



Welcome to the Highway Materials Engineering Course Module A, Lesson 5: Organizing Data. This lesson provides an understanding of the basic elements of a statistically-based quality assurance (QA) program and includes an introduction to one phase of quality assurance: techniques for organizing 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:

- Create a frequency table using the guidelines provided
- Plot a frequency histogram using a given data set

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



This lesson will take approximately 40 minutes to complete.



U.S. Department of Transportation  
Federal Highway Administration

MODULE A  
LESSON 5

ORGANIZING DATA

Resources  
Glossary  
Help



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

- Create a frequency table using the guidelines provided; and
- Plot a frequency histogram using a given data set.

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

This lesson will take approximately 40 minutes to complete.

## Document Needed For This Lesson



Take a moment to download and print the Module A Lesson 5 Exercises PDF document by selecting the paperclip icon.

For this lesson, you will need the following exercises:

- Module A, Lesson 5: Exercise 1
- Module A, Lesson 5: Exercise 2



To open the exercises document, select the paperclip icon.



During this lesson, you will be prompted to reference the lesson exercise document. The referenced document 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



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



This lesson will cover the second phase: Organize the data. Statistics is a process where the data are organized in a meaningful way to understand trends, consistency, and variability. Organizing data provides information for writing and evaluating specifications and assembling data into systematic groups or classifications from which logical conclusions can be drawn.

## Is Collecting Data Enough?



- Data cannot provide useful information until it is organized in preparation for analysis

**Data or Just Numbers?**

---

22	34	54	26	45	56	72	29
55	23	70	82	35	23	19	92
54	62	16	34	88	45	61	34
33	71	28	36	62	33	91	18
71	46	34	76	52	22	49	60
27	81	37	92	55	68	14	41
66	33	41	57	99	28	75	42
44	23	75	88	61	16	27	59



In Lesson 4, we learned the importance of the proper way to collect data. Is that enough? No. Once the data are collected, we have a large group of measurements and test results. The data cannot provide any useful information until they are organized in preparation for analysis. This must be done in a form that is intelligible and understandable, otherwise they are just a collection of numbers. The human mind cannot easily comprehend a large series of separate facts or numbers.

Image description: Eight columns of eight random numbers in each column.

## Organization Formats



Tabular

VS

Graphical

- Data must first be organized into a format that will allow for an understanding of the trends that are present
- This is necessary if further analysis is to be conducted and inferences drawn from the sample data

Select each term to learn more.



U.S. Department of Transportation  
Federal Highway Administration

MODULE A  
LESSON 5

ORGANIZING DATA

Resources  
Glossary  
Help



For data to be meaningfully evaluated, they must first be organized into a format that will allow for an understanding of the trends that are present among the data. This is required if further analysis is to be conducted and inferences drawn from the sample data, which is necessary when we start to develop specifications. Data can be organized in two formats: tabular and graphical.

Select each term to learn more.

## Organization Formats



Tabular

VS

Graphical

Table 3.1

X CLOSE

A	B	C	D	E
3	7	5	8	9
5	9	6	5	8
4	8	6	7	8
5	7	5	6	7
4	8	5	8	9
3	6	6	7	9
5	6	7	8	9
4	7	6	7	8
5	8	5	7	9

Here is an example of a tabular format. Tabular is the simplest way to present the data and is usually one of the first steps in organizing them.

Tabular format is simply information presented in the form of a table with rows and columns. Most office productivity software programs, such as word processors and spreadsheets, include tools for entering text and data in tabular format. You can then enhance the appearance of the table by adjusting fonts, borders, and backgrounds, and other visual features.

Image description: Numbers in tabular format.



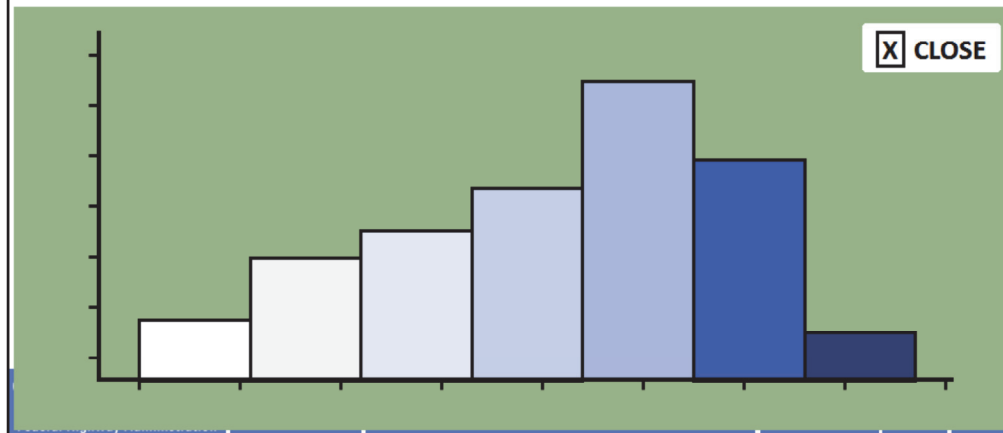
## Organization Formats



Tabular

VS

Graphical



Here is an example of a graphical format. A graphical form of your data not only provides a nice picture of the data, but it is also quick. Using a graphical form makes it much easier to see differences over time or some other factor. Also, looking at differences in data can be seen and viewed as profound, while just skimming text and numbers makes it easier to miss significant data.

Image description: Graphical format chart.

## Phase 1: Raw Air Content Data



2.1	4.1	3.4	3.7	7.3
5.0	5.2	4.0	2.0	5.3
4.8	5.9	5.4	4.7	5.1
6.0	7.1	1.9	5.2	6.8
4.9	6.0	4.2	6.2	5.8
3.0	5.0	3.9	7.7	6.3

Raw Air Content

- These data cannot easily determine:
  - Minimum
  - Maximum
  - Trends
- Is difficult to gather any summary information from this display




This table presents a set of 30 air content test results, in percent, for Portland cement concrete (PCC). These data are presented in the sequence they were collected, which is the most simplistic tabular format. However, you cannot easily determine the minimum, maximum, or trends in the data. They are presented as a set of individual test values.

The test results represent raw data, which means that they have not been grouped or organized. Presenting them in this fashion does not provide any specific information about the data or provide appropriate meaning in any fashion that will allow a better understanding of the results. It is difficult to gather any summary information from this display.

Image description: Table with a set of 30 air content test results.

## Phase 2: Organize Raw Air Content, % Data




<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>2.1</td><td>4.1</td><td>3.4</td><td>3.7</td><td>7.3</td></tr> <tr><td>5.0</td><td>5.2</td><td>4.0</td><td>2.0</td><td>5.3</td></tr> <tr><td>4.8</td><td>5.9</td><td></td><td></td><td></td></tr> <tr><td>6.0</td><td>7.1</td><td></td><td></td><td></td></tr> <tr><td>4.9</td><td>6.0</td><td>4.2</td><td>6.2</td><td>5.8</td></tr> <tr><td>3.0</td><td>5.0</td><td>3.9</td><td>7.7</td><td>6.3</td></tr> <tr><td colspan="5" style="text-align: center;">Raw Air Content</td></tr> </table>	2.1	4.1	3.4	3.7	7.3	5.0	5.2	4.0	2.0	5.3	4.8	5.9				6.0	7.1				4.9	6.0	4.2	6.2	5.8	3.0	5.0	3.9	7.7	6.3	Raw Air Content					<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;"><b>Smallest</b></td> <td colspan="4"></td> <td style="text-align: center;">Range (largest minus smallest) 7.7% - 1.9% = 5.8%</td> </tr> <tr> <td style="text-align: center;">1.9</td> <td>2.0</td><td>4.0</td><td>5.0</td><td>5.4</td><td>6.3</td> </tr> <tr> <td></td> <td>2.1</td><td>4.1</td> <td style="text-align: center;"><b>Middle</b></td> <td colspan="2" style="text-align: center;">Center of data about 5.0% or 5.1%.</td> </tr> <tr> <td></td> <td>3.0</td><td>4.2</td> <td style="text-align: center;">5.1</td> <td colspan="2"></td> </tr> <tr> <td></td> <td>3.4</td><td>4.7</td><td>5.2</td><td>6.0</td><td>7.3</td> </tr> <tr> <td></td> <td>3.7</td><td>4.8</td><td>5.2</td><td>6.0</td> <td style="text-align: center;">7.7</td> </tr> <tr> <td></td> <td colspan="4"></td> <td style="text-align: center;"><b>Largest</b></td> </tr> </table>	<b>Smallest</b>					Range (largest minus smallest) 7.7% - 1.9% = 5.8%	1.9	2.0	4.0	5.0	5.4	6.3		2.1	4.1	<b>Middle</b>	Center of data about 5.0% or 5.1%.			3.0	4.2	5.1				3.4	4.7	5.2	6.0	7.3		3.7	4.8	5.2	6.0	7.7						<b>Largest</b>
2.1	4.1	3.4	3.7	7.3																																																																										
5.0	5.2	4.0	2.0	5.3																																																																										
4.8	5.9																																																																													
6.0	7.1																																																																													
4.9	6.0	4.2	6.2	5.8																																																																										
3.0	5.0	3.9	7.7	6.3																																																																										
Raw Air Content																																																																														
<b>Smallest</b>					Range (largest minus smallest) 7.7% - 1.9% = 5.8%																																																																									
1.9	2.0	4.0	5.0	5.4	6.3																																																																									
	2.1	4.1	<b>Middle</b>	Center of data about 5.0% or 5.1%.																																																																										
	3.0	4.2	5.1																																																																											
	3.4	4.7	5.2	6.0	7.3																																																																									
	3.7	4.8	5.2	6.0	7.7																																																																									
					<b>Largest</b>																																																																									

**Organize Numbers:**  
Smallest to Largest

Does having the data organized in this fashion provide any greater meaning than what was provided in the previous slide?

**ANSWER**




U.S. Department of Transportation  
Federal Highway Administration

**MODULE A**  
**LESSON 5**

**ORGANIZING DATA**

Resources  
Glossary  
Help



For the air content data in the previous slide, the first step that could be taken in an effort to make the numbers more intelligible is to order the data from smallest to largest. This provides more information than was provided by the unsorted raw data.


The largest and smallest values and the maximum spread, or range of the data, can now be easily observed. The smallest is 1.9%, the largest is 7.7%, and the range is 7.7% - 1.9%, which is 5.8%.

The middle of the data also provides an estimate of the center of the data, which is about 5.0% or 5.1%. However, we are still faced with 30 separate facts to consider. Experience has shown that the mind grasps a pictorial presentation of data faster than any other form. Consequently, several methods of grouping data for presentation have been developed based upon charts, graphs, and curves.

Select the box to answer the question: Does having the data organized in this fashion provide any greater meaning than what was provided in the previous slide?

Image description: Table with a set of 30 air content test results.

## Phase 2: Organize Raw Air Content, % Data



2.1	4.1	3.4	3.7	7.3	2.0	4.0	5.0	5.4	6.3
5.0	5.2	4.0	2.0	5.3	4.8	6.0	4.9	3.0	7.3


**Smallest** 1.9

**Range (largest minus smallest)**  
 $7.7\% - 1.9\% = 5.8\%$

X CLOSE

Yes, it is easier to see the range of data and to get an idea of the center.


?
Does having the data organized in this fashion provide any greater meaning than what was provided in the previous slide?
ANSWER



MODULE A  
LESSON 5

ORGANIZING DATA

Resources  
Glossary  
Help



The answer is yes, it is easier to see the range of data and to get an idea of the center.

Image description: Table with a set of 30 air content test results.

**Relative Frequency =**  
frequency divided by  
total number of observations  
(or data points)

### Present Data: Frequency Table

1.9 A	3.9 B	4.9	5.3	6.2
2.0	4.0	5.0	5.4	6.3
2.1 B	4.1	5.0 C	5.8	6.8 D
3.0	4.2	5.1	5.9	7.1
3.4	4.7	5.2	6.0	7.3
3.7	4.8	5.2	6.0	7.7

Raw Air Content

Class	Class Limits	Frequency	Relative Frequency
A	0-2	2	0.07
B	2-4	6	0.20
C	4-6	16	0.53
D	6-8	6	0.20

In this example, the class limit (or class interval) was set to 2.

Each number in the raw air content table is a data point. In this example, there is a total number of 30 data points.

U.S. Department of Transportation  
Federal Highway Administration

**MODULE A**  
**LESSON 5**

**ORGANIZING DATA**

Resources  
Glossary  
Help

Using the same air content data from the previous slide, and arranging the numbers from lowest to highest, helped us to visualize the range and center but it was not very meaningful. However, after ordering the data from lowest to highest, they can now be easily organized and presented in a number of ways. Looking at the set of data of 30 air content test results, the first presentation method that we will consider is the frequency table. A frequency table shows the frequency of occurrence of the values of the variable (that is, air content) in ordered classes.


As its name implies, a frequency table is a tabular method: one that is a necessary first step for the graphical presentation methods discussed later. Each group of observations is called a class. The frequency for any class is the number of observations (frequency) of measurements falling within that class, while the relative frequency for any class is the frequency for that class divided by the total number of observations (data points).

For class A, the frequency is 2 and the total number of data points is 30, so the relative frequency is 2 divided by 30 = 0.066 or rounded, 0.07. The class limits are the values that determine the upper and lower limits of a class. The interval between two class limits is called the class interval. Although it is not shown on this slide, the sum of the relative frequencies should total 1.0 (100%).

In this frequency table, it is obvious that we have a conflict with the value 2. How do we decide which class a value, such as 2, be placed in, since it is shown in Class A and Class B? That will be clarified in the next slide when the frequency table guidelines are given.

Image description: Table with a set of 30 air content test results.

## Frequency Table Guidelines



- Step 1:** Determine the Range of Data
- Step 2:** Decide on Number of Classes
- Step 3:** Determine Class Interval
- Step 4:** Establish Class Limits
- Step 5:** Ensure Classes Are Mutually Exclusive



Select each step to learn more.

U.S. Department of Transportation  
Federal Highway Administration

MODULE A  
LESSON 5

ORGANIZING DATA

Resources  
Glossary  
Help



Here are guidelines for constructing a frequency table. This is often a trial-and-error process and there is no singular right way of doing this. However, some frequency tables can provide a better visualization of the data than others.

The following guidelines illustrate the process and provide a pathway to frequency table development:

1. Determine the range of data;
2. Decide on the number of classes;
3. Determine the class interval;
4. Establish the class limits; and
5. Ensure that all classes are mutually exclusive.

Select each step to learn more.

**Frequency**

**Step 1: Determine the Range of Data** X CLOSE

- Range (largest minus smallest value)

<b>Smallest</b> 1.9	3.9	4.9	5.3	6.2
2.0	4.0	5.0	5.4	6.3
2.1	4.1	5.0	5.8	6.8
3.0	4.2	5.1	5.9	7.1
3.4	4.7	5.2	6.0	7.3
3.7	4.8	5.2	6.0	<b>Largest</b> 7.7

Range (largest minus smallest value)  
 $7.7\% - 1.9\% = 5.8\%$

U.S. Department of Transportation  
 Federal Highway Administration

The first step is to determine the range of data. The range of data is the largest value minus the smallest value. For this case, it is  $7.7\% - 1.9\% = 5.8\%$ .

Image description: Table with a set of 30 air content test results.



**Frequency**

**Step 2: Decide on Number of Classes** X CLOSE

- Number of classes
  - Trial-and-error process
  - Rule of thumb:
    - Five to 15 depending on the number of data points
  - We will choose eight

Class
1
2
3
4
5
6
7
8

Select each step to learn

U.S. Department of Transportation  
Federal Highway Administration

Step 2: Decide on the number of classes. This is a somewhat arbitrary decision but the concept is to present as clear an idea of the data as possible. Try to avoid a small number of classes, such as two or three that have such a high frequency that they do not give a true perspective of the spread. Likewise, avoid a large number of classes that may spread the data too much. This is part of the trial-and-error process. A rule of thumb is five to 15 classes depending on the number of data points. For this case, we will choose eight classes.

Table description: Class table.

**Step 3: Determine Class Interval** X CLOSE

Class interval is determined by:

- Dividing the range of values by the number of classes to be used
- Range (5.8) divided by class (8) = 0.725
- Round up (0.725 to 1.0)

Smallest				
1.9	3.9	4.9	5.3	6.2
2.0	4.0	5.0	5.4	6.3
2.1	4.1	5.0	5.8	6.8
3.0	4.2	5.1	5.9	7.1
3.4	4.7	5.2	6.0	7.3
3.7	4.8	5.2	6.0	7.7
Largest				

Range (largest minus smallest value)  
7.7% - 1.9% = 5.8%

Class
1
2
3
4
5
6
7
8

U.S. Department of Transportation  
Federal Highway Administration

Step 3: Determine the class interval. The class interval is determined by dividing the range of values by the number of classes to be used. To make the table as clear as possible, the class intervals should be the same width. This often requires rounding the number off. If this is necessary, always round up to ensure a sufficient range of classes to encompass all of the data points. The range of our data from Step 1 is 5.8; the number of classes from Step 2 is eight. So our class interval is  $5.8/8 = 0.725$ . This would be an awkward value to use, so to keep the intervals even, round off to 1.0 (1.0%). An interval of 0.75%, although not quite as “neat” would be another possible choice.

Table description: Class table.

Image description: Table with a set of 30 air content test results.

## Frequency

X CLOSE

**Step 4: Establish Class Limits**

- Establish the class limits so that the first lower limit (of the lowest class) is less than the lowest data value
- We can set the first lower limit at 1.0% so as to have class limits fall on integer values
  - Example:
    - 1.0 to 2.0
    - 2.0 to 3.0

Class Limits
1.0–2.0
2.0–3.0
3.0–4.0
4.0–5.0
5.0–6.0
6.0–7.0
7.0–8.0
8.0–9.0

Select each step to learn more

MO  
LE

Step 4: Establish the class limits so that the first lower limit (of the lowest class) is less than the lowest data value. The lowest data point is 1.9%, so the first lower class limit must be set below this point. Since we rounded up from 0.725% to 1.0% for the class intervals, we have some flexibility and can set the lower class limit at 1.0% so as to have class limits fall on integer values, i.e., 1.0 to 2.0, 2.0 to 3.0, etc. from Step 3.

Table description: Class limits table.

**Frequency**

**Step 5: Ensure Classes Are Mutually Exclusive** X CLOSE

- The class limits are defined so that a value can be placed in only one class
- We must ensure a value such as 2.0 is placed in only one class
- Establish the class limits halfway between two possible observations
  - Example:
    - 0.95 to 1.95
    - 1.95 to 2.95

Class Limits	Class Limits
1.0–2.0	0.95–1.95
2.0–3.0	1.95–2.95
3.0–4.0	2.95–3.95
4.0–5.0	3.95–4.95
5.0–6.0	4.95–5.95
6.0–7.0	5.95–6.95
7.0–8.0	6.95–7.95
8.0–9.0	7.95–8.95

Select each step to learn more

U.S. Department of Transportation  
Federal Highway Administration

Step 5: Ensure that all classes are mutually exclusive. The class limits must be defined so that a value can be placed in only one class. Looking at the data in Step 1, we must assure that a value of 2 is placed in the correct class. There are two ways of doing this. One is to make an arbitrary decision to always place a value on a class limit in the lower class (or upper class), but keep in mind this arbitrary decision must apply to all values on a class limit.

In other words, you cannot place one value on a class limit in a lower class and another in an upper class. This is called a decision rule. Another way to ensure that the classes are mutually exclusive is to establish the class limits halfway between two possible observations. In the air content example, the data is reported to 0.1%. So if we add a significant figure and set the class limits at 0.05%, there can be no conflict. For example, we can use class limits of 0.95 to 1.95, 1.95 to 2.95, etc. as we see on the next slide.

Table description: Class limits tables.

## Exercise 1: Create a Frequency Table



Class Limits	Tally	Frequency	Relative Frequency
1.0–2.0			
2.0–3.0			
3.0–4.0			
4.0–5.0			
5.0–6.0			
6.0–7.0			
7.0–8.0			
8.0–9.0			

To open the exercises, select the paperclip icon.



We are now ready to develop the frequency table. From Step 4, we set the class limits at 1.0 to 2.0, 2.0 to 3.0, 3.0 to 4.0. etc. with the decision rule that any value on a class limit will be placed in the “lower” class.

Now take a moment to create a frequency table on your own. Use the exercises document that you were provided from the paperclip icon. There you will find the steps to complete this exercise. Recall that relative frequency is frequency divided by the total number of observations, or data points.

Please allow 5 minutes to conduct this exercise.

## Exercise 1: Recap



Class Limits	Tally	Frequency	Relative Frequency
1.0–2.0	/	1	0.033
2.0–3.0	//	2	0.067
3.0–4.0	////	4	0.133
4.0–5.0	/// /	6	0.200
5.0–6.0	/// ////	9	0.300
6.0–7.0	///	5	0.167
7.0–8.0	///	3	0.100
8.0–9.0		0	0.000



In the frequency table and the graphical approaches that are discussed next, certain information about the individual data points is lost. For example, from the frequency table it is no longer possible to know the values of the highest and lowest data points.

However, the advantages gained by grouping the data far outweigh this loss of individual detail. For example, the frequency table provides us with an appreciation for the center (the value around which the data are centered) and the spread of the data.

## Frequency Histogram

- An extension of the frequency table

If the class intervals are equal in width, the area of an individual rectangle represents the number of observations within the class, while the total area under the figure represents the total number of data values.

Class Limits	Tally	Frequency	Relative Frequency

Height = Frequency in Class

Equal in Width

Base = Class Interval

U.S. Department of Transportation  
Federal Highway Administration

MODULE A  
LESSON 5

ORGANIZING DATA

Resources  
Glossary  
Help

Now that we have covered the tabular form of presenting data, and because the mind grasps pictorial representations of data faster than other forms, let's consider the graphical forms of data presentation.

The frequency histogram is an extension of the frequency table. Individual rectangles whose heights are the frequencies in each class are erected on the horizontal axis whose base is set equal to the class intervals. If the class intervals are equal in width, the area of an individual rectangle represents the number of observations within the class, while the total area under the figure represents the total number of data values. This provides the best visual presentation of the data.

Image description: Histogram.

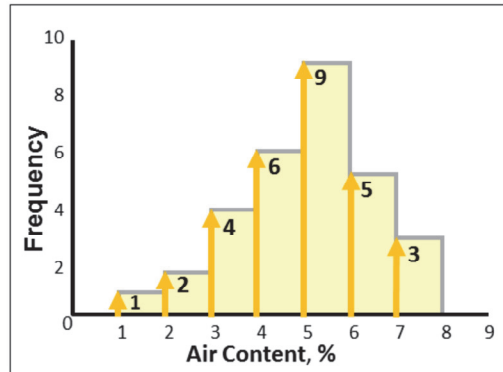
## Frequency Histogram



- Graphical representation of the "tally" and "frequency" columns in the frequency table

Class Limits	Tally	Frequency	Relative Frequency
1.0 – 2.0		1	0.033
2.0 – 3.0		2	0.067
3.0 – 4.0		4	0.133
4.0 – 5.0		6	0.200
5.0 – 6.0		9	0.300
6.0 – 7.0		5	0.167
7.0 – 8.0		3	0.100
8.0 – 9.0		0	0.000

This figure shows clearly that most of the air contents are between 5% and 6% and range from 1% to 8%



Remember that the individual low and high values are lost, but the overall picture more than compensates for that.



The histogram is a graphical representation of the "tally" and "frequency" columns in the frequency table with the vertical axis representing the frequency and the horizontal axis representing the variable, which is, in this case, concrete air content. A frequency table is usually developed as the first step in preparing a graphical presentation.

The frequency table in the slide above shows the results in the air content frequency histogram. In this frequency table, there is one observation between the class limits of 1.0 to 2.0, thus, the first rectangle on the left side is one unit high. This figure shows clearly that most of the air contents are between 5% and 6% and range from 1% to 8%. Remember from the frequency table and the frequency histogram, the individual low and high values are lost, but the overall picture more than compensates for that.

Image description: Histogram.



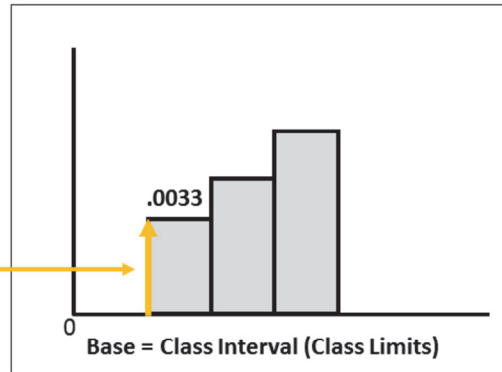
## Relative Frequency Histogram



- Height = Relative Frequency

Class Limits	Tally	Frequency	Relative Frequency
1.0 - 2.0	/	1	0.033
2.0 - 3.0	//	2	0.067
3.0 - 4.0	////	4	0.133
4.0 - 5.0	//// /	6	0.200
5.0 - 6.0	/// ////	9	0.300
6.0 - 7.0	///	5	0.167
7.0 - 8.0	///	3	0.100
8.0 - 9.0		0	0.000

Relative frequency in the first class (1.0 to 2.0) is  
 $1 \text{ (Frequency)} / 30 \text{ (Observations)} = 0.033$



Remember that relative frequency equals frequency divided by total number of observations (or data points).



The next step in the presentation (organization) of data is to divide the frequency in each class by the total number of observations in all classes, which produces the relative frequency for each class. These relative frequencies can then be plotted to form a relative frequency histogram. The relative frequency for any class is the frequency for that class divided by the total number of observations (data points).

For example, for the air content data, the relative frequency in the first class (1.0 to 2.0) is  $1/30$  or 0.033. And since all the data will be accounted for, the sum of the relative frequencies in all classes must add up to 1.0 (100%).

Image description: Histogram.

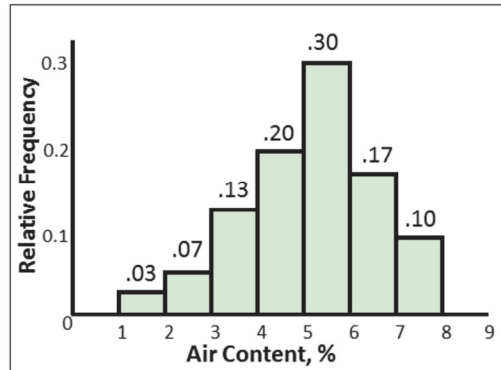
## Relative Frequency Histogram



- Provides us with a relatively simple definition of the term probability

Class Limits	Tally	Frequency	Relative Frequency
1.0 – 2.0	/	1	0.033
2.0 – 3.0	//	2	0.067
3.0 – 4.0	////	4	0.133
4.0 – 5.0	/// /	6	0.200
5.0 – 6.0	/// ////	9	0.300
6.0 – 7.0	///	5	0.167
7.0 – 8.0	///	3	0.100
8.0 – 9.0		0	0.000

In this case, we can say that the probability is 30% that a single test result will be between 5% and 6%.



The concept of relative frequency is extremely important because it provides a relatively simple definition of the term probability.



The concept of relative frequency is extremely important because it provides us with a relatively simple definition of the term probability (i.e., probability is the relative frequency of a particular occurrence). The relative frequency histogram for the concrete air content example shows that 9/30 (0.30%) of the test results were between 5% and 6%. In this case, we can say that the probability is 30% that a single test result will be between 5% and 6%. It is also common for specifications to require an estimate of the percentage of the material that is within certain limits. The relative frequency histogram provides this type of information for a data set.

Also from the relative frequency histogram, 0.50 (0.20 + 0.30) or 50% (20% + 30%) of the results are between 4% and 6% air.

It's important to know that the probability is 50% that a single test result will be between 4% and 6%. Suppose 4% and 6% were spec limits, and suppose the contractor was trying to reduce his/her variability, would he/she be satisfied?

Probability is the primary basis of the statistical analysis.

Image description: Histogram.

## Exercise 2: Using Density Data, pcf



1. Prepare a frequency table and relative frequency table
2. Construct a relative frequency histogram for the data

107.5	109.4	110.5	107.0	112.0
110.1	95.5	101.3	104.3	99.7
101.5	98.0	106.0	102.5	103.3
101.3	93.5	107.0	114.0	110.8
124.0	105.0	111.4	103.5	94.0
113.2	90.8	97.9	106.7	100.1

To open the exercises, select the paperclip icon.



Now take a moment to create a frequency table and relative frequency table. Use the exercises document that you were provided from the paperclip icon. Please allow 10 minutes to conduct this exercise. Remember to establish a decision rule.

Using the 30 density data values in the slide, you will organize the data to create a frequency table and histogram. The following results will complete this exercise:

1. Prepare a frequency and relative frequency table (using seven classes) for the data in the table; and
2. Construct a relative frequency histogram for the data using the frequency table found in the exercise.

Image description: Table with a set of 30 air content test results.



## Exercise 2: Recap

1. Prepare a frequency table and relative frequency table
2. Construct a relative frequency histogram for the data

Class Limits	Tally	Frequency	Relative Frequency
90.0–95.0	///	3	.10
95.0–100.0	////	4	.13
100.0–105.0		9	.30
105.0–110.0	/	6	.20
110.0–115.0	//	7	.24
115.0–120.0		0	0
120.0–125.0	/	1	.03

ANSWER



If the specifications require that 90% of the material exceed a density of 110 pcf, do you believe that the material sampled meets the specifications? Why or why not?



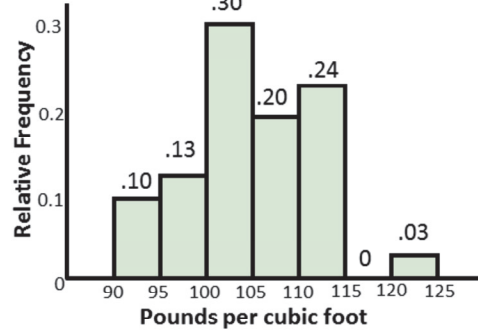
If the specifications require that 90% of the material exceed a density of 110 pounds per cubic foot (pcf) do you believe that the material sampled meets the specifications? Why or why not?

Select the answer button the learn more.

## Exercise 2: Recap



- 1.
- 2.



X CLOSE



If the specifications require that 90% of the material exceed a density of 110 pcf, do you believe that the material sampled meets the specifications? Why or why not?

ANSWER



The answer is no. Your histogram should show that much less than 90% of the data is above 110 pcf.

- Range =  $124.0 - 90.8 = 33.2$ ; and
- Class Interval =  $33.2/7 = 4.7$ , use 5.


Class Limits:

- 90.0–95.0
- 95.0–100.0
- 100.0–105.0
- 105.0–110.0


Decision rule: Place any value on a class limit in the lower class.




Approximately 27% of results exceeded 110 pcf, so the spec was not met.

Image description: Histogram.

Which of the following formats allows the mind to grasp the data fastest? 


- a) Raw data
- b) Sorted data
- c) Frequency table
- d) Frequency histogram

 Knowledge Check Try again Submit Clear



 **MODULE A**  
**LESSON 5** **ORGANIZING DATA** Resources  
Glossary  
Help  




Which of the following formats allows the mind to grasp the data fastest?

- a) Raw data;
- b) Sorted data;
- c) Frequency table; or
- d) Frequency histogram.

Which of the following formats allows the mind to grasp the data fastest? 

- a) Raw data
- b) Sorted data
- c) Frequency table
- d) Frequency histogram

  Knowledge Check Debrief

 **MODULE A**  
**LESSON 5** **ORGANIZING DATA** Resources  
Glossary  
Help  

The correct answer is d) Frequency histogram. The frequency histogram presents the data in a format that is easiest to get a visual idea of the data, that is, the lowest to highest, the center, and a general idea of the shape.

## Learning Outcomes Review



You are now able to:

- Create a frequency table using the guidelines provided
- Plot a frequency histogram using a given data set

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.



You have completed Module A, Lesson 5: Organizing Data. You are now able to:

- Create a frequency table using the guidelines provided; and
- Plot a frequency histogram using a given data set.

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.