Describing Data

Department of Mathematics

Salt Lake Community College

(Slides by Adam Wilson)

Definition

A **frequency table** is a table with two columns. One column lists the categories, and the other the frequencies with which the items in the categories occur. (i.e. how many items fit into each category.)

Definition

A **frequency table** is a table with two columns. One column lists the categories, and the other the frequencies with which the items in the categories occur. (i.e. how many items fit into each category.)

Example 1

An insurance company determines vehicle insurance premiums based on known risk factors. If a person is considered a higher risk, their premiums will be higher. One potential factor is the color of your car. The insurance company believes that people with some color cars are more likely to get in accidents.

To research this, they examined police reports for recent total-loss collisions. The data is summarized in the frequency table.

Color	**Frequency**
Blue	25
Green	52
Red	41
White	36
Black	39
Grey	23

Definition

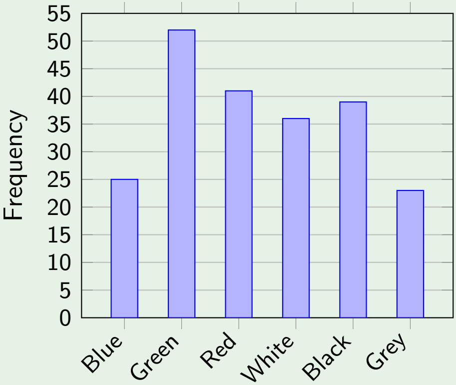
A **bar graph** is a graph that displays a bar for each category with the length of each bar indicating the frequency of that category.

Definition

A **bar graph** is a graph that displays a bar for each category with the length of each bar indicating the frequency of that category.

Example 2

Using the car data from Example 1, we have the following bar graph.



Definition

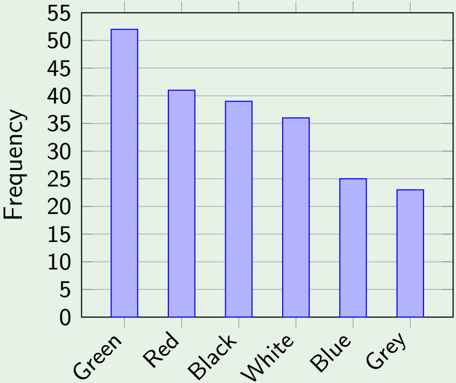
A **Pareto chart** is a bar graph ordered from highest to lowest frequency.

Definition

A **Pareto chart** is a bar graph ordered from highest to lowest frequency.

Example 3

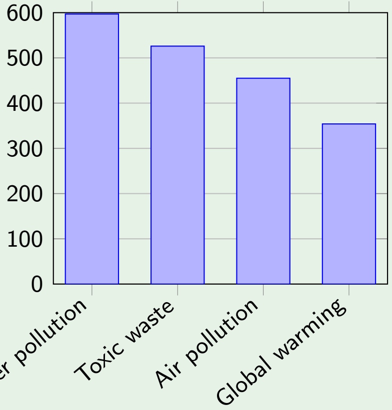
Using the car data from Example 1, we have the following Pareto chart.



Example 4

In a survey[1], adults were asked whether they personally worried about a variety of environmental concerns. The numbers (out of 1012 surveyed) who indicated that they worried "a great deal" about some selected concerns are summarized below.

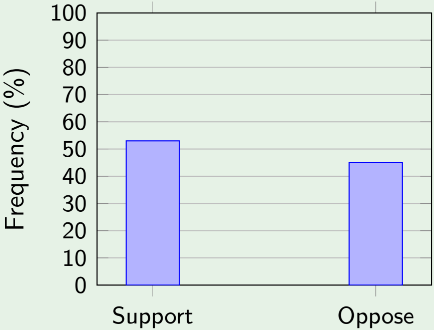
Environmental Issue	Frequency
Water pollution	597
Toxic waste	526
Air pollution	455
Global warming	354

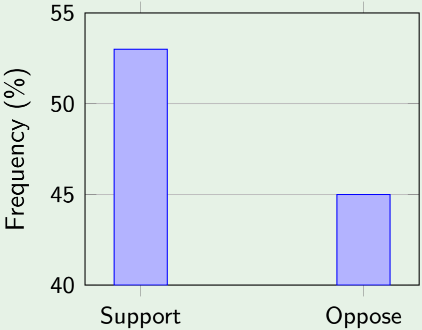


[1] Gallup Poll. March 5-8, 2009. http://www.pollingreport.com/enviro.htm

Example 5

Compare the two graphs below showing support for same-sex marriage rights from a poll taken in May 2013[2].





The difference in the vertical scale on the first graph suggests a different story than the true differences in percentages; the second graph makes it look like more than twice as many people support marriage rights as oppose it.

[2] Gallup Poll. May 2-7, 2013, from http://www.pollingreport.com/civil.htm

Definition

A **histogram** a bar graph, where the horizontal axis is a number line.

Definition

A **histogram** a bar graph, where the horizontal axis is a number line.

Definition

Class intervals are groupings of the data. In general, we define class intervals so that:

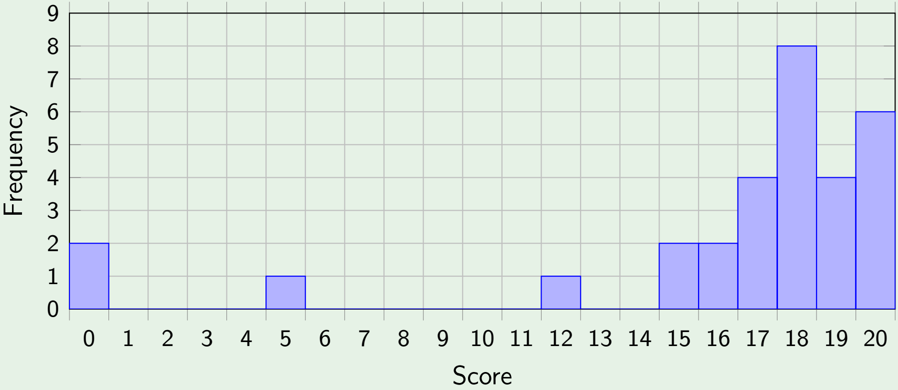
- Each interval is equal size. For example, if the first class contains values from 120-129, then the second class should include 130-139.
- We have somewhere between 5 and 20 classes, typically, depending upon the number of data we are working with.

Example 6

A teacher records scores on a 20-point quiz for the 30 students in his class.

19	20	18	18	17	18	19	17	20	18	20	16	20	15	17
12	18	19	18	19	17	20	18	16	15	18	20	5	0	0

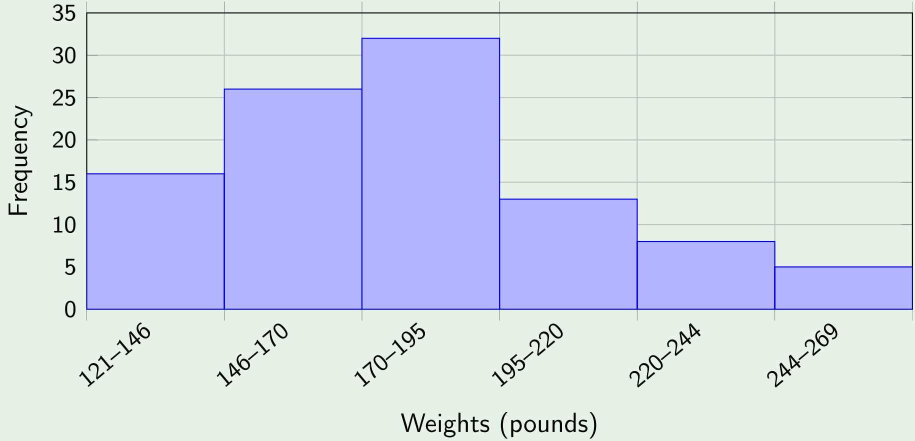
These scores could be summarized in the following histogram.



Example 7

Suppose that we have collected weights from 100 male subjects as part of a nutrition study. For our weight data, we have values ranging from a low of 121 pounds to a high of 269 pounds, giving a total span of 263-121 = 142.

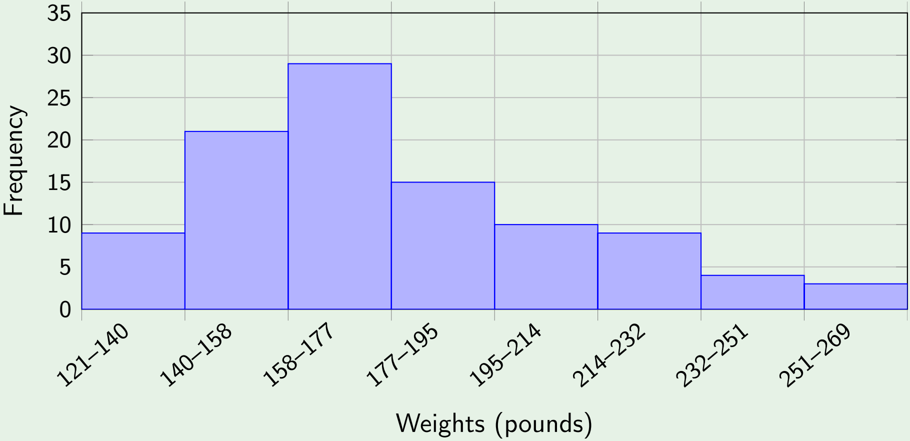
We could create 6 intervals with a width of around 20.



Example 7

Suppose that we have collected weights from 100 male subjects as part of a nutrition study. For our weight data, we have values ranging from a low of 121 pounds to a high of 269 pounds, giving a total span of 263-121 = 142.

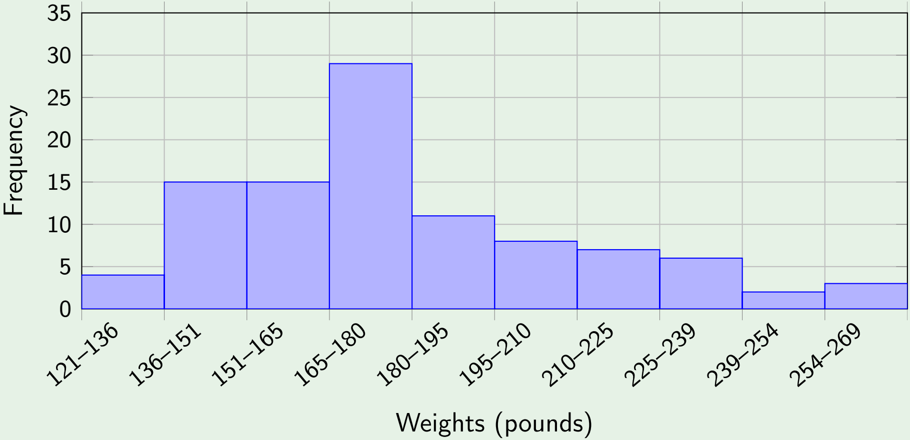
We could create 8 intervals with a width of around 18.



Example 7

Suppose that we have collected weights from 100 male subjects as part of a nutrition study. For our weight data, we have values ranging from a low of 121 pounds to a high of 269 pounds, giving a total span of 263-121 = 142.

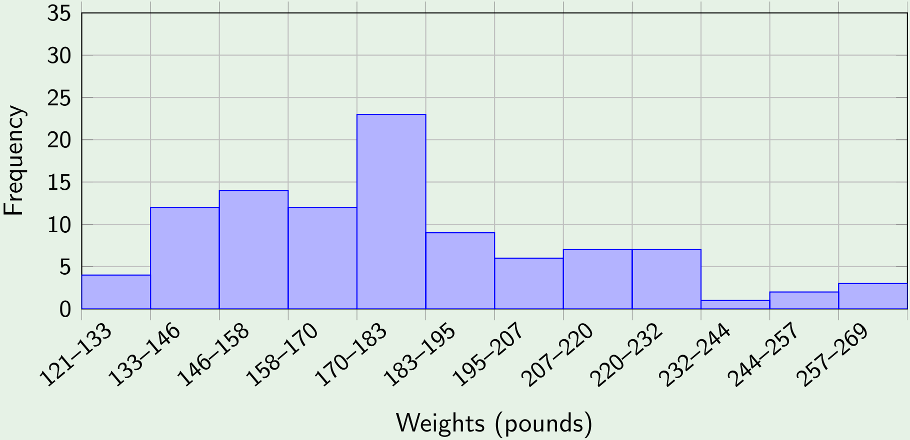
We could create 10 intervals with a width of around 14.



Example 7

Suppose that we have collected weights from 100 male subjects as part of a nutrition study. For our weight data, we have values ranging from a low of 121 pounds to a high of 269 pounds, giving a total span of 263-121 = 142.

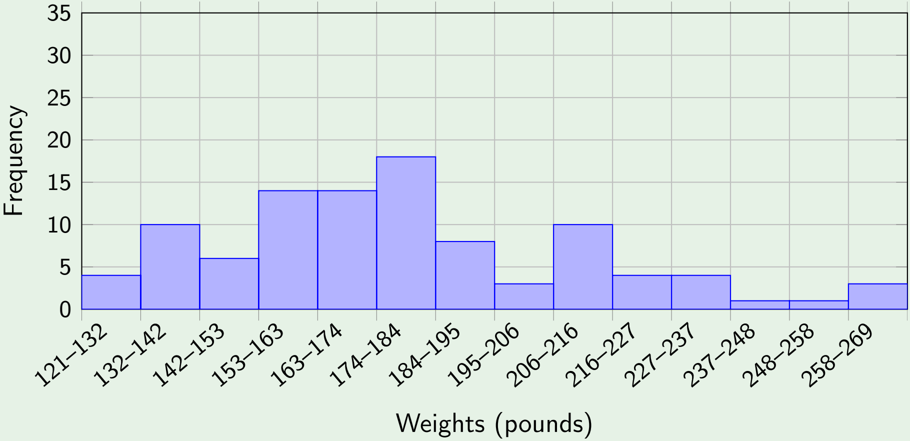
We could create 10 intervals with a width of around 12.



Example 7

Suppose that we have collected weights from 100 male subjects as part of a nutrition study. For our weight data, we have values ranging from a low of 121 pounds to a high of 269 pounds, giving a total span of 263-121 = 142.

We could create 14 intervals with a width of around 10.



Definition

The **mean** of a set of data is the sum of the data values divided by the number of values.

Definition

The **mean** of a set of data is the sum of the data values divided by the number of values.

Note

It is not uncommon to see the word "average" used instead of "mean".

Definition

The **mean** of a set of data is the sum of the data values divided by the number of values.

Note

It is not uncommon to see the word "average" used instead of "mean".

Example 8

Marci's exam scores for her last math class were: 79, 86, 82, 94.
What is the mean of these values?

Definition

The **mean** of a set of data is the sum of the data values divided by the number of values.

Note

It is not uncommon to see the word "average" used instead of "mean".

Example 8

Marci's exam scores for her last math class were: 79, 86, 82, 94.
What is the mean of these values?

$$\frac{79 + 86 + 82 + 94}{4} = \frac{341}{4} = 82.25 \approx 82.3$$

Typically we round means to one more decimal than the original data.

Example 9

The one hundred families in a particular neighborhood are asked their annual household income, to the nearest \$5 thousand dollars.

Income	Frequency	Income	Frequency
15	6	35	19
20	8	40	20
25	11	45	12
30	17	50	7

What is the mean of this data?

Example 9

The one hundred families in a particular neighborhood are asked their annual household income, to the nearest \$5 thousand dollars.

Income	Frequency	Income	Frequency
15	6	35	19
20	8	40	20
25	11	45	12
30	17	50	7

What is the mean of this data?

$$\frac{\overbrace{15+\cdots+15}^{6 \text{ terms}}+\overbrace{20+\cdots+20}^{8 \text{ terms}}+\overbrace{25+\cdots+25}^{11 \text{ terms}}+\cdots+\overbrace{50+\cdots+50}^{7 \text{ terms}}}{100}$$

$$=\frac{15\cdot 6+20\cdot 8+25\cdot 11+30\cdot 17+35\cdot 19+40\cdot 20+45\cdot 12+50\cdot 7}{100}$$

$$=\frac{3390}{100}=33.9$$

Definition

The **median** of a set of data is the value in the middle when the data is in numerical order.

Definition

The **median** of a set of data is the value in the middle when the data is in numerical order.

Note

To find the median, begin by listing the data in order from smallest to largest, or largest to smallest.

If the number of data values, N, is odd, then the median is the middle data value. This value can be found by rounding $\frac{N}{2}$ up to the next whole number.

If the number of data values is even, there is no one middle value, so we find the mean of the two middle values (values $\frac{N}{2}$ and $\frac{N}{2}+1$)

Example 10

Find the media of these quiz scores: 8 6 4 1 7 1 1

Example 10

Find the media of these quiz scores: 8 6 4 1 7 1 1

List the data in order: 1 1 1 $\underset{\Uparrow}{4}$ 6 7 8

Since we have an odd number of data, we see the median is 4.

Example 10

Find the media of these quiz scores: 8 6 4 1 7 1 1

List the data in order: 1 1 1 $\underset{\Uparrow}{4}$ 6 7 8

Since we have an odd number of data, we see the median is 4.

Example 11

Find the median of these quiz scores: 5 9 8 6 4 8 2 5 7 7

Example 10

Find the media of these quiz scores: 8 6 4 1 7 1 1

List the data in order: 1 1 1 $\underset{\Uparrow}{4}$ 6 7 8

Since we have an odd number of data, we see the median is 4.

Example 11

Find the median of these quiz scores: 5 9 8 6 4 8 2 5 7 7

List the data in order: 2 4 5 5 6 $\underset{\Uparrow}{\ }$ 7 7 8 8 9

Since we have an even number of data, there is no middle item. In this case, we say the median is the mean of the two middle numbers, 6 and 7, which is 6.5.

Example 12

The one hundred families in a particular neighborhood are asked their annual household income, to the nearest $5 thousand dollars.

Income	Frequency	Income	Frequency
15	6	35	19
20	8	40	20
25	11	45	12
30	17	50	7

What is the median of this data?

Example 12

The one hundred families in a particular neighborhood are asked their annual household income, to the nearest \$5 thousand dollars.

Income	Frequency	Income	Frequency
15	6	35	19
20	8	40	20
25	11	45	12
30	17	50	7

What is the median of this data?

We have 100 items, this means the median will be the mean of the 50^{th} and 51^{st} data values. So, we need to start counting up from the bottom:

Example 12

The one hundred families in a particular neighborhood are asked their annual household income, to the nearest $5 thousand dollars.

Income	Frequency	Income	Frequency
15	6	35	19
20	8	40	20
25	11	45	12
30	17	50	7

What is the median of this data?

We have 100 items, this means the median will be the mean of the 50^{th} and 51^{st} data values. So, we need to start counting up from the bottom:

There are 6 data values of 15. $\Rightarrow$ Values 1 to 6 are 15.

Example 12

The one hundred families in a particular neighborhood are asked their annual household income, to the nearest \$5 thousand dollars.

Income	Frequency	Income	Frequency
15	6	35	19
20	8	40	20
25	11	45	12
30	17	50	7

What is the median of this data?

We have 100 items, this means the median will be the mean of the 50^{th} and 51^{st} data values. So, we need to start counting up from the bottom:

There are 6 data values of 15. $\Rightarrow$ Values 1 to 6 are 15.
The next 8 data values are 20. $\Rightarrow$ Values 7 to (6+8)=14 are 20.

Example 12

The one hundred families in a particular neighborhood are asked their annual household income, to the nearest \$5 thousand dollars.

Income	Frequency	Income	Frequency
15	6	35	19
20	8	40	20
25	11	45	12
30	17	50	7

What is the median of this data?

We have 100 items, this means the median will be the mean of the 50^{th} and 51^{st} data values. So, we need to start counting up from the bottom:

There are 6 data values of 15. $\Rightarrow$ Values 1 to 6 are 15.
The next 8 data values are 20. $\Rightarrow$ Values 7 to (6+8)=14 are 20.
The next 11 data values are 25. $\Rightarrow$ Values 15 to (14+11)=25 are 25.

Example 12

The one hundred families in a particular neighborhood are asked their annual household income, to the nearest \$5 thousand dollars.

Income	Frequency	Income	Frequency
15	6	35	19
20	8	40	20
25	11	45	12
30	17	50	7

What is the median of this data?

We have 100 items, this means the median will be the mean of the 50^{th} and 51^{st} data values. So, we need to start counting up from the bottom:

There are 6 data values of 15.	$\Rightarrow$	Values 1 to 6 are 15.
The next 8 data values are 20.	$\Rightarrow$	Values 7 to (6+8)=14 are 20.
The next 11 data values are 25.	$\Rightarrow$	Values 15 to (14+11)=25 are 25.
The next 17 data values are 30.	$\Rightarrow$	Values 26 to (25+17)=42 are 30.

Example 12

The one hundred families in a particular neighborhood are asked their annual household income, to the nearest \$5 thousand dollars.

Income	Frequency	Income	Frequency
15	6	35	19
20	8	40	20
25	11	45	12
30	17	50	7

What is the median of this data?

We have 100 items, this means the median will be the mean of the 50^{th} and 51^{st} data values. So, we need to start counting up from the bottom:

There are 6 data values of 15.	$\Rightarrow$	Values 1 to 6 are 15.
The next 8 data values are 20.	$\Rightarrow$	Values 7 to (6+8)=14 are 20.
The next 11 data values are 25.	$\Rightarrow$	Values 15 to (14+11)=25 are 25.
The next 17 data values are 30.	$\Rightarrow$	Values 26 to (25+17)=42 are 30.
The next 19 data values are 35.	$\Rightarrow$	Values 43 to (42+19)=61 are 35.

Example 12

The one hundred families in a particular neighborhood are asked their annual household income, to the nearest \$5 thousand dollars.

Income	Frequency	Income	Frequency
15	6	35	19
20	8	40	20
25	11	45	12
30	17	50	7

What is the median of this data?

We have 100 items, this means the median will be the mean of the 50^{th} and 51^{st} data values. So, we need to start counting up from the bottom:

There are 6 data values of 15.	$\Rightarrow$	Values 1 to 6 are 15.
The next 8 data values are 20.	$\Rightarrow$	Values 7 to (6+8)=14 are 20.
The next 11 data values are 25.	$\Rightarrow$	Values 15 to (14+11)=25 are 25.
The next 17 data values are 30.	$\Rightarrow$	Values 26 to (25+17)=42 are 30.
The next 19 data values are 35.	$\Rightarrow$	Values 43 to (42+19)=61 are 35.

So, the median is $\frac{35+35}{2} = 35$, \$35 thousand dollars.

Definition

The **mode** is the element of the data set that occurs most frequently.

Definition

The **mode** is the element of the data set that occurs most frequently.

Note

The mode is useless with quantitative data. It is most commonly used for categorical data, for which median and mean cannot be computed.

Definition

The **mode** is the element of the data set that occurs most frequently.

Note

The mode is useless with quantitative data. It is most commonly used for categorical data, for which median and mean cannot be computed.

Example 13

In Example 1 we collected the data:

Color	Frequency	Color	Frequency
Blue	3	White	3
Green	5	Black	2
Red	4	Grey	3

For this data, Green is the mode.

Definition

The **mode** is the element of the data set that occurs most frequently.

Note

The mode is useless with quantitative data. It is most commonly used for categorical data, for which median and mean cannot be computed.

Example 13

In Example 1 we collected the data:

Color	**Frequency**	**Color**	**Frequency**
Blue	3	White	3
Green	5	Black	2
Red	4	Grey	3

For this data, Green is the mode.

Note

It is possible for a data set to have more than one mode.

Example 14

Consider these three sets of quiz scores:

Section A	5	5	5	5	5	5	5	5	5	5
Section B	0	0	0	0	0	10	10	10	10	10
Section C	4	4	4	5	5	5	5	6	6	6

All three sets of data have a mean of 5 and a median of 5, yet each section has quite different scores.

Example 14

Consider these three sets of quiz scores:

Section A	5	5	5	5	5	5	5	5	5	5
Section B	0	0	0	0	0	10	10	10	10	10
Section C	4	4	4	5	5	5	5	6	6	6

All three sets of data have a mean of 5 and a median of 5, yet each section has quite different scores.

Definition

The **range** is the difference between the maximum and minimum values.

Example 14

Consider these three sets of quiz scores:

Section A	5	5	5	5	5	5	5	5	5	5
Section B	0	0	0	0	0	10	10	10	10	10
Section C	4	4	4	5	5	5	5	6	6	6

All three sets of data have a mean of 5 and a median of 5, yet each section has quite different scores.

Definition

The **range** is the difference between the maximum and minimum values.

Example 15

Using the quiz scores from Example 14.
For Section A, the range is $5 - 5 = 0$.

Example 14

Consider these three sets of quiz scores:

Section A	5	5	5	5	5	5	5	5	5	5
Section B	0	0	0	0	0	10	10	10	10	10
Section C	4	4	4	5	5	5	5	6	6	6

All three sets of data have a mean of 5 and a median of 5, yet each section has quite different scores.

Definition

The **range** is the difference between the maximum and minimum values.

Example 15

Using the quiz scores from Example 14.
For Section A, the range is $5 - 5 = 0$.
For Section B, the range is $10 - 0 = 10$.

Example 14

Consider these three sets of quiz scores:

Section A	5	5	5	5	5	5	5	5	5	5
Section B	0	0	0	0	0	10	10	10	10	10
Section C	4	4	4	5	5	5	5	6	6	6

All three sets of data have a mean of 5 and a median of 5, yet each section has quite different scores.

Definition

The **range** is the difference between the maximum and minimum values.

Example 15

Using the quiz scores from Example 14.
For Section A, the range is $5 - 5 = 0$.
For Section B, the range is $10 - 0 = 10$.
For Section C, the range is $6 - 4 = 2$.

Definition

The **standard deviation** is a measure of variation based on measuring how far each data value deviates, or is different, from the mean.

The standard deviation

- is always positive.

Definition

The **standard deviation** is a measure of variation based on measuring how far each data value deviates, or is different, from the mean.

The standard deviation

- is always positive.
- will be zero if all the data values are equal.

Definition

The **standard deviation** is a measure of variation based on measuring how far each data value deviates, or is different, from the mean.

The standard deviation

- is always positive.
- will be zero if all the data values are equal.
- will get larger the more the data spreads out.

Definition

The **standard deviation** is a measure of variation based on measuring how far each data value deviates, or is different, from the mean.

The standard deviation

- is always positive.
- will be zero if all the data values are equal.
- will get larger the more the data spreads out.
- has the same units as the original data.

Definition

The **standard deviation** is a measure of variation based on measuring how far each data value deviates, or is different, from the mean.

The standard deviation

- is always positive.
- will be zero if all the data values are equal.
- will get larger the more the data spreads out.
- has the same units as the original data.
- can be highly influenced by outliers.

Computing Standard Deviations

To compute a standard deviation of a data set with size n.

1. Find the deviation of each data from the mean.

Computing Standard Deviations

To compute a standard deviation of a data set with size n.

1. Find the deviation of each data from the mean.
2. Square each deviation.

Computing Standard Deviations

To compute a standard deviation of a data set with size n.

1. Find the deviation of each data from the mean.
2. Square each deviation.
3. Add the squared deviations.

Computing Standard Deviations

To compute a standard deviation of a data set with size n.

1. Find the deviation of each data from the mean.
2. Square each deviation.
3. Add the squared deviations.
4. - If the data represents the whole population, divide the sum by n.

Computing Standard Deviations

To compute a standard deviation of a data set with size n.

1. Find the deviation of each data from the mean.
2. Square each deviation.
3. Add the squared deviations.
4. - If the data represents the whole population, divide the sum by n.
   - If the data is from a sample, divide the sum by $(n-1)$.

Computing Standard Deviations

To compute a standard deviation of a data set with size n.

1. Find the deviation of each data from the mean.
2. Square each deviation.
3. Add the squared deviations.
4. - If the data represents the whole population, divide the sum by n.
   - If the data is from a sample, divide the sum by $(n-1)$.
5. Compute the square root of the result.

Computing Standard Deviations

To compute a standard deviation of a data set with size n.

1. Find the deviation of each data from the mean.
2. Square each deviation.
3. Add the squared deviations.
4. - If the data represents the whole population, divide the sum by n.
   - If the data is from a sample, divide the sum by $(n-1)$.
5. Compute the square root of the result.

Note

Standard deviations are only computed by hand for very small data sets.

Example 16

Let us find the standard deviation for the quiz scores:

$$0 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 10$$

Example 16

Let us find the standard deviation for the quiz scores:

0 5 5 5 5 5 5 5 5 10

The mean is 5 and the deviations are:

Data Value	Deviation	Deviation squared
0		
5		
5		
5		
5		
5		
5		
5		
5		
10		

Example 16

Let us find the standard deviation for the quiz scores:

$$0 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 10$$

The mean is 5 and the deviations are:

Data Value	Deviation	Deviation squared
0	$0-5 = -5$	
5	$5-5 = 0$	
5	$5-5 = 0$	
5	$5-5 = 0$	
5	$5-5 = 0$	
5	$5-5 = 0$	
5	$5-5 = 0$	
5	$5-5 = 0$	
5	$5-5 = 0$	
10	$10-5 = 5$	

Example 16

Let us find the standard deviation for the quiz scores:

$$0 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 10$$

The mean is 5 and the deviations are:

Data Value	Deviation	Deviation squared
0	$0-5 = -5$	$(-5)^2 = 25$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
10	$10-5 = 5$	$(5)^2 = 25$

Example 16

Let us find the standard deviation for the quiz scores:

$$0 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 10$$

The mean is 5 and the deviations are:

Data Value	Deviation	Deviation squared
0	$0-5 = -5$	$(-5)^2 = 25$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
5	$5-5 = 0$	$(0)^2 = 0$
10	$10-5 = 5$	$(5)^2 = 25$

The standard deviation is $\sqrt{\dfrac{50}{10}} = \sqrt{5} \approx 2.2$.

Example 17

The standard deviations for the quiz scores from Example 14 and Example 16 are:

Data Set											Std. Dev.
Section A:	5	5	5	5	5	5	5	5	5	5	0
Section B:	0	0	0	0	0	10	10	10	10	10	5
Section C:	4	4	4	5	5	5	5	6	6	6	0.8
Section D:	0	5	5	5	5	5	5	5	5	10	2.2

Example 18

The price of a jar of peanut butter at 5 stores were: \$3.29, \$3.59, \$3.75, \$3.79, and \$3.99. Let us find the standard deviation of the prices.

Example 18

The price of a jar of peanut butter at 5 stores were: \$3.29, \$3.59, \$3.75, \$3.79, and \$3.99. Let us find the standard deviation of the prices.
The mean is 3.682 and the deviations are:

Data Value	Deviation	Deviation squared
3.29		
3.59		
3.75		
3.79		
3.99		

Example 18

The price of a jar of peanut butter at 5 stores were: \$3.29, \$3.59, \$3.75, \$3.79, and \$3.99. Let us find the standard deviation of the prices.
The mean is 3.682 and the deviations are:

Data Value	Deviation	Deviation squared
3.29	$3.29 - 3.682 = -0.391$	
3.59	$3.59 - 3.682 = -0.092$	
3.75	$3.75 - 3.682 = 0.068$	
3.79	$3.79 - 3.682 = 0.108$	
3.99	$3.99 - 3.682 = 0.308$	

Example 18

The price of a jar of peanut butter at 5 stores were: \$3.29, \$3.59, \$3.75, \$3.79, and \$3.99. Let us find the standard deviation of the prices.
The mean is 3.682 and the deviations are:

Data Value	Deviation	Deviation squared
3.29	$3.29 - 3.682 = -0.391$	$(-0.391)^2 = 0.153664$
3.59	$3.59 - 3.682 = -0.092$	$(-0.092)^2 = 0.008464$
3.75	$3.75 - 3.682 = 0.068$	$(0.068)^2 = 0.011664$
3.79	$3.79 - 3.682 = 0.108$	$(0.108)^2 = 0.004624$
3.99	$3.99 - 3.682 = 0.308$	$(0.308)^2 = 0.094864$

Example 18

The price of a jar of peanut butter at 5 stores were: \$3.29, \$3.59, \$3.75, \$3.79, and \$3.99. Let us find the standard deviation of the prices.
The mean is 3.682 and the deviations are:

Data Value	Deviation	Deviation squared
3.29	$3.29 - 3.682 = -0.391$	$(-0.391)^2 = 0.153664$
3.59	$3.59 - 3.682 = -0.092$	$(-0.092)^2 = 0.008464$
3.75	$3.75 - 3.682 = 0.068$	$(0.068)^2 = 0.011664$
3.79	$3.79 - 3.682 = 0.108$	$(0.108)^2 = 0.004624$
3.99	$3.99 - 3.682 = 0.308$	$(0.308)^2 = 0.094864$

Note that we are dealing with a sample. The standard deviation is:

$$\sqrt{\frac{0.153664 + 0.008464 + 0.011664 + 0.004624 + 0.094864}{4}}$$

$$= \sqrt{\frac{0.27328}{4}} = \sqrt{0.06832} \approx \$0.261$$

Definition

Quartiles are values that divide the data into quarters.

Definition

Quartiles are values that divide the data into quarters.

The first quartile, Q_1, is the data value so that 25% of the data fall below.

The third quartile, Q_3, is the data value so that 75% of the data fall below.

Definition

Quartiles are values that divide the data into quarters.

The first quartile, Q_1, is the data value so that 25% of the data fall below.

The third quartile, Q_3, is the data value so that 75% of the data fall below.

This divides the data into quarters:

- 25% of the data is between the minimum and Q_1.
- 25% of the data is between Q_1 and the median.
- 25% of the data is between the media and Q_3.
- 25% of the data is between Q_3 and the maximum.

Definition

Quartiles are values that divide the data into quarters.

The first quartile, Q_1, is the data value so that 25% of the data fall below.

The third quartile, Q_3, is the data value so that 75% of the data fall below.

This divides the data into quarters:

- 25% of the data is between the minimum and Q_1.
- 25% of the data is between Q_1 and the median.
- 25% of the data is between the media and Q_3.
- 25% of the data is between Q_3 and the maximum.

Five-Number Summary

A common summary of a data set takes the form:
Minimum, Q_1, Median, Q_3, Maximum

How to Calculate Q_1

1. Order the data from smallest to largest.

How to Calculate Q_1

1. Order the data from smallest to largest.
2. Compute the locator: $L = 0.25n$

How to Calculate Q_1

1. Order the data from smallest to largest.
2. Compute the locator: $L = 0.25n$
3.
   - If L is a decimal value:
     - Round up to the next whole number, $L+$.
     - Q_1 is the data value at position $L+$.

How to Calculate Q_1

1. Order the data from smallest to largest.
2. Compute the locator: $L = 0.25n$
3.
    - If L is a decimal value:
        - Round up to the next whole number, $L+$.
        - Q_1 is the data value at position $L+$.
    - If L is a whole number:
        - Q_1 is the mean of the data values in position L and $L + 1$.

How to Calculate Q_1

1. Order the data from smallest to largest.
2. Compute the locator: $L = 0.25n$
3.
   - If L is a decimal value:
     - Round up to the next whole number, $L+$.
     - Q_1 is the data value at position $L+$.
   - If L is a whole number:
     - Q_1 is the mean of the data values in position L and $L + 1$.

How to Calculate Q_3

Use the same procedure as for Q_1, but with locator: $L = 0.75n$.

Example 19

Suppose we have measured the height (in inches) of 9 women.
The data, sorted from smallest to largest, is:

$$59 \quad 60 \quad 62 \quad 64 \quad 66 \quad 67 \quad 69 \quad 70 \quad 72$$

Example 19

Suppose we have measured the height (in inches) of 9 women.
The data, sorted from smallest to largest, is:

59 60 62 64 66 67 69 70 72

Finding Q_1: The locator is $L = 0.25 \cdot 9 = 2.25$, so we need to round up, which gives $L+ = 3$. Thus, $Q_1 = 62$.

Example 19

Suppose we have measured the height (in inches) of 9 women.
The data, sorted from smallest to largest, is:

$$59 \quad 60 \quad 62 \quad 64 \quad 66 \quad 67 \quad 69 \quad 70 \quad 72$$

Finding Q_1: The locator is $L = 0.25 \cdot 9 = 2.25$, so we need to round up, which gives $L+ = 3$. Thus, $Q_1 = 62$.

Finding Q_2: This is just the median, which is 66.

Example 19

Suppose we have measured the height (in inches) of 9 women.
The data, sorted from smallest to largest, is:

$$59 \quad 60 \quad 62 \quad 64 \quad 66 \quad 67 \quad 69 \quad 70 \quad 72$$

Finding Q_1: The locator is $L = 0.25 \cdot 9 = 2.25$, so we need to round up, which gives $L+ = 3$. Thus, $Q_1 = 62$.

Finding Q_2: This is just the median, which is 66.

Finding Q_3: The locator is $L = 0.75 \cdot 9 = 6.75$, so we need to round up, which gives $L+ = 7$. Thus, $Q_3 = 69$.

Example 19

Suppose we have measured the height (in inches) of 9 women.
The data, sorted from smallest to largest, is:

$$59 \quad 60 \quad 62 \quad 64 \quad 66 \quad 67 \quad 69 \quad 70 \quad 72$$

Finding Q_1: The locator is $L = 0.25 \cdot 9 = 2.25$, so we need to round up, which gives $L+ = 3$. Thus, $Q_1 = 62$.

Finding Q_2: This is just the median, which is 66.

Finding Q_3: The locator is $L = 0.75 \cdot 9 = 6.75$, so we need to round up, which gives $L+ = 7$. Thus, $Q_3 = 69$.

The five-number summary for this data set is: 59, 62, 66, 69, 72

Example 20

Suppose, as in Example 19, we have measured the height (in inches) of 8 women. The data, sorted from smallest to largest, is:

$$59 \quad 60 \quad 62 \quad 64 \quad 66 \quad 67 \quad 69 \quad 70$$

Example 20

Suppose, as in Example 19, we have measured the height (in inches) of 8 women. The data, sorted from smallest to largest, is:

$$59 \quad 60 \quad 62 \quad 64 \quad 66 \quad 67 \quad 69 \quad 70$$

Finding Q_1: The locator is $L = 0.25 \cdot 8 = 2$, so Q_1 will be the median of the second and third data values. Thus, $Q_1 = \dfrac{60 + 62}{2} = 61$.

Example 20

Suppose, as in Example 19, we have measured the height (in inches) of 8 women. The data, sorted from smallest to largest, is:

$$59 \quad 60 \quad 62 \quad 64 \quad 66 \quad 67 \quad 69 \quad 70$$

Finding Q_1: The locator is $L = 0.25 \cdot 8 = 2$, so Q_1 will be the median of the second and third data values. Thus, $Q_1 = \dfrac{60 + 62}{2} = 61$.

Finding Q_2: This is just the median, which is 65.

Example 20

Suppose, as in Example 19, we have measured the height (in inches) of 8 women. The data, sorted from smallest to largest, is:

$$59 \quad 60 \quad 62 \quad 64 \quad 66 \quad 67 \quad 69 \quad 70$$

Finding Q_1: The locator is $L = 0.25 \cdot 8 = 2$, so Q_1 will be the median of the second and third data values. Thus, $Q_1 = \dfrac{60+62}{2} = 61$.

Finding Q_2: This is just the median, which is 65.

Finding Q_3: The locator is $L = 0.75 \cdot 8 = 6$, so Q_3 will be the median of the sixth and seventh data values. Thus, $Q_3 = \dfrac{67+69}{2} = 68$.

Example 20

Suppose, as in Example 19, we have measured the height (in inches) of 8 women. The data, sorted from smallest to largest, is:

$$59 \quad 60 \quad 62 \quad 64 \quad 66 \quad 67 \quad 69 \quad 70$$

Finding Q_1: The locator is $L = 0.25 \cdot 8 = 2$, so Q_1 will be the median of the second and third data values. Thus, $Q_1 = \dfrac{60+62}{2} = 61$.

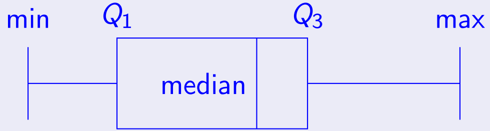
Finding Q_2: This is just the median, which is 65.

Finding Q_3: The locator is $L = 0.75 \cdot 8 = 6$, so Q_3 will be the median of the sixth and seventh data values. Thus, $Q_3 = \dfrac{67+69}{2} = 68$.

The five-number summary for this data set is: 59, 61, 65, 68, 70

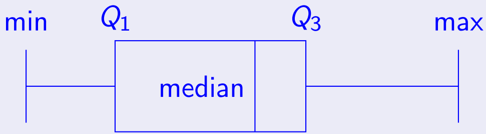
Definition

A **box plot** is a graphical representation of a five-number summary.



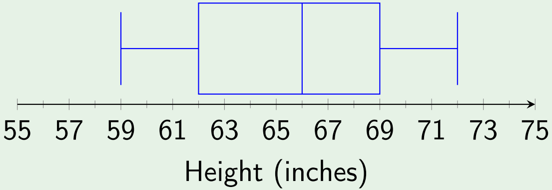
Definition

A **box plot** is a graphical representation of a five-number summary.



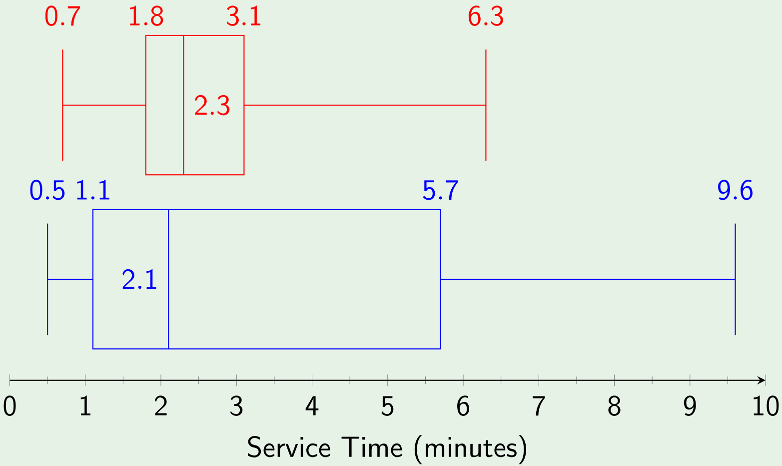
Example 21

Example 19 has the five-number summary: 59, 62, 66, 69, 72



Example 22

The box plot shows the service times for two fast-food restaurants.



What comparisons between the two restaurants can we make?
If you are in a hurry, which restaurant should you go to?