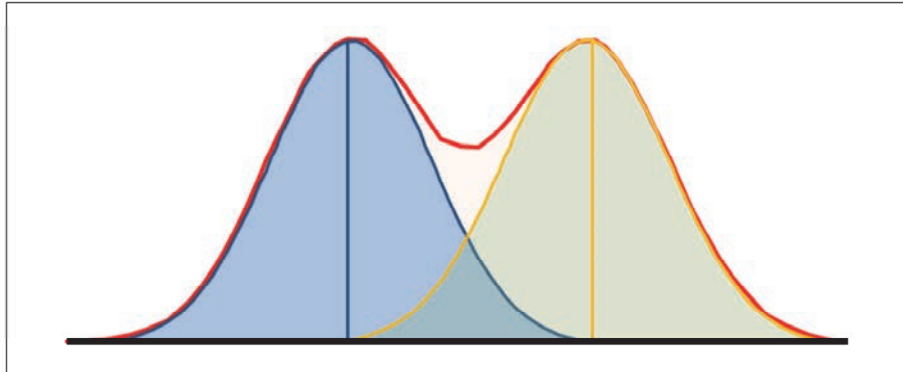
Image 1 description: Distribution curve.

Image 2 description: Distribution curve.

Bimodal Distribution: Obvious



- Data from two separate populations were mistakenly combined



ANSWER



How can this happen?




Here we see an obvious case of a bimodal distribution. There are two normal distributions side-by-side. One of the primary reasons for concern if a probability distribution appears bimodal is that sampling was probably not performed on a single controlled process. When we say that sampling was probably not performed on a single controlled process, that means that data from two separate populations was mistakenly combined.


Select the box to answer the question, how can this happen?

Image description: Two distribution curves overlapping.

Bimodal Distribution: Obvious



- Data from two separate populations were mistakenly combined



Supposed we are producing aggregate base through an aggregate plant and inadvertently the operator changes the JMF, but the sampling continues without realizing this.

If the sampling data from prior to the change are combined with the data after the change, the analysis of the entire data set would likely indicate a bimodal distribution.

In the case here, it would be obvious, indicating a large change.

X CLOSE


Q&A How can this happen? **ANSWER**

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



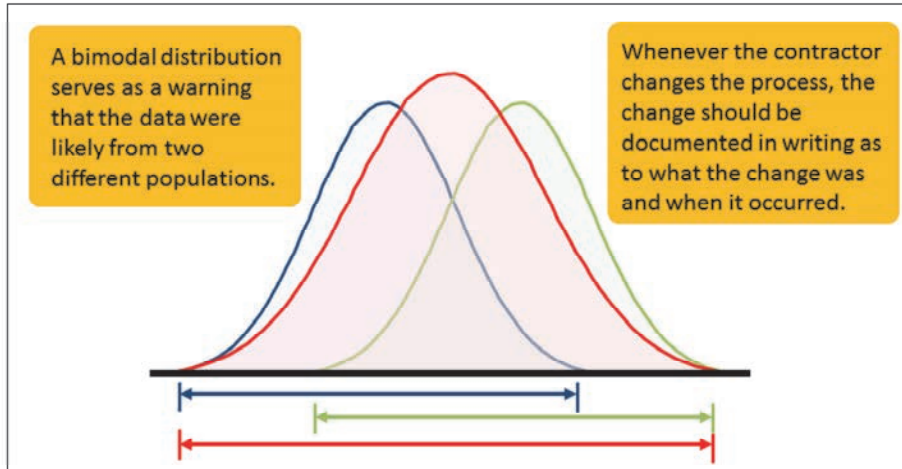
The figure illustrates how data from two different populations might combine to form a single bimodal distribution. How can this happen? Supposed we are producing aggregate base through an aggregate plant and inadvertently the operator changes the JMF, but the sampling continues without realizing this. If the sampling data from prior to the change are combined with the data after the change, the analysis of the entire data set would likely indicate a bimodal distribution. In the case here, it would be obvious, indicating a large change.

Image description: Two distribution curves overlapping.

Unimodal Distribution, but...?



- But what would happen in the previous scenario if the change were small?




But what would happen in the previous scenario if the change were small? A bimodal distribution serves as a warning that the data were likely from two different populations. Unfortunately, however, data taken from two different processes will not always plot to form a bimodal distribution.

Here we see a blue normal curve with the blue line underneath indicating the spread and next to it, the green normal curve with its spread underneath. The combined distribution here shown by the red normal curve with its spread including the combined spread may still be unimodal, but with a broader base (higher standard deviation) than either of the two individual populations.

The incorrect conclusion will then be that the data are from a single population with a high degree of variability, when in reality it is two different populations. As a point of emphasis, which will be discussed in detail in Lesson 9: Quality Control, whenever the contractor changes the process, the change should be documented in writing as to what the change was and when it occurred.

Image description: Three distribution curves overlapping.

Match each term with its definition.







Terms:

- c Statistical inference
- a Probability
- d Probability distribution
- b Normal distribution

Definitions:

- a. Measure of chance
- b. Specific continuous distribution (curve) having a bell shape that is determined by values of μ and σ
- c. Estimate or decision
- d. Continuous distribution that may take many shapes

 **Knowledge Check** Try again Submit Clear

 **MODULE A**
LESSON 7 **INTERPRETING DATA** Resources
Glossary
Help  

Match each term with its definition.

Terms:

- Statistical inference;
- Probability;
- Probability distribution; and
- Normal distribution.

Definitions:

a) Measure of chance;


b) Specific continuous distribution (curve) having a bell shape that is determined by values of μ

and σ ;

c) Estimate or decision; and

d) Continuous distribution that may take many shapes.

Match each term with its definition.




Terms:

- c Statistical inference
- a Probability
- d Probability distribution
- b Normal distribution

Definitions:

- a. Measure of chance
- b. Specific continuous distribution (curve) having a bell shape that is determined by values of μ and σ
- c. Estimate or decision
- d. Continuous distribution that may take many shapes



 Knowledge Check Debrief

U.S. Department of Transportation
Federal Highway Administration

**MODULE A
LESSON 7**

INTERPRETING DATA

Resources
Glossary
Help



The correct answers are:

- Statistical Inference is c) Estimate or decision;
- Probability is a) Measure of chance;
- Probability distribution is d) Continuous distribution that may take many shapes; and
- Normal distribution is b) Specific continuous distribution (curve) having a bell shape that is determined by values of μ and σ .

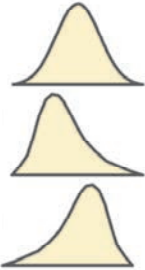
Match the graph to its probability distribution.

Probability Distributions:

- b Positive
- c Negative
- a Normal

Column 2

- a. Graph with values evenly distributed on both sides
- b. Graph with a longer tail extending toward the higher values on the right side
- c. Graph with a longer tail extending toward the higher values on the left side




Knowledge Check Try again Submit Clear

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



Match the graph to its probability distribution.

Probability Distributions:

- Positive;
- Normal; and
- Negative.

Graph:

- a) Graph with values evenly distributed on both sides;
- b) Graph with a longer tail extending toward the higher values on the right side; and
- c) Graph with a longer tail extending toward the higher values on the left side.

Image 1 description: Distribution curve.

Image 2 description: Distribution curve.

Image 3 description: Distribution curve.

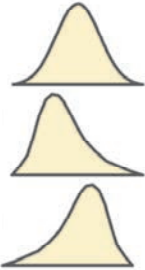
Match the graph to its probability distribution.

Probability Distributions: Column 2

b Positive a. Graph with values evenly distributed on both sides

c Negative b. Graph with a longer tail extending toward the higher values on the right side

a Normal c. Graph with a longer tail extending toward the higher values on the left side




Knowledge Check Debrief

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



The correct answers are b, c, a.

- Positive probability distributions have b) a longer tail extending toward the higher values on the right side;
- Negative probability distributions have c) a longer tail extending toward the higher values on the left side; and
- Normal probability distributions have a) values evenly distributed on both sides.

Image 1 description: Distribution curve.

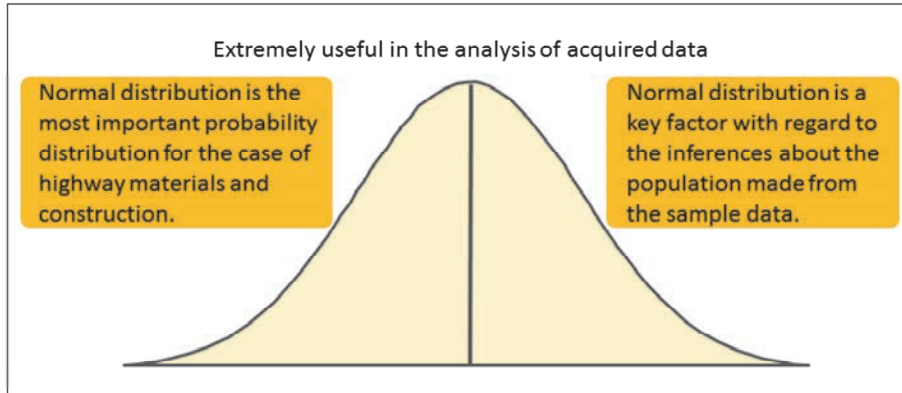
Image 2 description: Distribution curve.

Image 3 description: Distribution curve.

Normal Probability Distribution



- Normal curve



The normal distribution is the most important probability distribution for the case of highway materials and construction.



U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



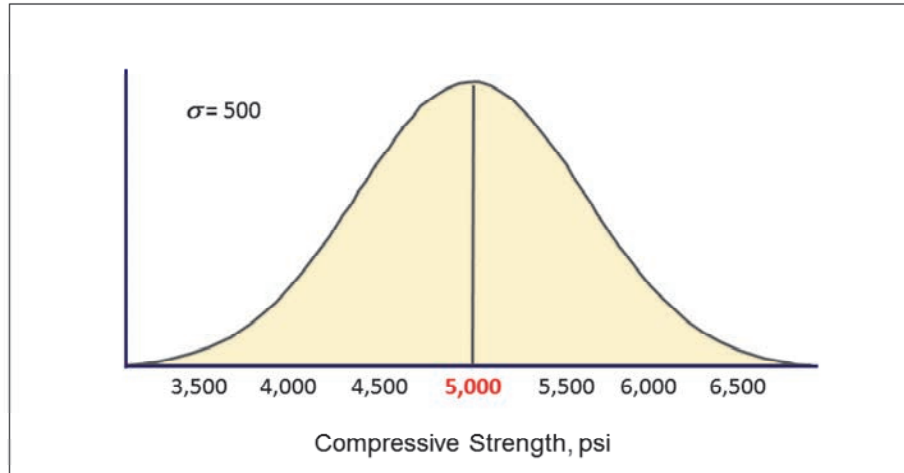
Now we will concentrate on the most common distribution found in materials and construction: the normal curve. From extensive research, it has been concluded that numerous measurements that occur in highway construction, and in nature in general, distribute themselves about some average value with the majority of the measurements grouped near the mean and with progressively fewer results recorded as one proceeds away from the mean. The normal distribution, therefore, is the most important probability distribution for the case of highway materials and construction. Besides being extremely useful in the analysis of acquired data, the normal distribution is a key factor with regard to the inferences about the population that are made from the sample data.

Image description: Distribution curve.

Normal Distribution Properties



- What would you expect the variability to be in concrete like this?



Let's look at the important properties of the normal distribution, also commonly called the normal curve. We will begin our study of the normal distribution by looking at some of the properties or features of this important distribution. Some of these have already been presented, but are reiterated here along with some additional properties.

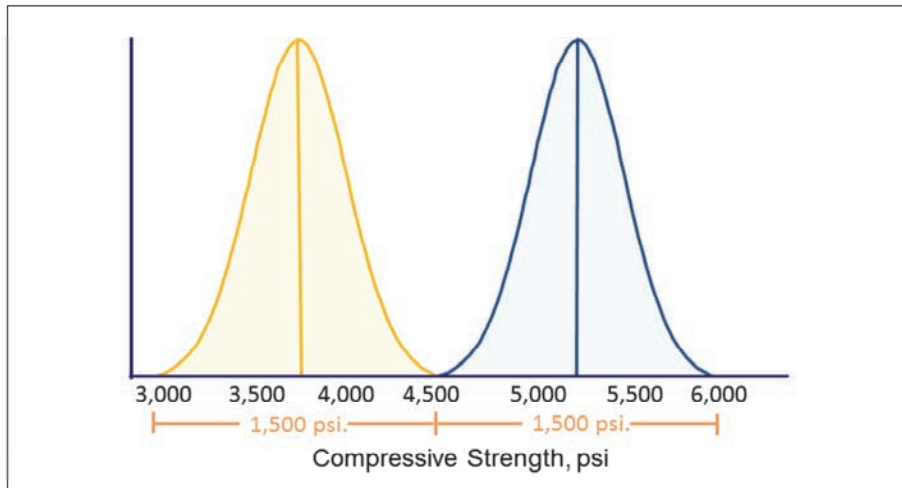
A concrete compressive strength example is used to illustrate the features of the normal distribution. This population, which is shown on the screen, has a mean of 5,000 psi and a standard deviation of 500 psi. From the plot it can be seen that the distribution is unimodal, that most of the compressive strength test results are clustered near the middle (the mean) of the distribution, and that the normal distribution is symmetric about the vertical axis through the mean. This is about any experience you may have with concrete compressive strengths. What would you expect the variability to be in concrete like this?

Image description: Distribution curve.

Normal Distribution Properties



- This figure shows two distributions with the same standard deviation



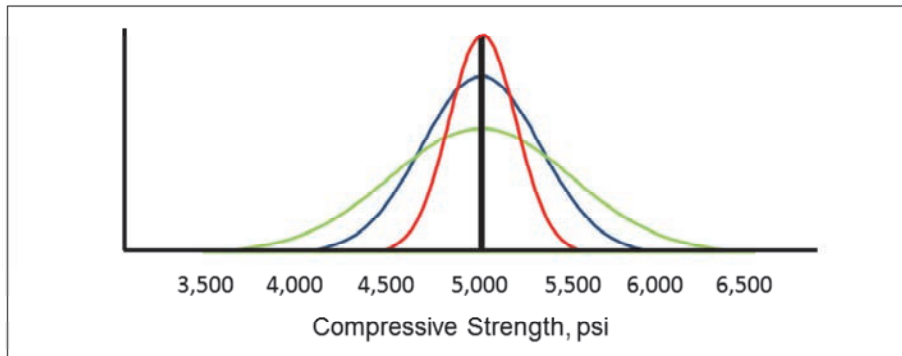
A particular normal distribution is uniquely and completely described if the mean and standard deviation associated with the distribution are known. A different normal distribution is created whenever the value of *either* the mean or standard deviation is changed. There are, therefore, an infinite number of possible normal distributions, each with a different combination of mean and standard deviation. This figure shows two distributions with the same standard deviation represented by a spread of about 1,500 psi, but with different means, that is, one centered at about 4,000 psi and the other centered at about 5,500 psi.

Image description: Two distribution curves overlapping.

Properties of the Normal Distribution



Three normal distributions with the same mean but each with a different standard deviation



ANSWER



Why do you think they have different heights?



Looking at properties of the normal curve again, the figure in this screen shows three normal distributions with the same mean, about 5,000 psi, but each with a different standard deviation as represented by the spread. The green normal curve has the largest spread of about 3,000 psi, the blue curve a smaller spread of about 2,000 psi and the red curve with the smallest spread of about 1,000 psi. But observe the heights of the three normal curves.

Select the box to answer the question: Why do you think they have different heights?

Image description: Three distribution curves overlapping.

Properties of the Normal Distribution

Three normal distributions with the same mean but each with a different standard deviation

Remember that each curve represents 100% of the population.

Therefore, if the base is reduced, the height has to increase. This is important visually, but as we said earlier, the height has no meaning statistically.

Why do you think they have different heights?

ANSWER

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help

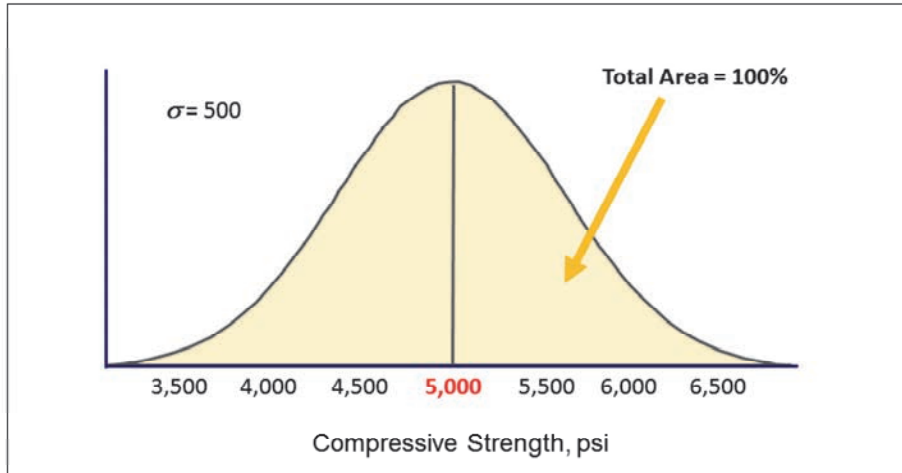
Remember that each curve represents 100% of the population. Therefore, if the base is reduced, the height has to increase. This is important visually, but as we said earlier, the height has no meaning statistically.

Image description: Three distribution curves overlapping.

Areas Under the Curve



- Areas under the normal curve



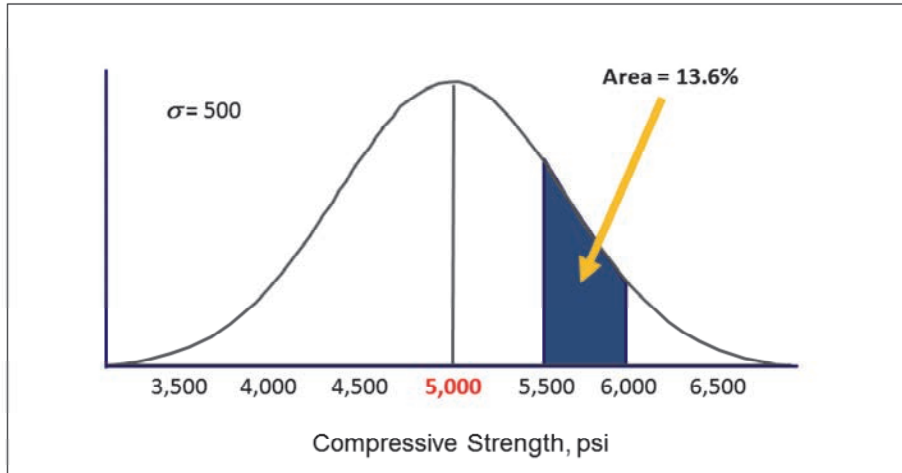
Now we will discuss areas under the normal curve—how we can find them, and how we will relate them to specifications. We will get a little deeper into the very useful properties on the normal curve and how we can use these properties to tell us more about the population. Like all other probability distributions, the total area under the normal distribution is equal to 100% (or 1.0). This follows from our discussion about frequency histograms and the transformation to the smooth curve discussed earlier in this lesson.

Image description: Distribution curve.

Interpreting the Area Under the Curve



- Understanding what the area means



In the previous screen, we discussed the importance of being able to determine areas under the normal curve. But before we understand how to do that we need to understand what the areas mean. The figure shows the population of concrete strengths that we have seen before with the area between 5,500 and 6,000 psi highlighted. The highlighted area under the normal distribution can be interpreted in two ways. (Think back to the example of trucks hauling to a project and the two ways to interpret the number of loads the trucks hauled.) The same interpretation exists for this example:

- 13.6% of the test results in the population occur between 5,500 psi and 6,000 psi; and
- If a randomly selected test result were obtained from a population (lot) with a mean concrete compressive strength of 5,000 psi and standard deviation of 500 psi, there is a 13.6% probability (i.e., chance) that the strength will be between 5,500 psi and 6,000 psi.

Image description: Distribution curve.

Normal Distributions

• All probability distributions can be described by a mathematical formula

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help

← →

All probability distributions can be described by a mathematical formula, otherwise they could not be generalized. The normal distribution is no exception and it has what looks to be a complicated equation that can be used to represent it in terms of its mean and standard deviation. Remember, they are the two population parameters that fully describe the normal distribution and separates one from another. One way to calculate the area under the curve, therefore, would be to integrate the equation between the limits in which we are interested.

Fortunately, we do not need to do this because all normal distributions can be transformed into a standard normal distribution for which areas are readily available in the form of a reference table. Note: In the upper right-hand corner of the screen, the area in parentheses is X minus mu divided by sigma. This is called a function and we will use the next screen to simplify the normal distribution equation.

Image 1 description: Equation.

Image 2 description: Distribution curve.

Image 3 description: Distribution curve.

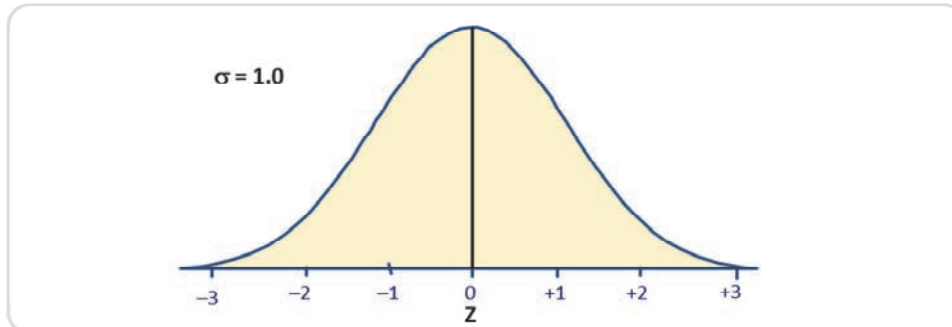
Image 4 description: Distribution curve.

Standard Normal Distribution



- A normal distribution with mean = 0 and standard deviation = 1

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(Z)^2}$$




This screen shows what looks like a normal distribution. In fact, it is but it is a very specific normal distribution; it is a normal distribution with mean = 0, and standard deviation = 1. This may seem odd. Do you know of any highway materials that have a mean of zero and a standard deviation of one? There is likely not one, so what is the usefulness of a distribution such as this? We will see that next and become comfortable with its usage. But first some more details about the standard normal curve. The horizontal axis for the standard normal distribution is usually designated as the Z-axis. The values on the Z-axis are equal to the number of standard deviations above (positive Z values), or below (negative Z values) the mean of 0. The equation for the Z statistic is $Z = X \text{ minus } \mu \text{ divided by } \sigma$, so this simplifies the function that describes the distribution. We will see how this also provides a generalization that can be used with any normal distribution.

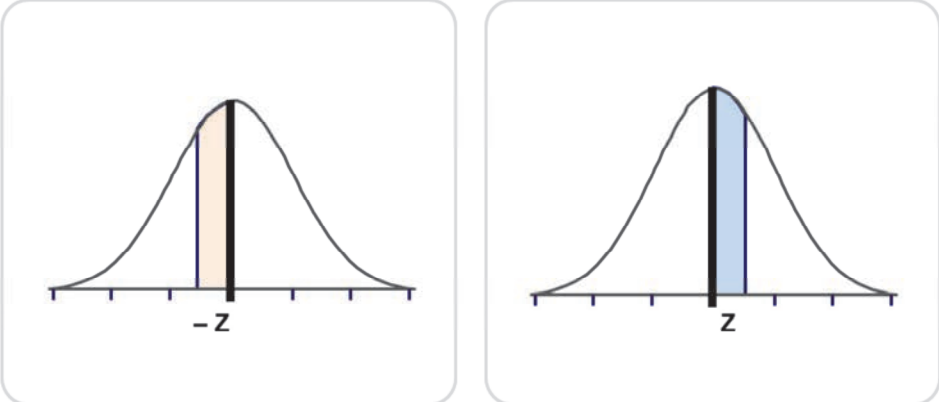
Image 1 description: Equation.

Image 2 description: Distribution curve.

Exercise 1: Familiarization with the Z Table



- The standard normal distribution is also called the standard normal curve




To open the exercises document, select the paperclip icon.

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



Now let's take a moment for an exercise. Use the exercises that were provided in the document from the paperclip icon. Please allow 5 minutes to conduct this exercise.

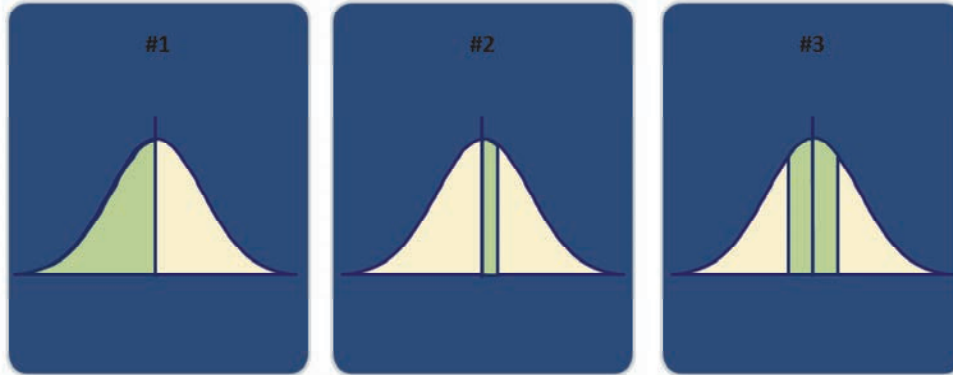
The standard normal distribution is also called the standard normal curve. Once we see how we use the Z-value to determine the area under the curve we will see how it is useful for applications to the normal curve. Z-values have been developed in a table showing the areas under the standard normal distribution, and the significance of the results that can be obtained by using the table are illustrated in the examples that follow.

Please review the tabulation of the areas under the standard normal distribution in the exercise before moving onto the next screen.

Image 1 description: Distribution curve.

Image 2 description: Distribution curve.

Calculate Area Examples



Select each example to learn more.



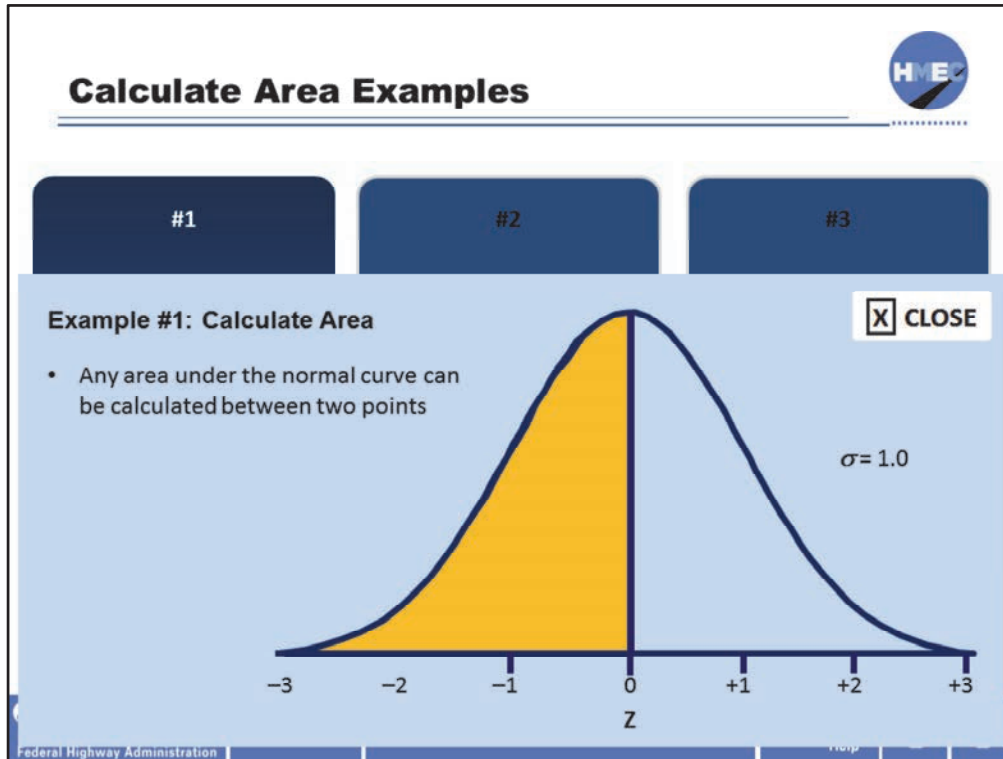
Calculate the area examples.

Select each example to learn more.

Image 1 description: Distribution curve.

Image 2 description: Distribution curve.

Image 3 description: Distribution curve.



This is the first example to determine the area under the standard normal curve. Any area under the normal curve can be calculated between two points. (Think back to where we discussed the area under a frequency histogram or curve.)

This first example is to determine the area under the standard normal distribution that is less than or equal to $Z = 0$, which means we are looking at the $-Z$ side of the curve. This problem can be solved by inspection of the figure. We know a distribution is divided in half by the mean, so it is obvious that 50% (0.50) of the area is less than or equal to a $Z = 0$. If we choose to use the table, we would define the yellow shaded area as that between a $Z = 0$ and a $Z = 3.409$ (since that is as far as this table goes). The answer would be .4998 (49.98%).

In theory, the standard normal curve and thus, the normal distribution on which it is based, go to infinity in both directions. So tables of areas under the normal curve approach 50%, do not reach it mathematically, but do from a practical standpoint.

Image description: Distribution curve.

Calculate Area Examples



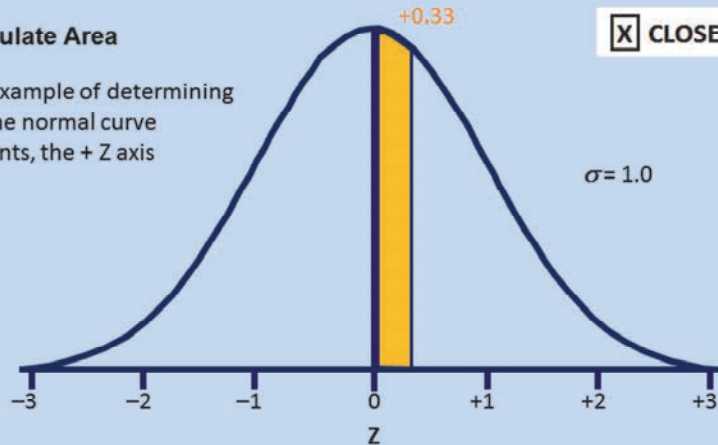
#1

#2

#3

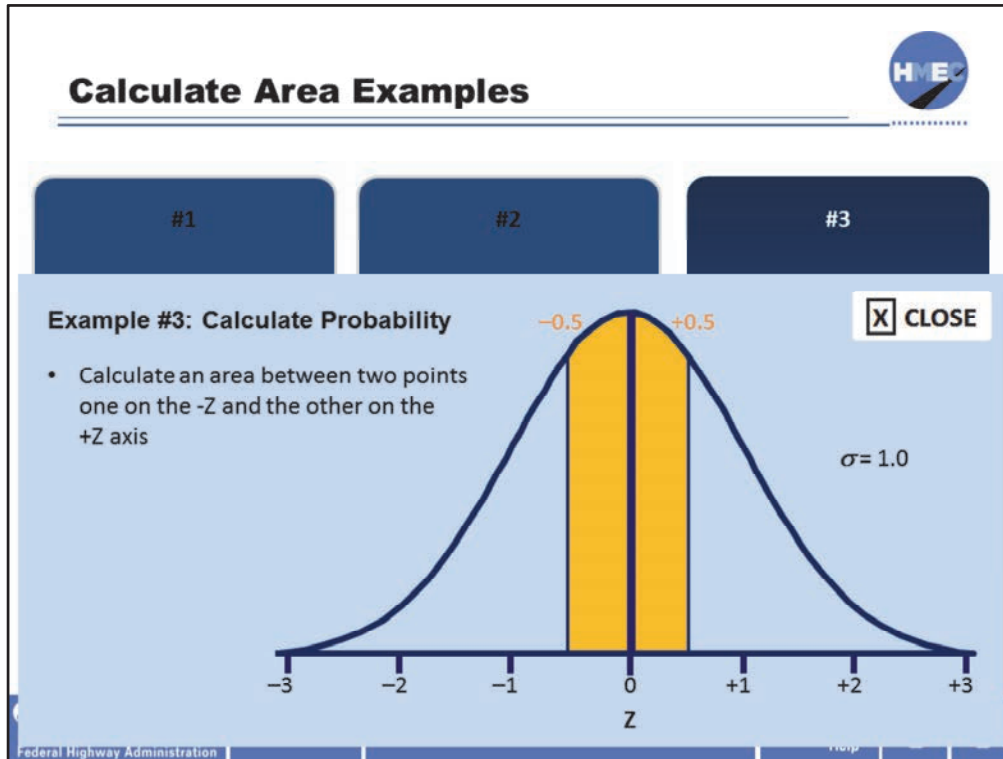
Example #2: Calculate Area

- Here is another example of determining the area under the normal curve between two points, the + Z axis



Here is another example of determining the area under the normal curve between two points, the + Z axis. We want to determine the area under the standard normal distribution between $Z = 0$ and $Z = + 0.33$. Using the table of areas under the normal curve, we look up a $Z = 0.33$ (Remember a + Z and - Z have the same value). You see the area is .1293 (12.93%).

Image description: Distribution curve.

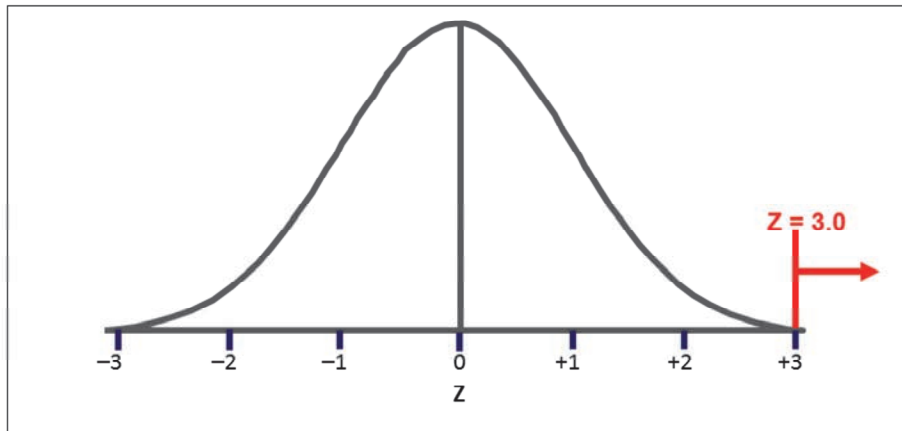


Now let's calculate an area between two points one on the -Z and the other on the +Z axis. We will also ask the question differently. If one Z value were randomly obtained from the standard normal distribution, what are the chances (that is, what is the probability) that it will be between $Z = -0.5$ and $Z = +0.5$?

Again we refer to the earlier statement, that said there are two ways of interpreting the probability. Remember, the areas have to be calculated from the mean value. Using the table of areas under the normal curve, we look up a $Z = 0.500$ (remember +Z and -Z have the same value). You see the area is .1915 (19.15%). This is for one side, but since the curve is symmetrical, we double the value (.1915 + .1915) to get .3830 (38.30%).

Image description: Distribution curve.

Example: Calculate Area



ANSWER



Determine the area under the standard normal distribution above $Z = +3.0$.



Now it's your turn. Look at the figure. We want to determine the area under the standard normal distribution above $Z = +3.0$.

Select the box for the answer.

Image description: Distribution curve.

Example: Calculate Area



Reading from the table of areas under the standard normal curve, $Z = 0$ to $Z = 3.0$ is .4987; this is the area to the left of the red line.

CLOSE

Since there is .5000 above the mean, $.5000 - .4987 = .0013$ (0.13%) above a $Z = 3.0$.

It is important to keep in mind that the area that lies above 3 Zs is a very small area. This knowledge will be useful later.

ANSWER



Determine the area under the standard normal distribution above $Z = +3.0$.



U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



Reading from the table of areas under the standard normal curve, $Z = 0$ to $Z = 3.0$ is .4987; this is the area to the left of the red line.

Since there is .5000 above the mean, $.5000 - .4987 = .0013$ (0.13%) above a $Z = 3.0$.

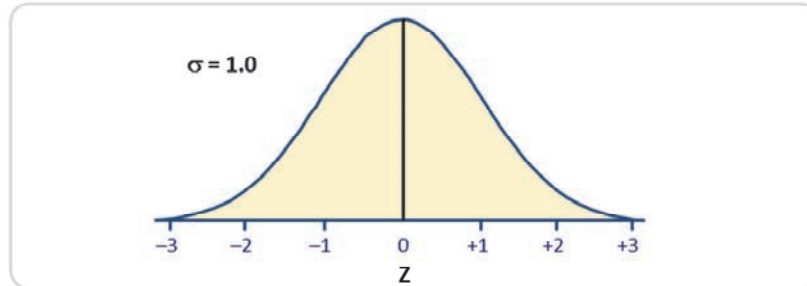
It is important to keep in mind that the area that lies above 3 Zs is a very small area. This knowledge will be useful later.

Image description: Distribution curve.

Exercise 2: Normal Distribution



- Become familiar with using Z values of 1.0, 2.0, and 3.0.; this information very useful as we proceed through this lesson, and in fact will define a “statistical rule”
- Use your table of areas under the standard normal curve to solve the three knowledge checks on the following screens



To open the exercises document, select the paperclip icon.



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

This exercise is for you to do to become familiar with using Z values of 1.0, 2.0. and 3.0. You will find this information very useful as we proceed through this lesson, and in fact will define a “statistical rule.”

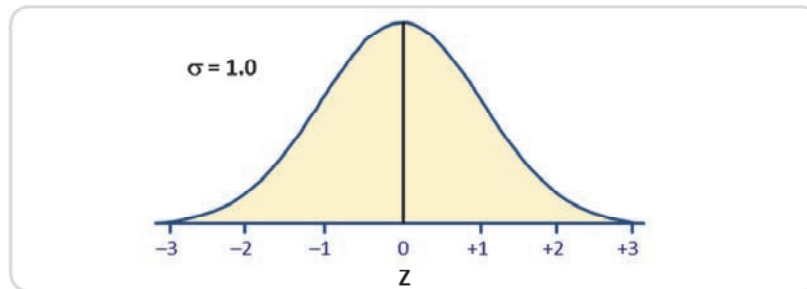
Use your table of areas under the standard normal curve to solve the following three knowledge checks.

Image description: Distribution curve.

Select the best answer. What percentage of the area under the standard normal distribution is between $Z = -1.0$ and $Z = +1.0$?



- a) 0.2456 (24.56%)
- b) 0.6826 (68.26%)
- c) 0.8826 (88.26%)



Knowledge Check

Try again

Submit

Clear



Select the best answer. What percentage of the area under the standard normal distribution is between $Z = -1.0$ and $Z = +1.0$?

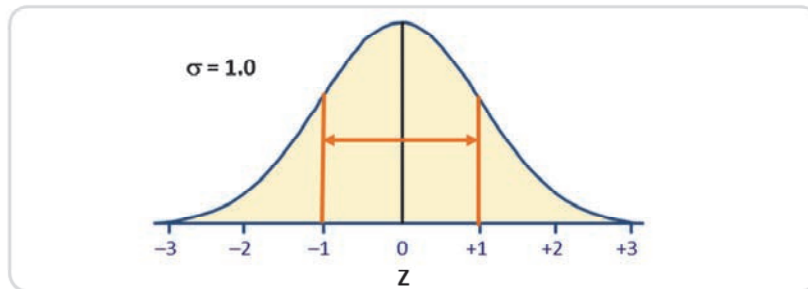
- a) 0.2456 (24.56%);
- b) 0.6826 (68.26%); or
- c) 0.8826 (88.26%).

Image description: Distribution curve.

Select the best answer. What percentage of the area under the standard normal distribution is between $Z = -1.0$ and $Z = +1.0$?



- a) 0.2456 (24.56%)
- b) 0.6826 (68.26%)
- c) 0.8826 (88.26%)



Knowledge Check Debrief



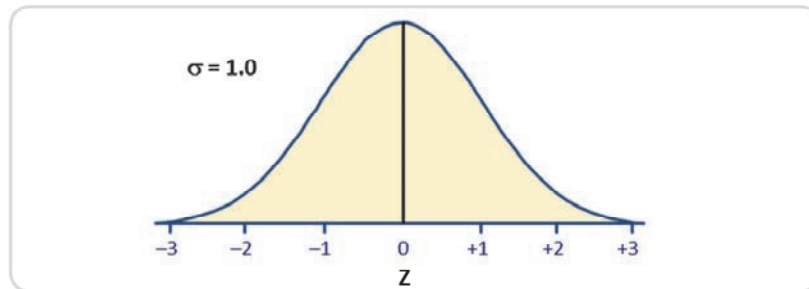
The correct answer is b) 0.6826 (68.26%).

Image description: Distribution curve.

Select the best answer. What percentage of the area under the standard normal distribution is between $Z = -2.0$ and $Z = +2.0$?



- a) 0.4426 (44.26%)
- b) 0.6826 (68.26%)
- c) 0.9544 (95.44%)



Knowledge Check

Try again

Submit

Clear



Select the best answer. What percentage of the area under the standard normal distribution is between $Z = -2.0$ and $Z = +2.0$?

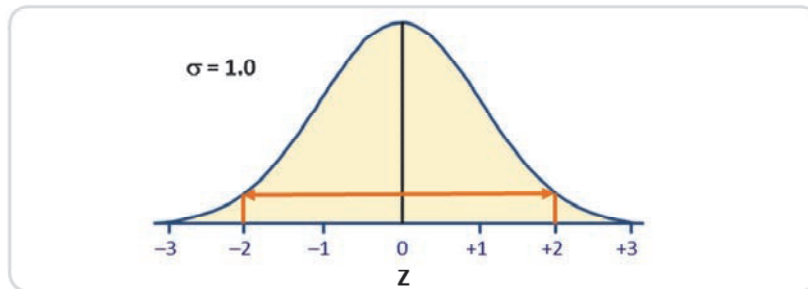
- a) 0.4426 (44.26%);
- b) 0.6826 (68.26%); or
- c) 0.9544 (95.44%).

Image description: Distribution curve.

Select the best answer. What percentage of the area under the standard normal distribution is between $Z = -2.0$ and $Z = +2.0$?



- a) 0.4426 (44.26%)
- b) 0.6826 (68.26%)
- c) 0.9544 (95.44%)



Knowledge Check Debrief



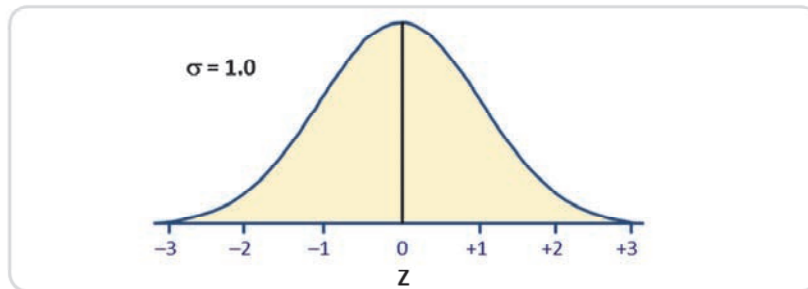
The correct answer is c) 0.9544 (95.44%).

Image description: Distribution curve.

Select the best answer. What percentage of the area under the standard normal distribution is between $Z = -3.0$ and $Z = +3.0$?



- a) 0.9974 (99.74%)
- b) 0.9544 (95.44%)
- c) 0.9826 (98.26%).



Knowledge Check

Try again

Submit

Clear



Select the best answer. What percentage of the area under the standard normal distribution is between $Z = -3.0$ and $Z = +3.0$?

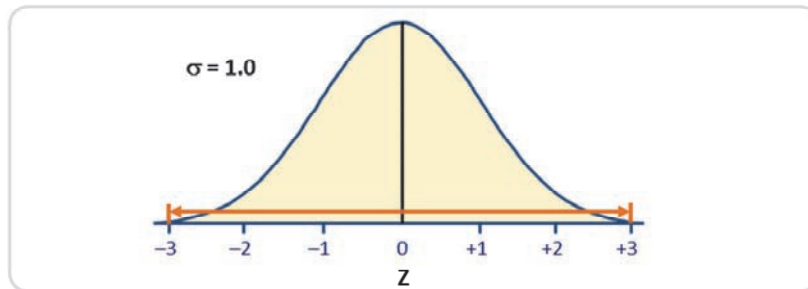
- a) 0.9974 (99.74%);
- b) 0.9544 (95.44%); or
- c) 0.9826 (98.26%).

Image description: Distribution curve.

Select the best answer. What percentage of the area under the standard normal distribution is between $Z = -3.0$ and $Z = +3.0$?



- a) 0.9974 (99.74%)
- b) 0.9544 (95.44%)
- c) 0.9826 (98.26%).



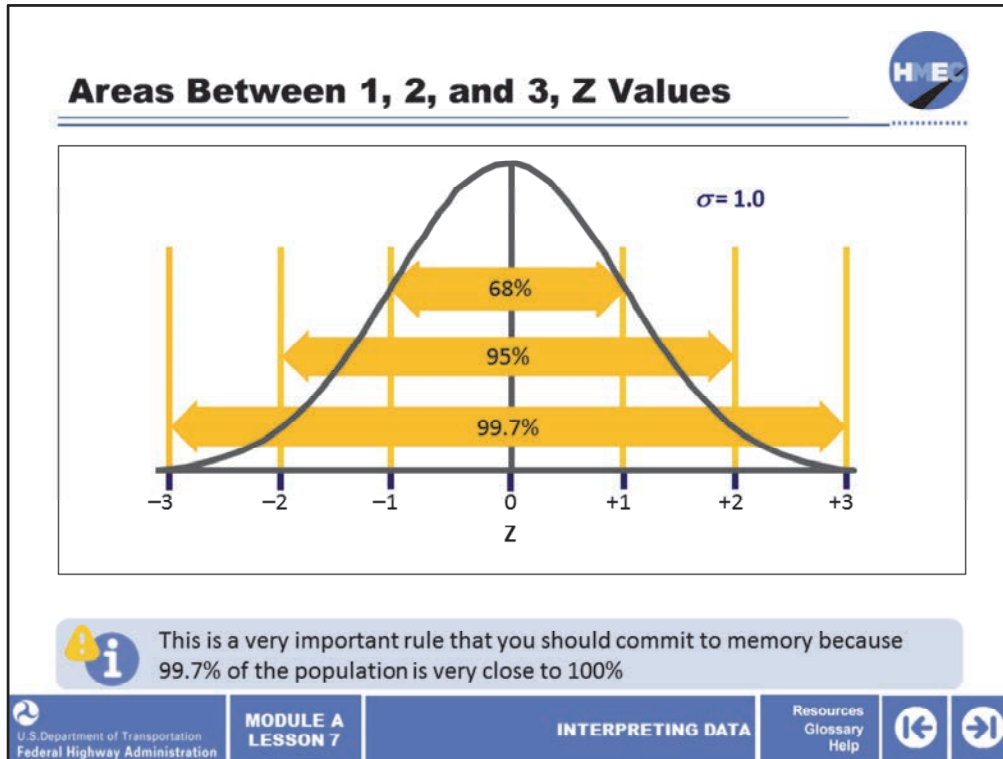
Knowledge Check Debrief



The correct answer is a) 0.9974 (99.74%).

These three knowledge checks give us 0.6826 (68.26%), 0.9544 (95.44%), 0.9974 (99.74%). This is what is called the 68-95-99.7 Rule.

Image description: Distribution curve.

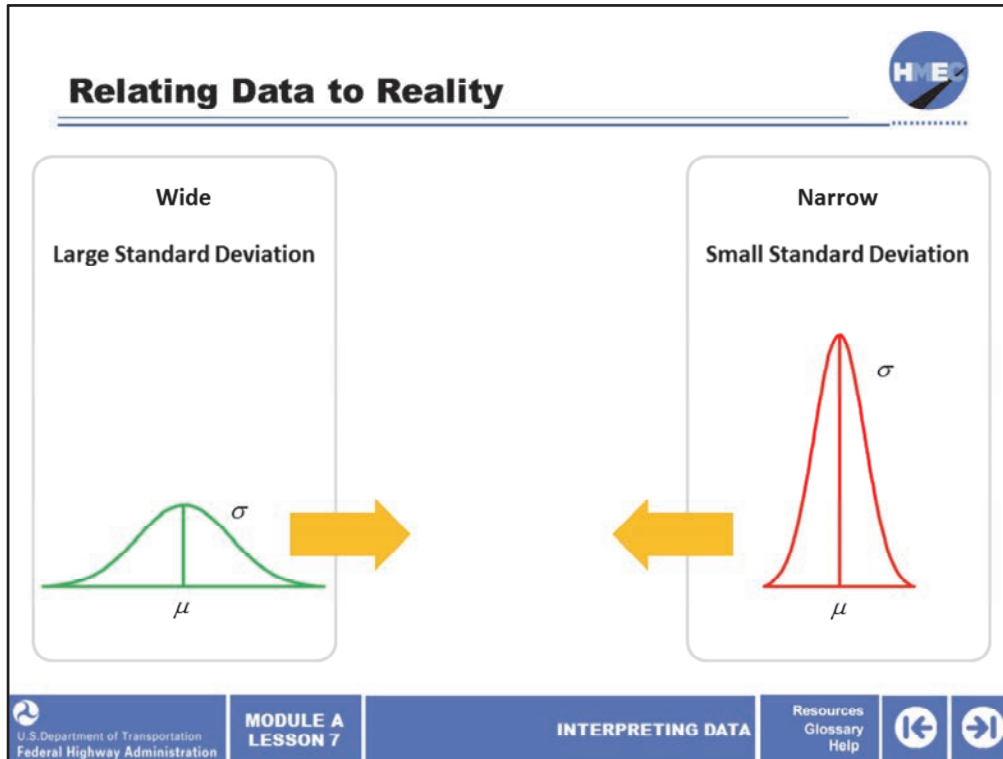


Using the areas you calculated in the previous exercise, we apply them to the standard normal curve.

1. What percentage of the area under the standard normal distribution is between $Z = -1.0$ and $Z = +1.0 = 68.26\%$, rounded to 68%?
2. What percentage of the area under the standard normal distribution is between $Z = -2.0$ and $Z = +2.0 = 95.44\%$, rounded to 95%?
3. What percentage of the area under the standard normal distribution is between $Z = -3.0$ and $Z = +3.0 = 99.74\%$, rounded to 99.7%?

These three areas constitute what is called the 68-95-99.7 rule. This is a very important rule that you should commit to memory. Also note that because 99.7% of the population is very close to 100%, we often say three Z limits constitute “all” of the population.

Image description: Distribution curve.



At this point you may question what the standard normal distribution is useful for, since as we said, it has a mean of zero and a standard deviation of 1.0. They are very unusual population properties. It was noted previously that there is a family of normal distributions that can all be transformed to the standard normal distribution.

In this screen, we see a red normal distribution that is narrow, that it has a small standard deviation and a green normal distribution that is wide. Both of these can be easily transformed in a standard normal distribution to which the table of areas under the standard normal distribution apply. This relationship allows the standard normal distribution to be applied to practical construction situations.

Image 1 description: Distribution curve.

Image 2 description: Distribution curve.

Calculating The “Z” Value



$$Z = \frac{(X - \mu)}{\sigma}$$

Equation for the Z Value

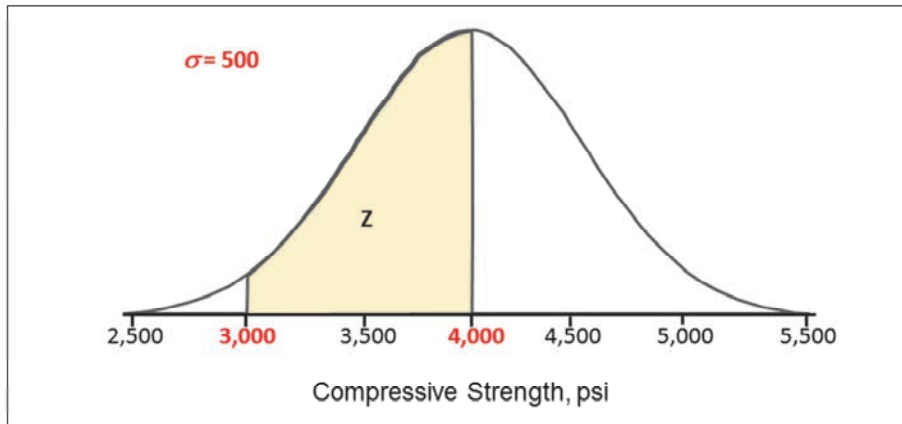
Z	Areas under the Standard Normal Curve
X	Point of Interest (Often a Spec Limit)
μ (mu)	Mean
σ (sigma)	Standard Deviation



We have used the Z value to find areas under the standard normal curve. Here we reintroduce the equation for the Z value, $Z = X$ minus μ (mu) divided by σ (sigma). X is the point of interest (often a spec limit), μ (mu) is the mean and σ (sigma) is the standard deviation. We will now see how any normal distribution can be transformed into a standard normal distribution. We remember that the Z-axis is in units of standard deviations from the mean. Therefore, a value from a normal distribution with any given mean and standard deviation can be transformed into the standard normal distribution by this equation.

Image description: Equation.

Example 1: Area Under the Normal Curve



What percentage of the distribution is between 3,000 psi and 4,000 psi?

ANSWER



U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA


Resources
Glossary
Help



Let's look at a situation where we have a concrete batching process that is assumed to follow a normal probability distribution with a mean of 4,000 psi and a standard deviation of 500 psi. Suppose we want to know what percentage of the distribution is between 3,000 psi and 4,000 psi, which is the green shaded area shown here.

Select the box to answer the question, what percentage of the distribution is between 3,000 psi and 4,000 psi?

Image description: Distribution curve.

Example 1: Area Under the Normal Curve 

$\sigma = 500$

$Z = \frac{(X - \mu)}{\sigma} \rightarrow Z = \frac{(3,000 - 4,000)}{500} = -2.0$

2,500 3,000 3,500 4,000 4,500 5,000 5,500
Compressive Strength, psi



Q&A What percentage of the distribution is between 3,000 psi and 4,000 psi? **ANSWER**

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help

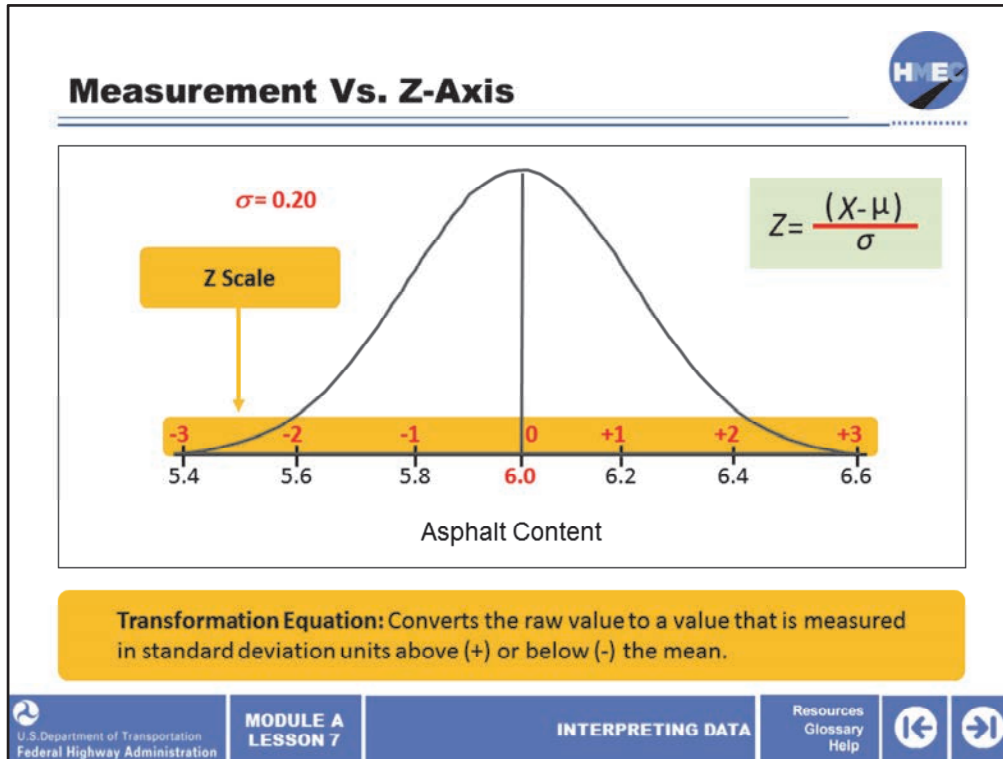
 

This is an example of calculating the area for a concrete compressive strength problem. Before we can use the table of areas under the standard normal distribution we have to calculate the Z value. The Z equation is: $Z = X$ minus μ (mu) divided by σ (sigma); substituting the mean and standard deviation into the equation $Z = (3,000 \text{ minus } 4,000)$ divided by $500 = \text{negative } 2.0$. We calculated the area for $Z = \text{negative } 2.0$ previously and found the area to be .4772 (47.72%).

Image 1 description: Equation.

Image 2 description: Equation.

Image 3 description: Distribution curve.



Here we see a normal distribution that represents the asphalt contents of a lot of asphalt concrete. It has a mean of 6.0% and a standard deviation of 0.20%. The transformation equation converts the raw value to a value that is measured in standard deviation units above (+) or below (-) the mean.

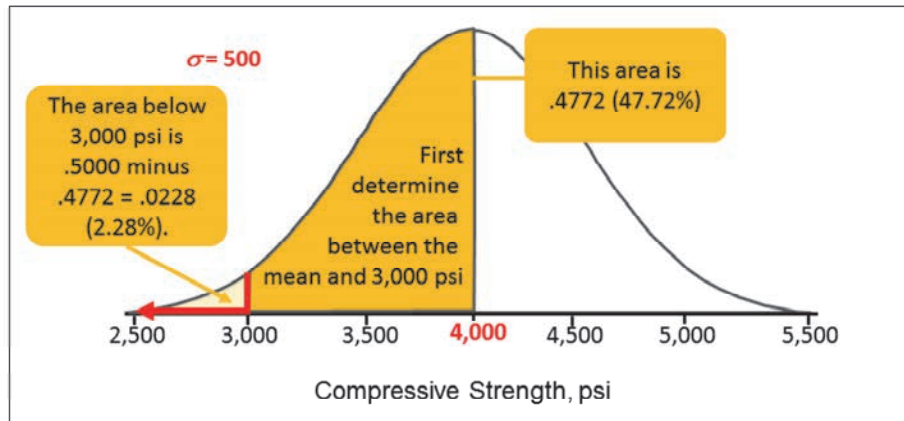
This is illustrated by the two horizontal scales shown on the figure. Once the Z value is known, the standard normal distribution can be used to determine areas and hence probabilities. Any value of asphalt content can also be represented as being a given number of standard deviations above or below the mean (shown on the Z axis).

Image description: Distribution curve.

Example 2: Calculate Probability



- The shaded area we are interested in is now the area below 3,000 psi, so we cannot calculate the area below 3,000 psi directly

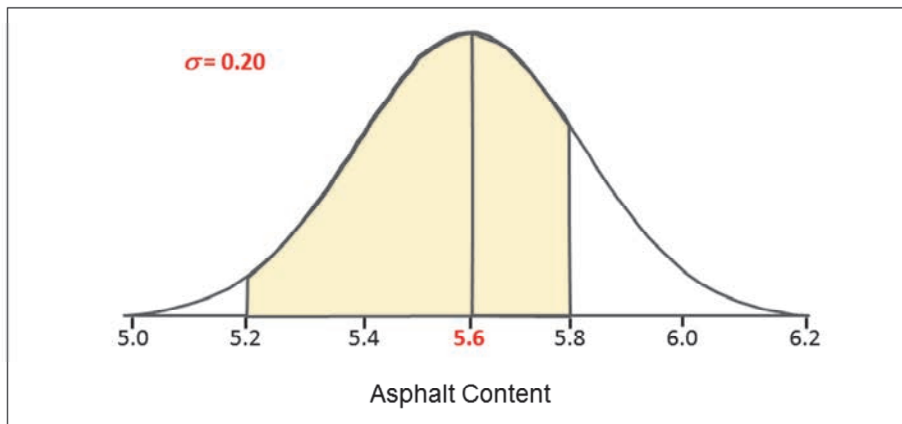


Here we have the same normal distribution as the previous screen but the shaded area we are interested in is now the area below 3,000 psi. This would be of interest if the specification limit for concrete strength were 3,000 psi. We know we always calculate Z and the area from the mean value. So we cannot calculate the area below 3,000 psi directly.

We first have to determine the area between the mean and 3,000 psi. We did this in the previous screen and found this area to be .4772 (47.72%). So the area below 3,000 psi is .5000 minus .4772 = .0228 (2.28%). If the specification limit for the concrete strength were 3,000 psi, we say there was 2.28% of the population outside the spec.

Image description: Distribution curve.

Example 3: Areas Under the Normal Curve



What percentage of the asphalt concrete will be within the allowable tolerance limits?

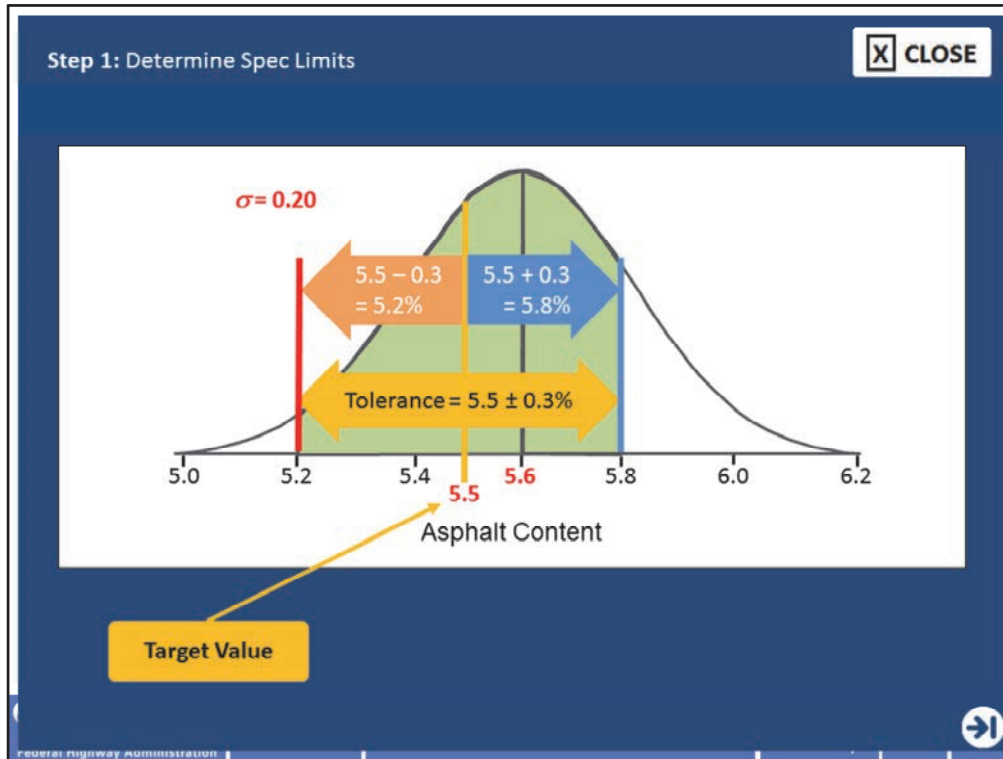
ANSWER



In this problem, we see another population of asphalt contents. A contractor is producing asphalt concrete with an asphalt content that is assumed to be normally distributed with mean = 5.6% and standard deviation = 0.20%. If the target value for the job mix formula (JMF) is 5.5%, with an allowable tolerance of $\pm 0.3\%$, we want to know what percentage of the asphalt concrete will be within the allowable tolerance limits.

Select the box to answer the question, what percentage of the asphalt concrete will be within the allowable tolerance limits?

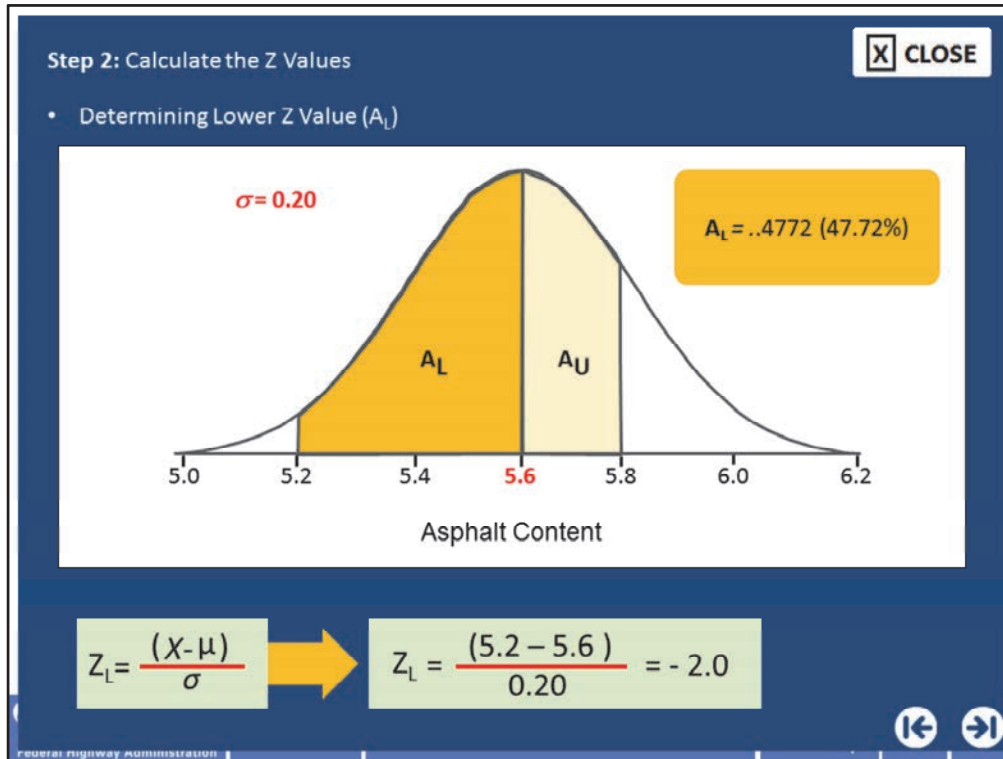
Image description: Distribution curve.



We cannot solve this problem until we determine the spec limits, so that is the first step. Since the target value is 5.5% with a tolerance of $\pm 0.3\%$, the lower spec limit is $5.5 - 0.3 = 5.2\%$ and the upper spec limit is $5.5 + 0.3 = 5.8\%$. So this establishes the green shaded area we see in the screen.

The next step is on the next screen.

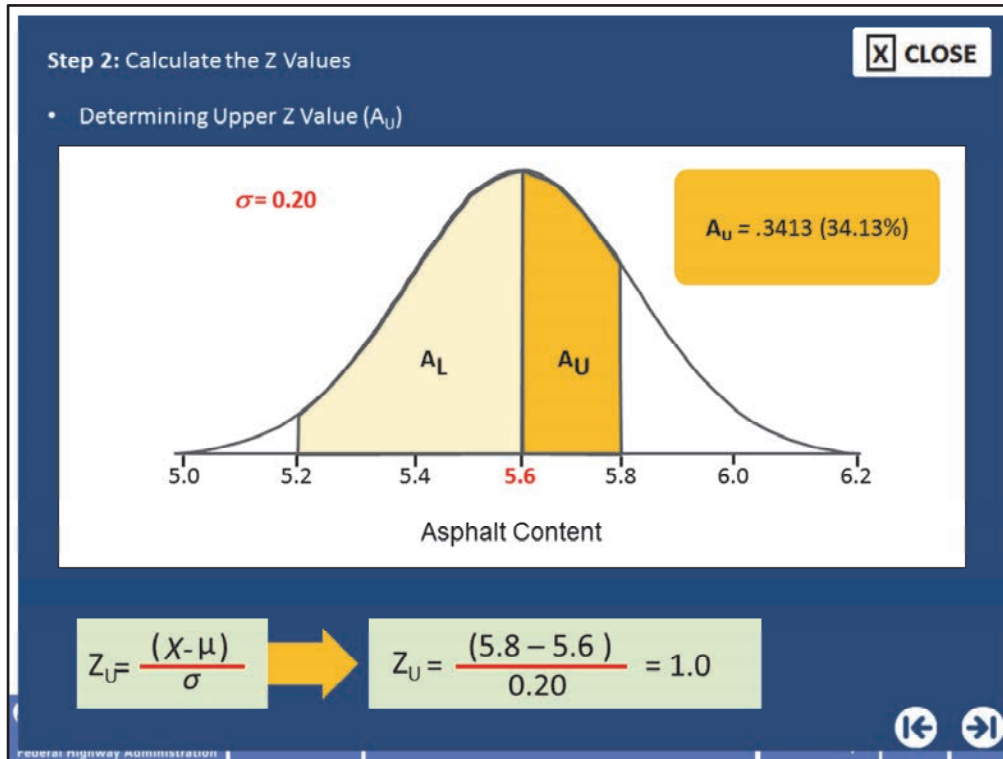
Image description: Distribution curve.



The next step is to calculate the Z values. Since we have two Z values and areas to find, we will designate the lower Z value and area Z_L and A_L and the upper Z value and area Z_U and A_U .

Z_L is X_L minus μ divided by σ equals $(5.2 - 5.6)$ divided by 0.20 equals -2.0 . So A_L (the area between 5.6 and 5.2) is $.4772$ (47.72%).

Image description: Distribution curve.

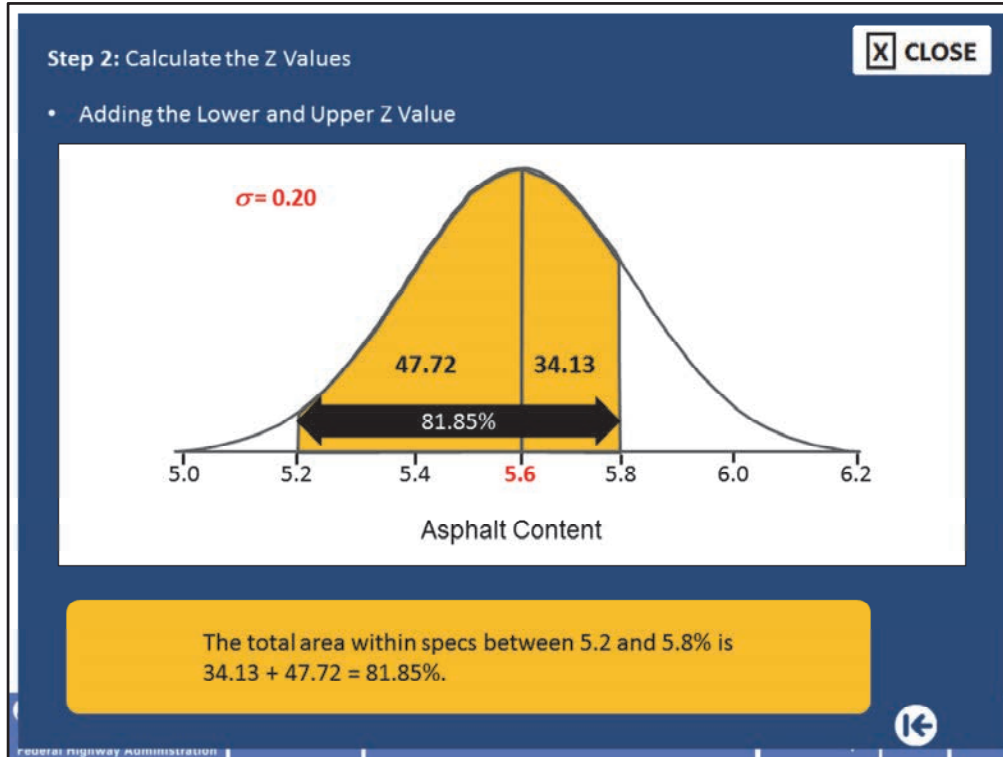


Z_U is X_U minus μ divided by σ equals $(5.8 - 5.6)$ divided by 0.20 equals 1.0 . So A_U (the area between 5.8 and 5.6) is $.3413$ (34.13%).

Image 1 description: Distribution curve.

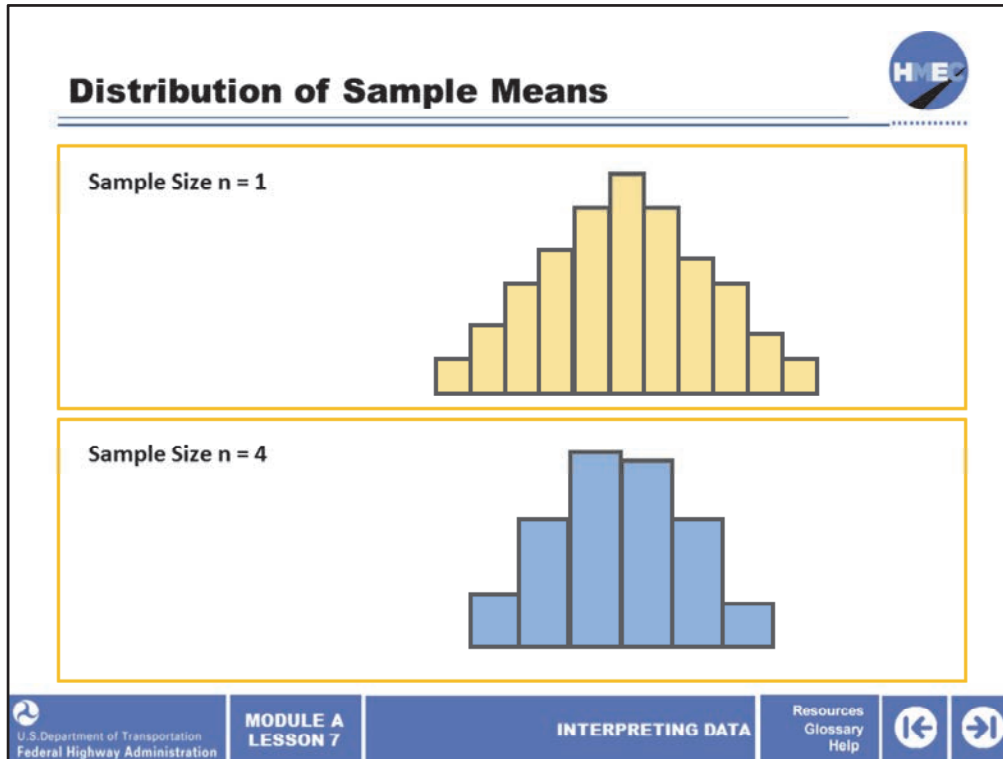
Image 2 description: Equation.

Image 3 description: Equation.



Thus the total area within specs between 5.2 and 5.8% is $34.13 + 47.72 = 81.85\%$.

Image description: Distribution curve.



At this point, we will start the discussion of a very important concept. It is called the distribution of sample means, or just distribution of means.

If we sample a population taking a sample size of one, that is $n=1$, say 40 times and plot the results, we will get a distribution that generally reflects that of the population. Let's assume the population has a mean, μ , of 50; then the sample mean \bar{X} will be around 50, a little higher or a little lower (assuming we sample correctly and randomly). If the population has a standard deviation, σ , of about 9, the sample standard deviation, s , will also be about 9, maybe a little higher or lower. But the sample statistics will be a close representation of the population parameters.

But suppose using the same population we take samples of say, size 4, that is $n=4$, 40 times and plot the averages, we will get a different looking distribution. The sample mean, \bar{X} , will be approximately the same as that of the population mean, μ , but the sample standard deviation, s , will be appreciably smaller than the population standard deviation, σ .

This would tend to imply that although the mean of a single sample of size $n = 4$ would not necessarily be equal to the population mean, it would have a better chance of being close to the population mean than would a single sample result of size $n = 1$.

Image 1 description: Histogram.

Image 2 description: Histogram.

Note that we are plotting frequency versus the chip number.

Sampling Experiment

- Samples from a bag that contains 10 chips of equal size and weight

Population Mean = 5.5

Population Range = 9

Frequency

Population

1

1 2 3 4 5 6 7 8 9 10

ANSWER

Q&A What are the differences if you draw 100 samples of size $n = 1$ and 100 samples of size $n = 4$?

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help

This screen and those that follow show a sampling experiment that helps explain the concept of the distribution of means. This is an explanation of a computer-simulated sampling experiment. We performed a simple sampling experiment to illustrate some important properties of the distribution of sample means.

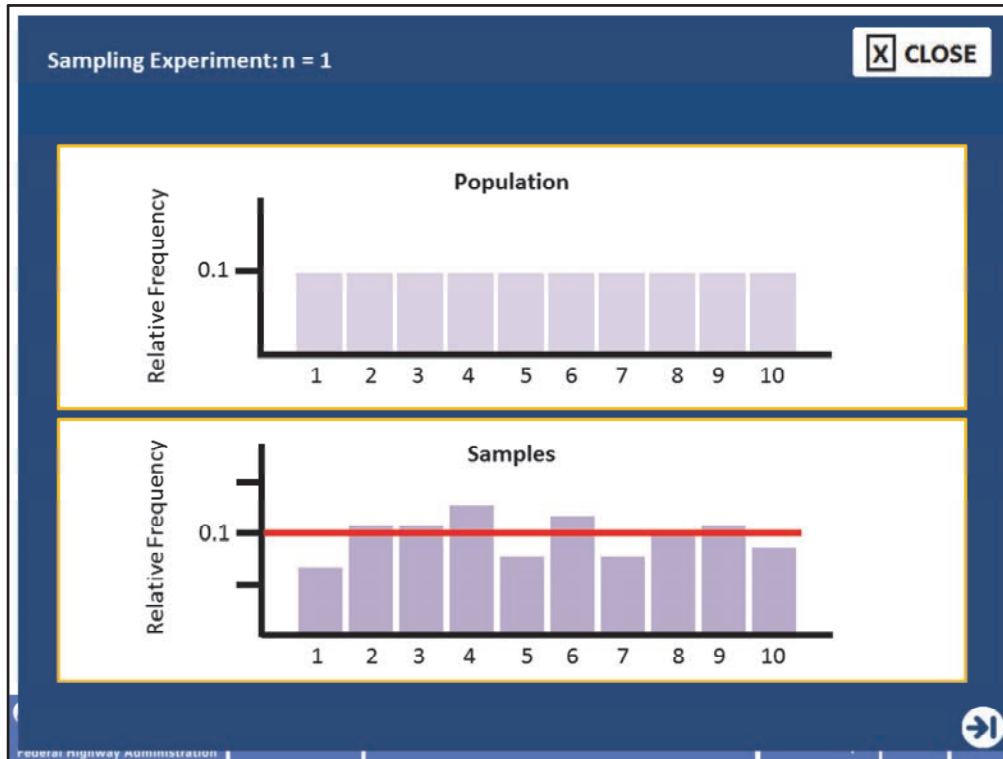
We sampled from a bag that contains 10 chips of equal size and weight. Each chip had a number on it. The 10 chips had the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10. Thus, the population mean is 5.5 and the population range is 9.

A histogram of the population of chips in the bag is shown in here. Every chip has a frequency of one and since there are 10 chips, the relative frequency is 1 divided by 10 or .10.

This is an example of a uniform distribution that we discussed earlier in the lesson. (Make special note that we are plotting frequency versus the chip number.)

Select the box to answer the question, what are the differences if you draw 100 samples of size $n = 1$ and 100 samples of size $n = 4$?

Image description: Histogram.



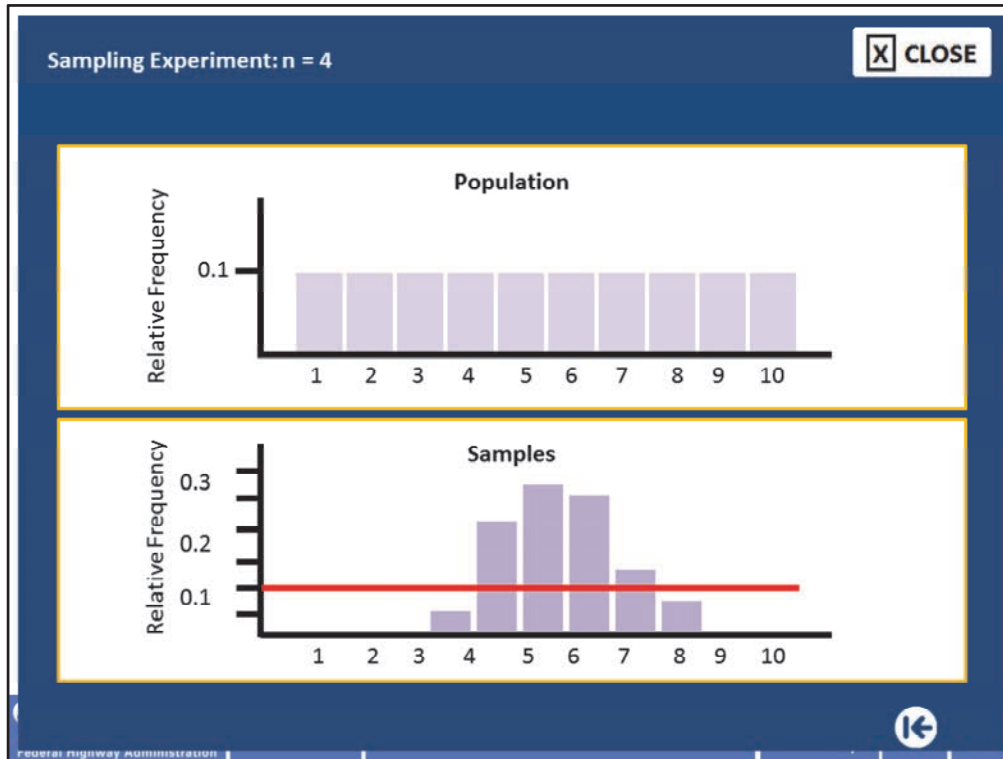
Here we introduce an example of the concept of distribution of means. We first do a sampling experiment using a sample size, $n = 1$. The process for this sampling experiment is to randomly draw one chip from the bag, record the number, and then replace the chip before drawing another chip. Two different sampling approaches are conducted.

First, 100 samples of size $n = 1$ were drawn. A histogram for these 100 samples of size $n = 1$ is shown on the screen, but notice we have to plot relative frequency versus the chip number so we can capture the difference between the sample and the population.

Two things are immediately obvious when the histogram for the samples is compared with the population histogram. First, the sample histogram has as much spread as the population. This is an indication that the variability of the sample generally matches that of the population. Second, with regard to shape (height), the sample histogram is roughly a “carbon copy” of the population histogram with some chips from the sample having a slightly higher frequency and some a slightly lower frequency than the population.

Image 1 description: Histogram.

Image 2 description: Histogram.



The second sampling approach was to draw a sample size of $n = 4$ chips (replacing each individual chip before drawing the next one so as to not bias the population). We averaged the four numbers that were obtained, and then we drew the histogram for the 100 samples of size $n = 4$. The resultant histogram for the distribution of sample means is quite different from the population histogram.

First, it is not spread out as much. In this case, the means only go from between 3 and 4 to between 8 and 9. Also, the shape changes where there is a smaller occurrence of both low and high values. This confirms that samples of size $n = 4$ have a better chance of yielding a sample mean that is close to the population mean (5.5) than do samples of size $n = 1$, which extended from 1 to 10. Second, the shape of the distribution of sample means is quite different than that of the population histogram. Both of these points are discussed next.

Let's think about why this happens. Why are there more results near the center than at the extremes? If we take a single chip and it is a 1 and replace it, the chances of getting the 1 again is only 1 out of 10, 0.1 (10%). It is much more likely that another numbered chip will be drawn. What if we take a sample of size 4; what is the likelihood of getting the 1 four times? The probability is $(0.1)(0.1)(0.1)(0.1)$ which is 0.01%. No wonder there were no averages at the sample location of 1. Suppose the computer randomly chose these numbers for the sample size of 4: 1, 3, 5, and 9. The average of this particular sample would be 4.5, much closer to the population mean of 5.5.

Now you can start to see why there are more values toward the center than on either end.

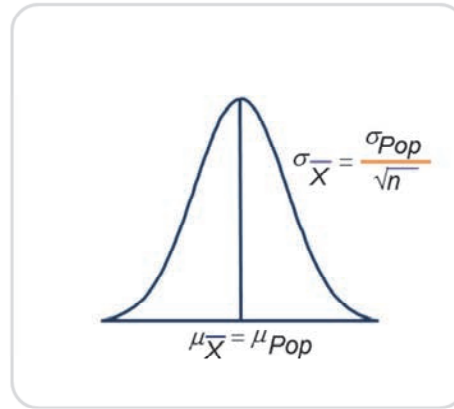
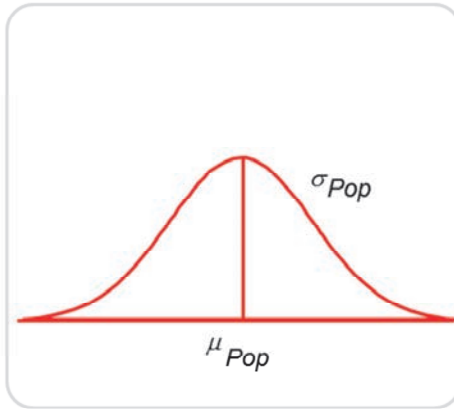
Image 1 description: Histogram.

Image 2 description: Histogram.

Distribution of Sample Means



- The mean and standard deviation of the distribution of sample means are related to the mean and standard deviation of the population




Getting back to the normal curve for the population and the distribution of means, we see here that the distribution of means also follows a normal distribution, but with a different shape (width). It has been proven mathematically that the mean and standard deviation of the distribution of sample means are related to the mean and standard deviation of the population. The mean of the distribution of samples means, $\mu_{\bar{X}}$, will be equal to the mean of the population, μ_{Pop} . We got an indication of that from the previous example of the chip sampling experiment. While it was not as obvious from the previous example, the standard deviation of the sample means $\sigma_{\bar{X}}$ will be equal to the population standard deviation divided by the square root of the sample size, this is σ_{Pop} divided by the square root of 4. It is important to understand that the sample size used here is the number of values used to calculate the sample mean, and not the number of sample means.

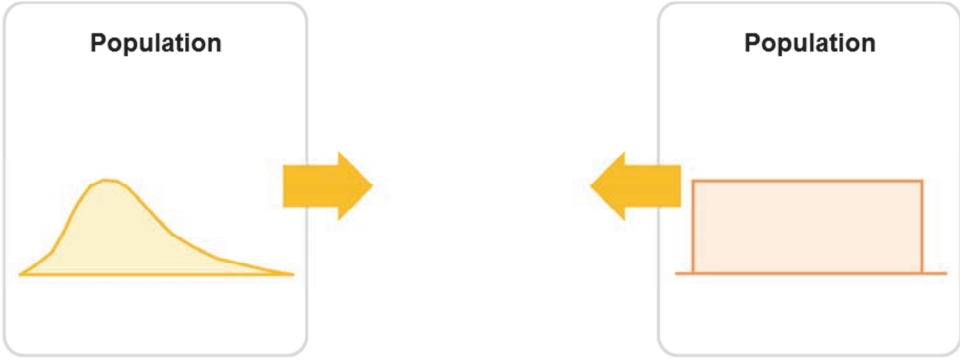
Image 1 description: Distribution curve.

Image 2 description: Distribution curve.

Central Limit Theorem



- The central limit theorem states that regardless of the shape of the population distribution, the means of samples of size n greater than 1 drawn from the population, will approach a normal distribution as n approaches infinity—this is true regardless of the shape of the population from which the samples are drawn




U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



So this leads up to the “central limit theorem.” The central limit theorem states that regardless of the shape of the population distribution, the means of samples of size n greater than 1 drawn from the population, will approach a normal distribution as n approaches infinity. This is true regardless of the shape of the population from which the samples are drawn.

As we saw in an earlier screen, the distribution of samples of size $n = 4$ starts to look normally distributed although it came from a uniform distribution. This theorem is extremely important, particularly if we may be dealing with a skewed distribution, which is more likely to occur than a uniform distribution. The shape of the distribution of sample means can be derived from the central limit theorem. Thus, the central limit theorem relationship to the distribution of sample means is one of the theoretical bases for the use of multiple sampling (that is, the use of sample sizes of n greater than 1).

Image 1 description: Distribution curve.

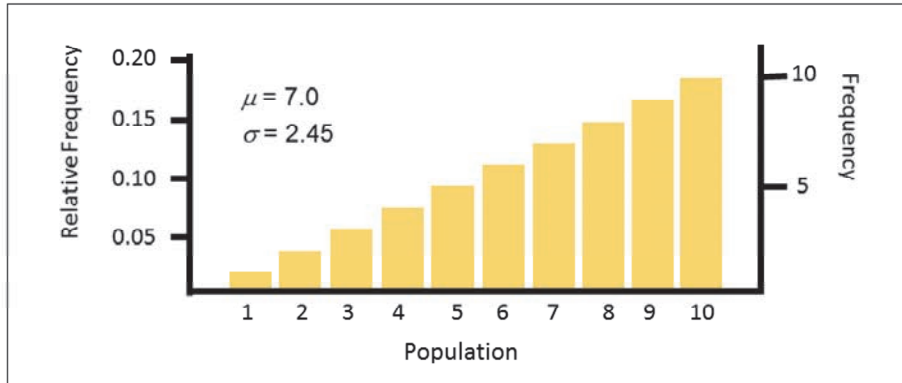
Image 2 description: Uniform distribution.

Central Limit Theorem Example



- Triangular distribution

Note that the shape of the population histogram is far from that of a normal distribution.



What are the differences if you draw 100 samples of size $n = 1$, $n = 5$, and $n = 10$?

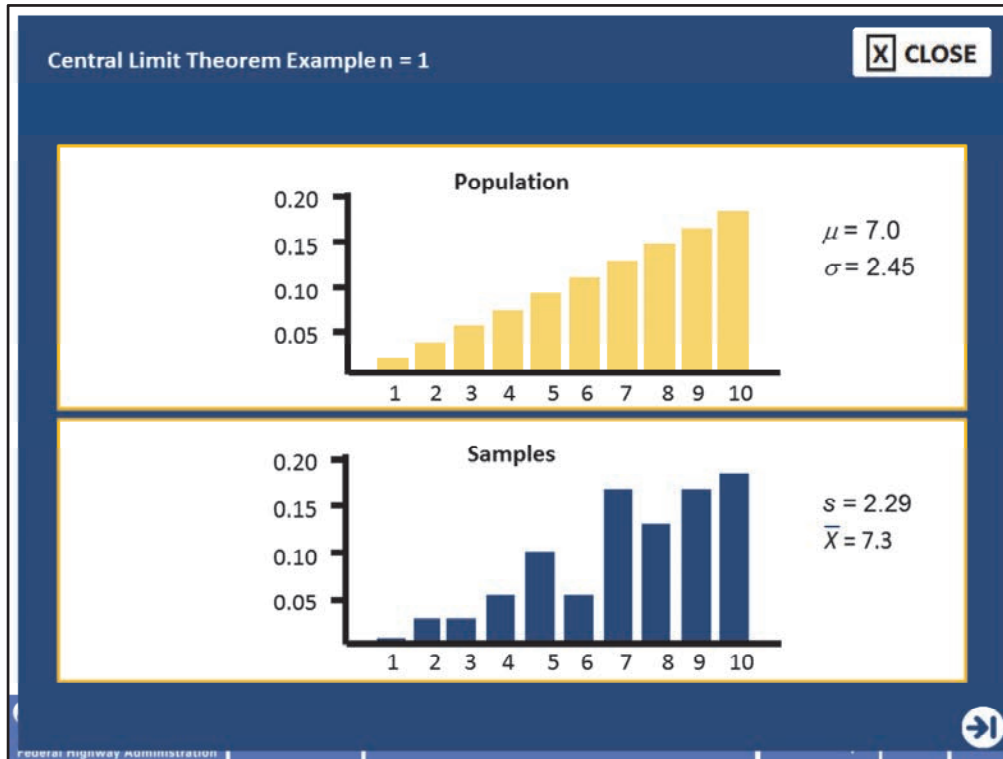
ANSWER



Here we see a triangular distribution. It is unlikely that any materials property would be distributed like this but this shape will emphasize that it would tend to be normal if larger sample sizes were used. As mentioned in the previous example, the concept of the central limit theorem is very important. On this screen we present another computer-simulated sampling experiment to illustrate the central limit theorem. In this experiment we draw 100 samples each of sizes $n = 1$, $n = 5$, and $n = 10$ from the triangular shaped distribution shown by the histogram. Note that the shape of the population histogram is far from that of a normal distribution. The mean of the population is 7.0, while the population standard deviation is 2.45.

Select the box to answer the question: What are the differences if you draw 100 samples of size $n = 1$, $n = 5$, and $n = 10$?

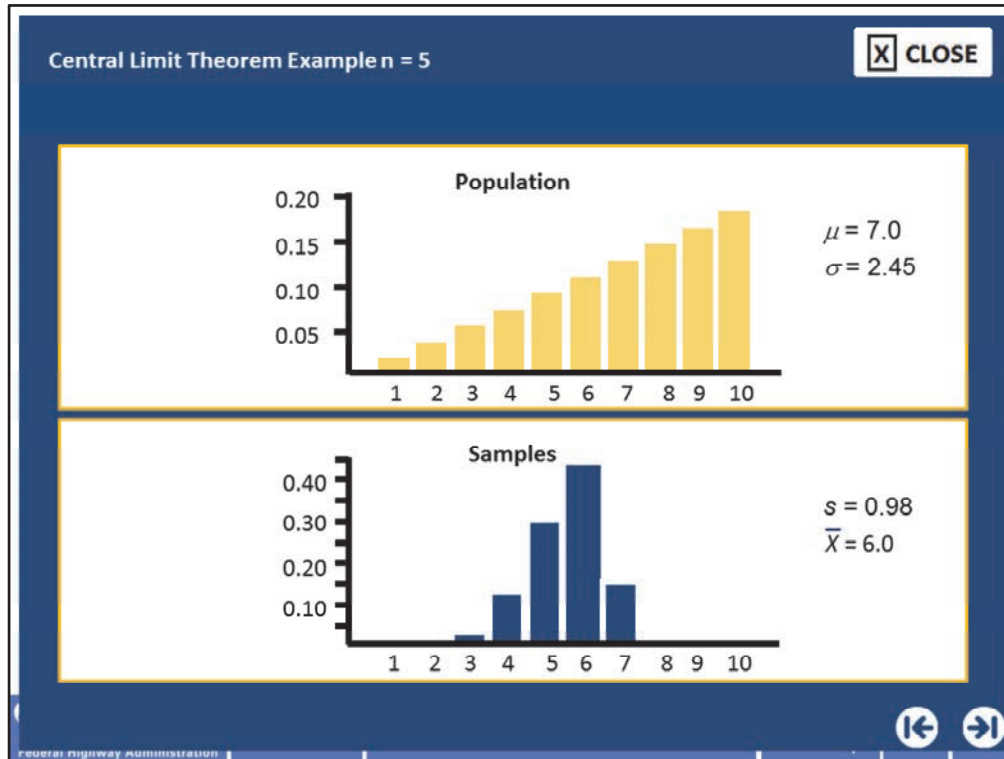
Image description: Histogram.



Continuing the example of the concept of the central limit theorem, we first examine the 100 samples of size $n = 1$ that were randomly drawn from the population shown at the top of the screen. The results for those 100 samples are shown at the bottom of the screen. It can be seen that the shape of the histogram for the 100 individual samples of size $n = 1$ generally follows the shape of the population from which the samples were drawn. There are some deviations from the population, particularly around a value of 6; however, given that 100 samples is not a particularly large number from a statistical standpoint, the results still indicate that samples of size $n = 1$ will have the same general distribution as the population from which they were drawn. Furthermore, the mean and standard deviation of the 100 samples of size $n = 1$ is about the same as the population from which they were drawn. This will always be the case for the distribution of samples of size $n = 1$.

Image 1 description: Histogram.

Image 2 description: Histogram.



Continuing the example of the concept of the central limit theorem, we next draw 100 samples of size $n = 5$ from the population at the top of the screen. For each sample, the mean of the five individual values was determined and recorded. These 100 sample means of size $n = 5$ were then plotted to form the histogram shown at the bottom of the screen. It can be seen that the histogram for the distribution of sample means of size $n = 5$ varies markedly from the population from which the samples were drawn.

From the previous discussion in the earlier screens, we know the following:

- That the distribution of sample means should approach a normal distribution,
- That the mean of the distribution of sample means should be the same as the population mean, and
- That the standard deviation of the distribution of sample means should equal the population standard deviation divided by the square root of n .

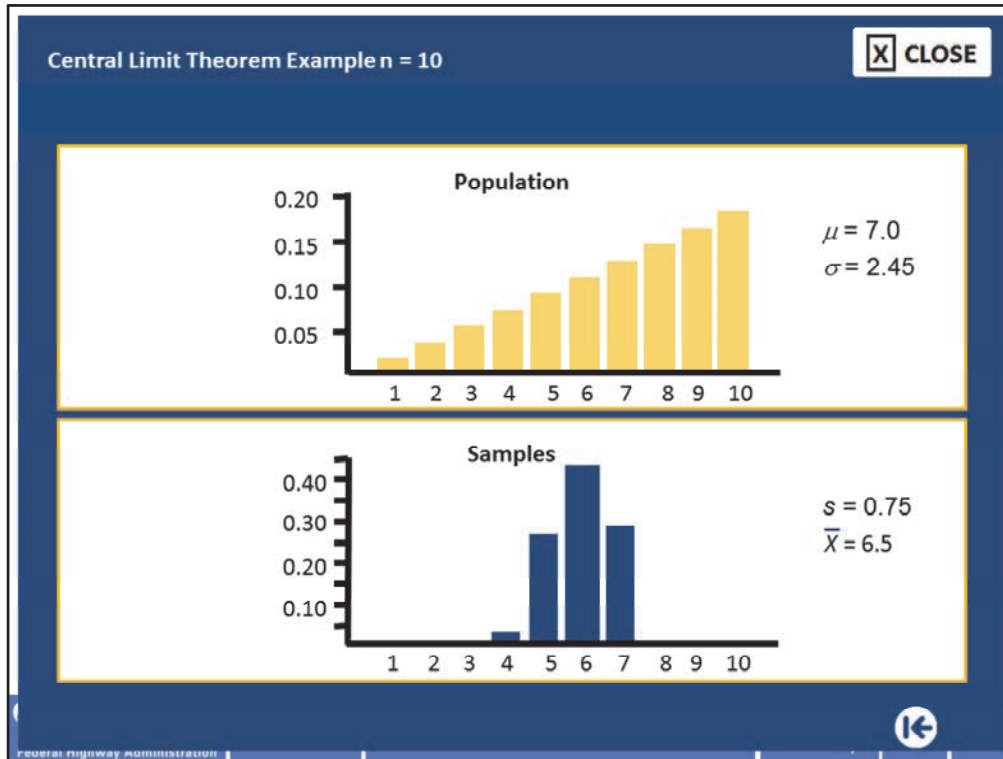
So for the first bullet, the shape of the sample distribution approaches normality, certainly more so than the population. For the second bullet, the population average was 7.0 and the sample average was 6.0, not exactly equal but not too far off. For the third bullet, the population standard deviation is 2.45 and the sample size is $n = 5$.

Carrying the calculation to its conclusion, we divide the sigma sub Pop, 2.45 by the square root of 5 which is 2.236. So 2.45 divided by 2.236 equals 1.096. We would expect the standard deviation (calculated from the population parameter) to be 1.10 (rounding) and

the sample experiment produced a standard deviation of 0.98. So how well do you think the histogram for the 100 sample means of sample size 5 meets these expectations?

Image 1 description: Histogram.

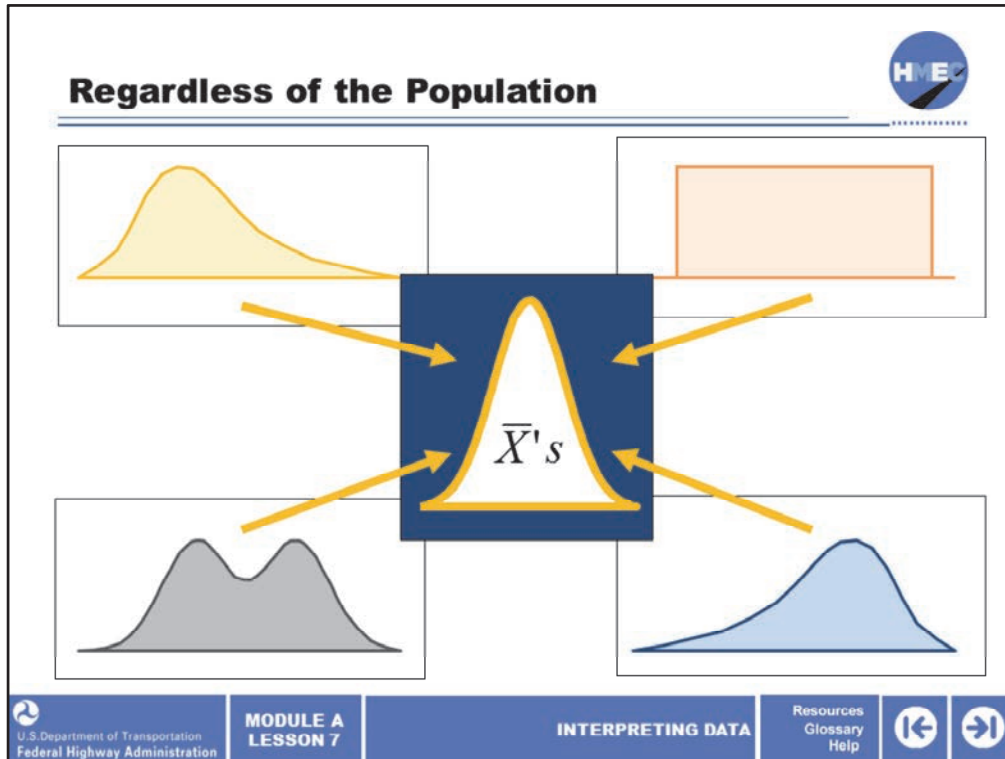
Image 2 description: Histogram.



Next, 100 samples of size $n = 10$ were drawn from the population. For each sample the mean of the 10 individual values was determined and recorded. These 100 sample means of size $n = 10$ were then plotted to form the histogram shown at right. This histogram appears to be even more normally shaped than the case where $n = 5$. Also, as the sample size has increased, the mean of the sample means has more closely approached the mean of the population. The standard deviation for the sample means is 0.75, which is almost exactly equal to the theoretical value of 0.77 (2.45 divided by the square root of 10).

Image 1 description: Histogram.

Image 2 description: Histogram.



Regardless of the shape of the population, it can be converted to a normal distribution by using averages of samples of size n greater than 1. From the relationships stated above, as the sample size increases, the spread of the distribution of sample means decreases. This is one justification for using a sample size of n greater than 1. Also, because of the central limit theorem, using sample averages allows us to approximate the distribution of sample means by the normal distribution without knowing the shape of the population distribution.

Image 1 description: Distribution curve.

Image 2 description: Uniform distribution.

Image 3 description: Distribution curve.

Image 4 description: Distribution curve.

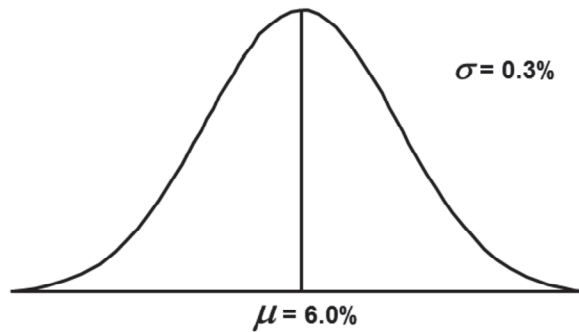
Image 5 description: Distribution curve.

Example: Applying the Distribution of Means



- Based on this sample mean, how likely is it that this sample can form this population?

For sample of $n = 4$, $\bar{x} = 6.5\%$



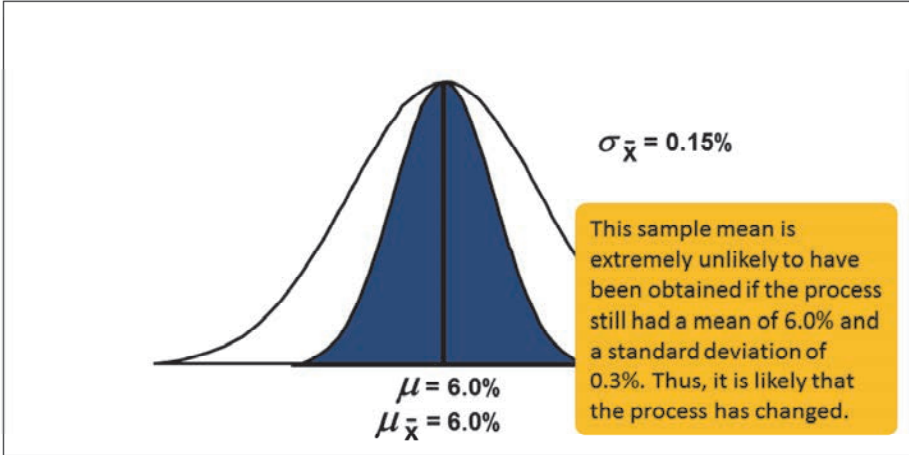
This is a statement of a practical problem of applying the distribution of means in which the question asked: Has the process average changed? Assume that an asphalt plant is supplying a mix that has an asphalt content with a mean of 6.0% and a standard deviation of 0.3%, and that it is assumed that the population is normally distributed.

After a plant breakdown, the inspector decides to take a sample of size $n = 4$ and because he/she knows that sample averages estimate the population average better than individuals, decides to use the sample average to determine if the process has changed. The sample mean is 6.5%. Based on this sample mean, how likely is it that this sample can form a population with a mean of 6.0%?

Image description: Distribution curve.

Example: Applying the Distribution of Means HME

- Remember we are still working with the distribution of means
- The solution to the problem is:



U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help

←
→

Remember we are still working with the Distribution of Means. The solution to the problem is:

- The distribution of sample means of samples of size $n = 4$ will follow a normal distribution;
- The mean of the samples will be equal to the population mean;
- The standard deviation of the sample means will be equal to the population standard deviation divided by the square root of 4. The standard deviation for the distribution of sample means would therefore be 0.30% divided by the square root of 4 which is 2. So the standard deviation of the sample means, $\sigma_{\bar{x}}$ is 0.15%;
- We learned from the 68-95-99.7 rule earlier in this lesson that if we go plus and minus three standard deviation from the means, we will include 99.7 percent of the population. These limits for the distribution of sample means would therefore be $6.0 \pm 3(0.15) = 5.55\%$ to 6.45%; and
- Our sample mean was 6.5; therefore, it is outside our three standard deviation limits. Calculating a Z value that corresponds to the sample mean of 6.5% leads to $Z = X \text{ minus } \mu \text{ divided by } \sigma_{\bar{x}}$ would be $Z = 6.5 \text{ minus } 6.0 \text{ divided by } 0.15 = 3.33$. Therefore, the sample mean is 3.33 standard deviations away from the population mean.


This sample mean is extremely unlikely to have been obtained if the process still had a mean of 6.0% and a standard deviation of 0.3%. Thus, it is likely that the process has changed.

Please note that we are calculating the standard deviation of the means here, not the

standard deviation of the sample.

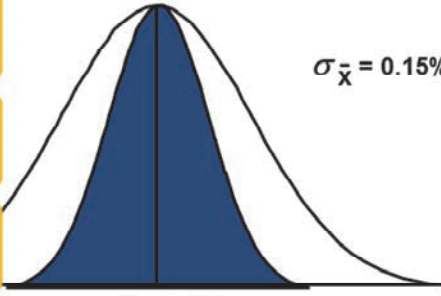
Image description: Distribution curve.

Exercise 3: Applying the Distribution of Means



- Looking at this same problem again

$$\sigma_{\bar{x}} = \frac{\sigma_{\text{Pop}}}{\sqrt{n}}$$





$$\sigma_{\text{Pop}} = 0.3\%$$

$$\sigma_{\bar{x}} = \frac{0.3}{\sqrt{4}}$$

$$\mu = 6.0\%$$

$$\mu_{\bar{x}} = 6.0\%$$



 To open the exercises document, select the paperclip icon.



MODULE A
LESSON 7

INTERPRETING DATA

Resources
 Glossary
 Help

Now let's take a moment for another exercise. Use the exercises document that is provided from the paperclip icon. Please allow 5 minutes to conduct this exercise.

Looking at this same problem again, we found the 3 σ limits that include 99.7 percent of the population to be $6.0 \pm 3(0.15) = 5.55\%$ to 6.45% . And we said that our sample mean was 6.5; therefore, it is outside our three standard deviation limits.

So, what is the probability of having a mean of $n = 4$ of 6.5%? To accomplish this, you need to:

1. Use the Z equation ($Z = (X - \mu) / \sigma_{\bar{x}}$) and to calculate a Z value that corresponds to the sample mean of 6.5%.
2. Then, go to the table of areas under the normal curve and look up the area that corresponds to that Z value.
3. Then, draw a conclusion of the likelihood of a sample average ($n = 4$) of 6.5% coming from the population with an average of 6.0% and standard deviation of 0.3%.
4. Next we will use the table of areas under the standard distribution to find the probability associated with this Z value.

Take some time to research this question and document your answers.

Image 1 description: Equation.

Image 2 description: Equation.

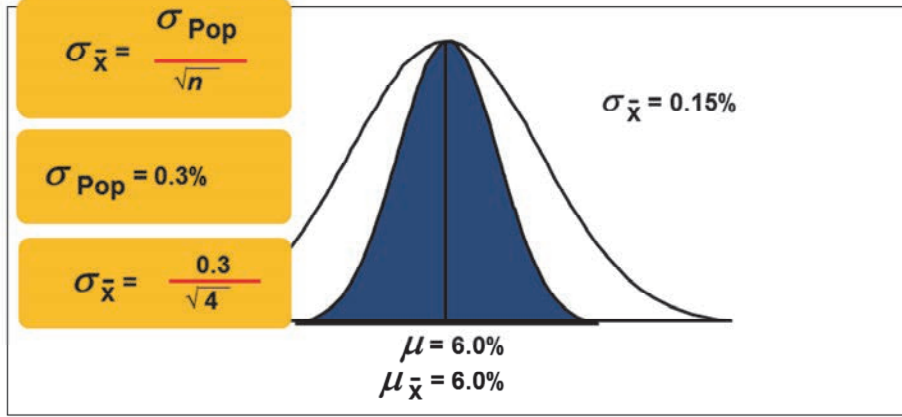
Image 3 description: Equation.

Image 4 description: Distribution curve.

Exercise 3: Recap



- Looking at this same problem again



The answer is:

$$Z = (X - \mu) / \sigma_{\bar{x}} = (6.5 - 6.0) / 0.15 = 3.3$$

Therefore, the sample mean is 3.33 standard deviations away from the population mean. And the area associated with this Z value is 0.4996.

This means there is only a $(0.5000 - 0.4996 = 0.0004)$.04% chance of the average of 6.5 coming from this population.

Image 1 description: Equation.

Image 2 description: Equation.

Image 3 description: Equation.

Image 4 description: Distribution curve.

Exercise 4: Developing Specs

Spec. Limits for: $n = 1$, $n = 4$, $n = 9$

$\sigma = 0.3\%$

$\mu = 6.0\%$

To open the exercises document, select the paperclip icon.

U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help

Now let's take a moment for another exercise. Use the exercises document that is provided from the paperclip icon. Please allow 5 minutes to conduct this exercise.

Here is an exercise to let you determine how the limits of a specification for different sample sizes might be established. The actual development of statistically based specifications is more complicated than the exercise presented here but should give you an idea of the process. The design and development of specifications are discussed in detail in the lesson on acceptance plans and risks.

Go to exercise 4 in the exercises document to determine the upper and lower specification limits using the following three scenarios.

1. The population for asphalt content is known to be normally distributed with mean = 6.0% and standard deviation = 0.3%. Based on this population, develop upper and lower specification limits within which all (99.7%) of asphalt content test results of sample size $n = 1$ must fall.
2. Develop upper and lower specification limits within which all (99.7%) the mean asphalt content of samples of size $n = 4$ must fall.
3. Develop upper and lower specification limits within which all (99.7%) the mean asphalt content of samples of size $n = 9$ must fall.

Answers:

- Spec limits for $n=1$: $\mu \pm 3\sigma = 6.0 \pm 3(0.3) = 6.0 \pm (0.9) = 5.1$ to 6.9

- Spec limits for $n=4$: $\sigma_{\text{sub } \bar{x}} = \sigma/\text{square root on } n = 0.3/2 = 0.15$: $\mu \pm 3\sigma_{\text{sub } \bar{x}} = 6.0 \pm 3(0.15) = 6.0 \pm 0.45 = 5.55 \text{ to } 6.45$
- Spec limits for $n=9$: $\sigma_{\text{sub } \bar{x}} = \sigma/\text{square root on } n = 0.3/3 = 0.10$: $\mu \pm 3\sigma_{\text{sub } \bar{x}} = 6.0 \pm 3(0.10) = 6.0 \pm 0.30 = 5.70 \text{ to } 6.30$

If you were a contractor, all three sets of limits are equal. There is no difference in terms of risk to the contractor among the limits.

Image description: Distribution curve.

Exercise 4: Recap



Spec. Limits for: $n = 1$, $n = 4$, $n = 9$

For $n = 1$
the 3σ limits for this sample size are $\mu \pm 3\sigma = 6.0 \pm 3(0.3) = 6.0 \pm 0.9 = 5.1\% - 6.9\%$. $\sigma = 0.3\%$

For $n = 4$
the $3\sigma_{\bar{x}}$ limits for this sample size are $\mu \pm 3\sigma_{\bar{x}} = 6.0 \pm 3(0.3/\sqrt{4}) = 6.0 \pm 3(0.3/2) = 6.0 \pm 3(.15) = 5.55\% - 6.45\%$

For $n = 9$
the $3\sigma_{\bar{x}}$ limits for this sample size are $\mu \pm 3\sigma_{\bar{x}} = 6.0 \pm 3(0.3/\sqrt{9}) = 6.0 \pm 3(0.3/3) = 6.0 \pm 3(.10) = 5.7\% - 6.3\%$



The answers are:

1. For $n = 1$, the 3σ limits for this sample size are $\mu \pm 3\sigma = 6.0 \pm 3(0.3) = 6.0 \pm 0.9 = 5.1\% - 6.9\%$.
2. For $n = 4$, the $3\sigma_{\bar{x}}$ limits for this sample size are $\mu \pm 3\sigma_{\bar{x}} = 6.0 \pm 3(0.3/\sqrt{4}) = 6.0 \pm 3(0.3/2) = 6.0 \pm 3(.15) = 5.55\% - 6.45\%$.
3. For $n = 9$, the $3\sigma_{\bar{x}}$ limits for this sample size are $\mu \pm 3\sigma_{\bar{x}} = 6.0 \pm 3(0.3/\sqrt{9}) = 6.0 \pm 3(0.3/3) = 6.0 \pm 3(.10) = 5.7\% - 6.3\%$.

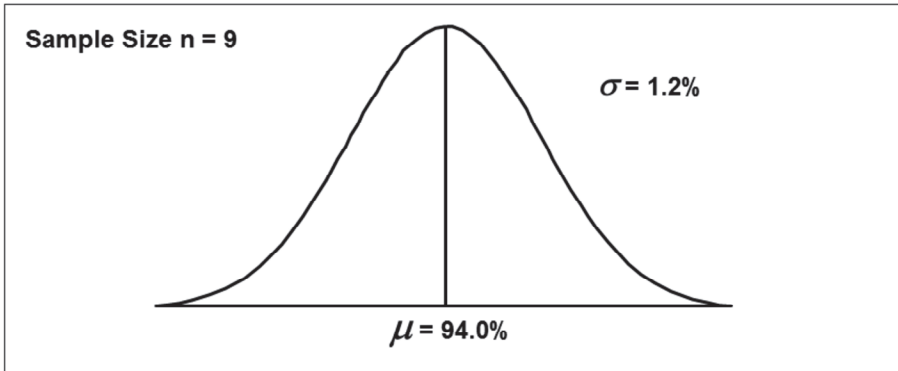
Since all of these are three sigma limits, the probability for the contractor of getting this material rejected (contractor's risk) is the same. That is, 0.13% on each end of the distribution, or 0.26%.

Image description: Distribution curve.

Exercise 5: Distribution of Means



- What is the probability that the mean of these nine test results will be between 93.4% and 94.4% of maximum theoretical density (MTD)?



To open the exercises document, select the paperclip icon.



Now let's take a moment for another exercise. Use the exercises document that is provided from the paperclip icon. Please allow 5 minutes to conduct this exercise.

Go to exercise 5 in the exercises document on the distribution of means. A contractor is building a project that includes 5 miles of asphalt pavement. Assume he will follow the same roller pattern throughout the project. After considerable testing on the first 2 miles of pavement, the contractor has established that his average in-place density, μ , is 94.0% maximum theoretical density (MTD) with a standard deviation, σ , of 1.2% MTD.

At the beginning of the third mile, the agency inspector requests that a new control strip be built. A total of nine nuclear density readings are obtained from the new control strip at random locations. What is the probability that the mean of these nine test results will be between 93.4% and 94.4% MTD?

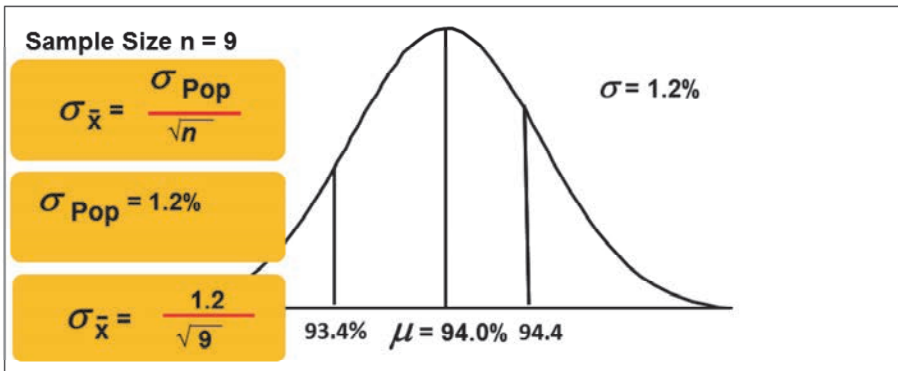
Take some time to research this question and document your answer.

Image description: Distribution curve.

Exercise 5: Recap



- What is the probability that the mean of these nine test results will be between 93.4% and 94.4% of maximum theoretical density (MTD)?



To open the exercises document, select the paperclip icon.



The distribution of sample means of samples of size $n = 9$ will follow a normal distribution. The mean of the samples will be equal to the population mean. The standard deviation of the sample means will be equal to the population standard deviation (1.2%) divided by the square root of 9. The standard deviation for the distribution of sample means would therefore be 1.2% divided by the square root of 9. So the standard deviation of the sample means, $\sigma_{\bar{x}}$ is 0.4%.

The next step is to calculate the Z values. Since we have two Z values and areas to find, we will designate the lower Z value and area Z_L and A_L and the upper Z value and area Z_U and A_U .

Z_L is X_L minus μ divided by $\sigma_{\bar{x}}$ equals $(93.4 - 94.0)$ divided by 0.4 equals -1.5. Use the Z table for a $Z = 1.5$; so A_L (the area between 93.4 - 94.0) is .4332 (43.32%).

Z_U is X_U minus μ divided by $\sigma_{\bar{x}}$ equals $(94.4 - 94.0)$ divided by 0.4 equals 1.0. Use the Z table for a $Z = 1.0$; so A_L (the area between 93.4 - 94.0) is .3413 (34.13%).

The probability that the mean of these nine test results will be between 93.4% and 94.4% MTD is $43.32 + 34.13 = 77.45\%$.

Image 1 description: Equation.

Image 2 description: Equation.

Image 3 description: Equation.

Image 4 description: Distribution curve.

Learning Outcomes Review



You are now able to:

- Define statistical inference, probability, and probability distributions
- Identify various probability distributions
- Explain the importance of the normal distribution
- Relate the concept of probability to the calculation of risks

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.



U.S. Department of Transportation
Federal Highway Administration

MODULE A
LESSON 7

INTERPRETING DATA

Resources
Glossary
Help



You have completed Module A, Lesson 7: Interpreting Data. You are now able to:

- Define statistical inference, probability, and probability distributions;
- Identify various probability distributions;
- Explain the importance of the normal distribution; and
- Relate the concept of probability to the calculation of risks.

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.