

Algebra 1
Unit 4
 Describing Data

Monday	Tuesday	Wednesday	Thursday	Friday
			April 15 th	April 16 th
			Summary Statistics	
April 19 th	April 20 th	April 21 st	April 22 nd	April 23 rd
Data Displays Data Distributions	Comparing Data Unit 4 Quiz (due at midnight)	Two Way Frequency Tables	Two Way Frequency Tables	
April 26 th	April 27 th	April 28 th		
Linear Regression	Review Unit 4 Test (opens)	Unit 4 Test (due at midnight)		

Summary Statistics

In Unit 4, we will be discussing data – both quantitative and categorical. First, we will begin with quantitative data. Quantitative data is data that is _____.

Some examples of quantitative data are:

When given a list of data, there are several different calculations we can give to describe the data.

The Five Number Summary

The five number summary is a numerical summary of your data. It has five parts:

Minimum: _____

Lower Quartile: _____

Median: _____

Upper Quartile: _____

Maximum: _____

Practice: Find the five number summary for each of the following data sets.

If you are calculating by hand, the first thing you must do is order your data from least to greatest

10 18 14 13 17 12 12 18 14

20 28 24 26 33 25 29 27 24 26

Measures of Center

Mean: _____

Median: _____

Mode: _____

Measures of Spread

Range: _____

Interquartile Range: _____

Mean Absolute Deviation: _____

Below is a list of mens heights. Find the five number summary, mean, mode, range, and interquartile range for this data.

71	63	67	69	76	76	73	68
71	75	69	62	69	73	64	72

Ordered data:

Minimum: _____

Lower Quartile: _____

Median: _____

Upper Quartile: _____

Maximum: _____

Mean: _____

Mode: _____

Range: _____

Interquartile Range: _____

Summary Statistics – Using Calculator

To calculate summary statistics on your graphing calculator, follow these steps.

1) Press **STAT**, then **ENTER**.

2) To clear L₁, press **↑** to go to the very top, **CLEAR**, **↓**.

★NEVER PRESS DELETE TO CLEAR THE LIST.★

3) Type your data into L₁. Be sure to hit **ENTER** after each number. You do NOT need to order your data first.

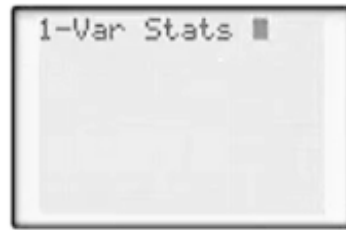
★ALWAYS TYPE YOUR DATA INTO L1.★

4) Once all of your data is entered, press **STAT**, then **→** and **ENTER**.

5) Your screen will then look like one of the following:

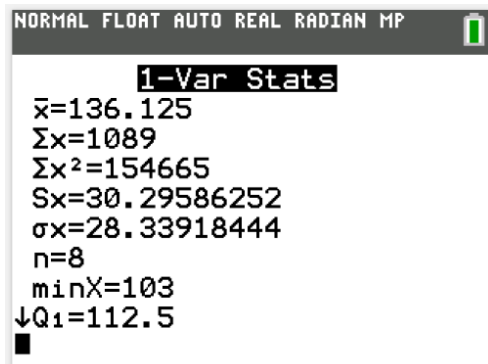


Make sure it says List: L₁ and that the FreqList is blank. Press **ENTER** until screen changes.



Press **ENTER**. Your calculator defaults to using L₁.

6) Your screen will then change to this:



Press **↓** and **↑** to scroll down and up to find the information that you need.

\bar{x} = mean

min = minimum

Q1 = lower quartile

med = median

Q3 = upper quartile

max = maximum

You will have to calculate the range and interquartile range.

Using your calculator, find the summary statistics listed below for the following data set.

Below is a list of hours slept on school nights for a sample of high school students. Round the nearest tenth if necessary.

9 6 6 5 7 12 7 5 7 6 8.5 11 9 6.5

Minimum		Lower Quartile		Median		Upper Quartile	
Maximum		Average		Range		Interquartile Range	

Summary Statistics – Practice

For the following data sets, find the five number summary, mean, range, and interquartile range. Round the nearest hundredth if needed.

1) Academy Awards

Movie	# Awards	Movie	# Awards
The Bridge on the River Kwai	7	Lord of the Rings: Return of the King	11
On the Waterfront	8	The French Connection	5
Ordinary People	4	All Quiet on the Western Front	2
A Beautiful Mind	4	My Fair Lady	8
Crash	3	Ben-Hur	11
The Deer Hunter	5	The English Patient	9
Mutiny on the Bounty	1	Titanic	11
West Side Story	10		

Minimum		Lower Quartile		Median		Upper Quartile	
Maximum		Average		Range		Interquartile Range	

2) Age at First Job (in years)

15 17 13 14 18 14 15 16
 20 21 15 17 17 15 13

Minimum		Lower Quartile		Median		Upper Quartile	
Maximum		Average		Range		Interquartile Range	

3) Test Scores (points)

50 39 50 49 47 55 44 39 41 48 48 53 52 39

Minimum		Lower Quartile		Median		Upper Quartile	
Maximum		Average		Range		Interquartile Range	

4) Life Expectancy (in years)

State	Years	State	Years	State	Years	State	Years
New Jersey	82.4	Ohio	81	Louisiana	78.2	Massachusetts	83.8
Florida	81.7	Tennessee	77.9	Texas	80.3	New Mexico	77.7
District of Columbia	77.9	South Carolina	78.3	West Virginia	74.1	Minnesota	80.3
South Dakota	74.3	New York	82.5	Mississippi	74.2		

Minimum		Lower Quartile		Median		Upper Quartile	
Maximum		Average		Range		Interquartile Range	

5) Mountain Heights

Name	Feet	Name	Feet	Name	Feet	Name	Feet
Saser Kangri II E	24,649	Annapurna Dakshin	23,684	Rimo I	24,229	Diran	23,839
Chogolisa	25,148	Sia Kangri	24,350	Saser Kangri III	24,590	Singhi Kangri	23,629
Malubiting	24,469	Mamostong Kangri	24,659	Teram Kangri III	24,219	Baintha Brakk	23,901
Jongsong Peak	24,482	Kamet	25,446	Annapurna I	26,545	Ngadi Chuli	25,823
Annapurna II	26,040						

Minimum		Lower Quartile		Median		Upper Quartile	
Maximum		Average		Range		Interquartile Range	

6) Number of Runs in a Baseball Game

12 8 19 4 5 4 12 3 1 8 5 6 7 9

Minimum		Lower Quartile		Median		Upper Quartile	
Maximum		Average		Range		Interquartile Range	

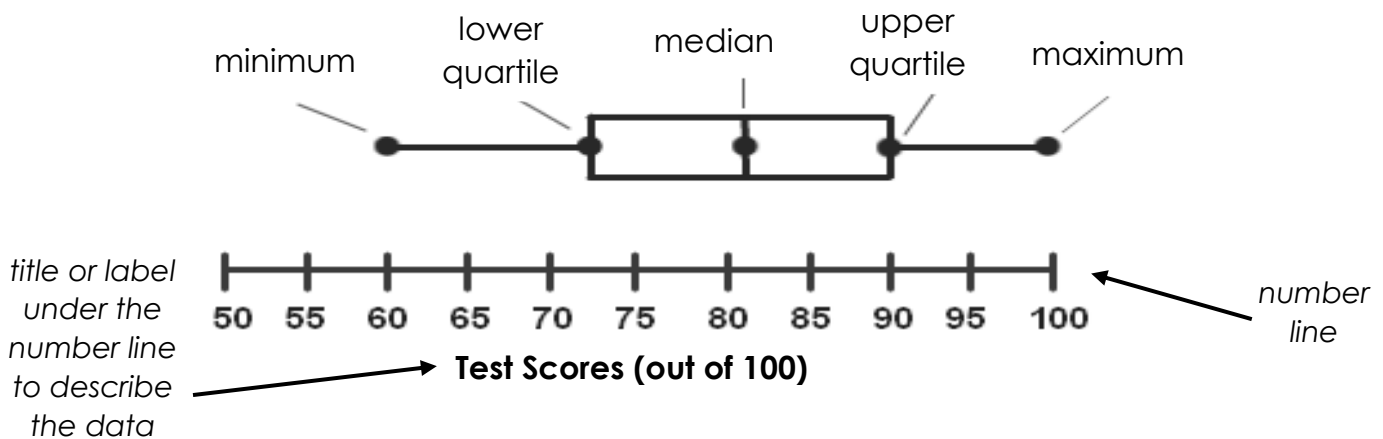
Data Displays

Box Plots

a graphical representation of the five number summary

A **box plot** (also called box and whisker plot) is used to show how data values are distributed. They are created using five important numbers that show the minimum, maximum, median, lower quartile, and upper quartile.

In a box plot, a rectangle is drawn starting at the first quartile and ending at the third quartile. The rectangle shows the middle 50% of the data set. The median is represented by a line. Whiskers are drawn from the rectangle to the minimum and maximum data values. An example of a box plot is below:

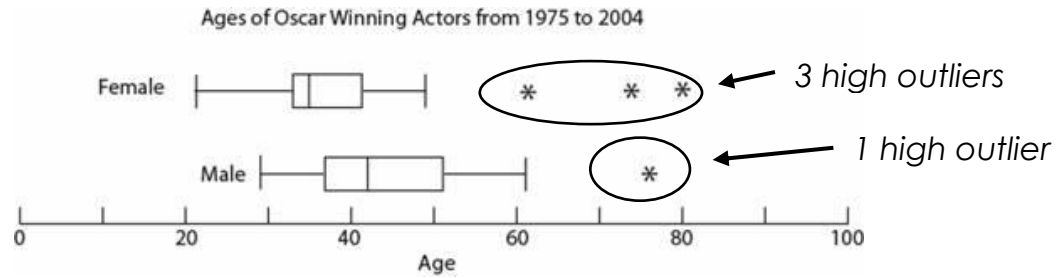


Shapes of Box Plot Distributions (Types of Distributions)

SHAPE/TYPE	DESCRIPTION	PICTURE
SYMMETRIC	When graphed, a vertical line drawn at the center will form mirror images. This shape is referred to as the bell shaped curve or normal curve The median and mean will be approximately equal.	
UNIFORM	The data is spread equally (or very close to equally) across the range. Uniform distributions are a type of symmetric distributions. The median and mean will be approximately equal.	
SKewed LEFT (NEGATIVE SKEW)	Fewer data points are found to the left of the graph. The "tail" of the graph is to the left. The mean is less than the median.	
SKewed RIGHT (POSITIVE SKEW)	Fewer data points are found to the right of the graph. The "tail" of the graph is to the right. The mean is greater than the median.	

Outliers: A data value that lies on the outside of all the other data values. It is denoted by an asterisk (*) or dot.

Data sets can have no outliers, one outlier, or multiple outliers. Outliers can be found on the low end of the data, the high end of the data, or both.



Picking the Best Measure of Center

Mean: _____

Median: _____

Mode: _____

Picking the Best Measure of Spread

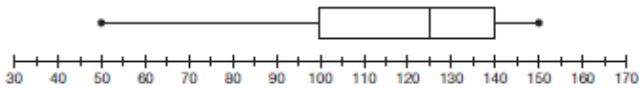
Interquartile Range: _____

Mean Absolute Deviation: _____

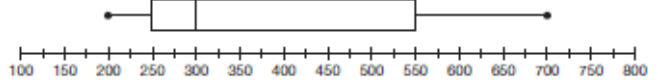
Identifying Distributions

Identify the type of distribution of the following box plots. Then identify the best measure of center and spread.

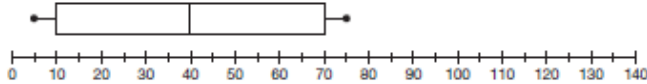
a.



b.



c.



Creating Box Plots

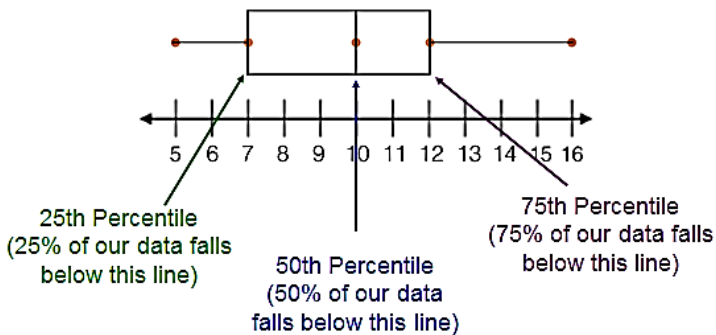
The data below represents the age at which various students obtained their first job. Calculate the five number summary (by hand or with your calculator) and create a box plot to represent this data.

{15, 11, 12, 18, 10, 21, 11, 17, 13}

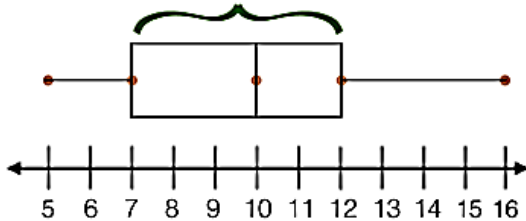
Min: _____ Q1: _____ Med: _____ Q3: _____ Max: _____



Interpreting Box Plots



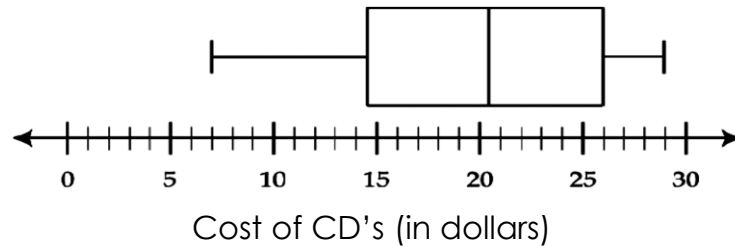
Interquartile Range (IQR)
Middle 50% of data
Q1 - Q3



- 1) 25% of data values fall at or below _____.
- 2) 25% of data values fall at or above _____.
- 3) 50% of data values fall at or above _____.
- 4) 75% of data values fall at or above _____.
- 5) 75% of data values fall at or below _____.
- 6) The middle 50% of data values fall between _____ and _____.
- 7) 100% of data values fall between _____ and _____.

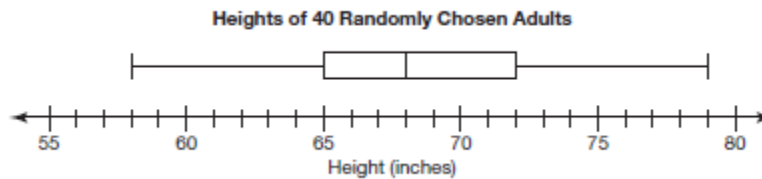
Practice with Box Plots

Example 1: Analyze the box plot below about the cost, in dollars, of 12 CD's. Answer the questions.



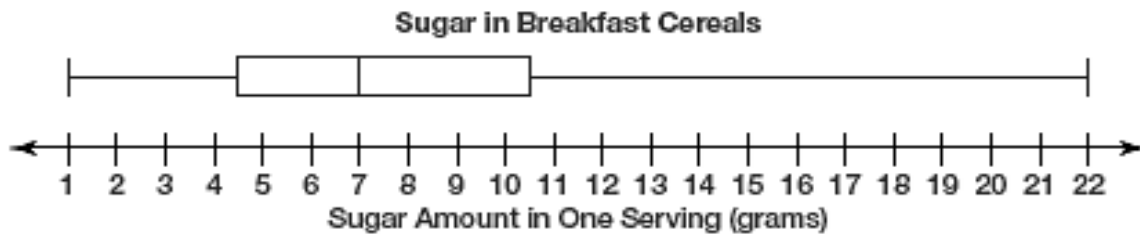
- | | |
|--|--|
| A. Which cost is the upper quartile? | B. What is the range? |
| C. What is the median? | D. Which cost represents the 100 th percentile? |
| E. How many CD's cost between \$14.50 and \$26.00? | F. How many CD's cost less than \$14.50? |

Example 2: Analyze the box plot below and answer the following questions:



- | | |
|--|---|
| A. What is the height range of the middle 50 percent of the surveyed adults? | B. How many of the surveyed adults are between 72 and 79 inches? |
| C. What percent of the surveyed adults are 72 inches or shorter? | D. What is the height of the tallest adult surveyed? |
| E. About 10 people have a height below what amount? | F. About 20 people have a height above amount? |
| G. How many of the surveyed adults are at least 58 inches tall? | H. Describe the distribution. Is the median or mean best describe the data? |

Example 3: Jamie has organized the amount of sugar, per serving, in 60 different cereals and created a box plot of his data below:



a. State the numbers (including what they represent) for the five number summary.

b. Give three conclusions that can be made about the sugar amount in one serving of breakfast cereal.

□ _____

□ _____

□ _____

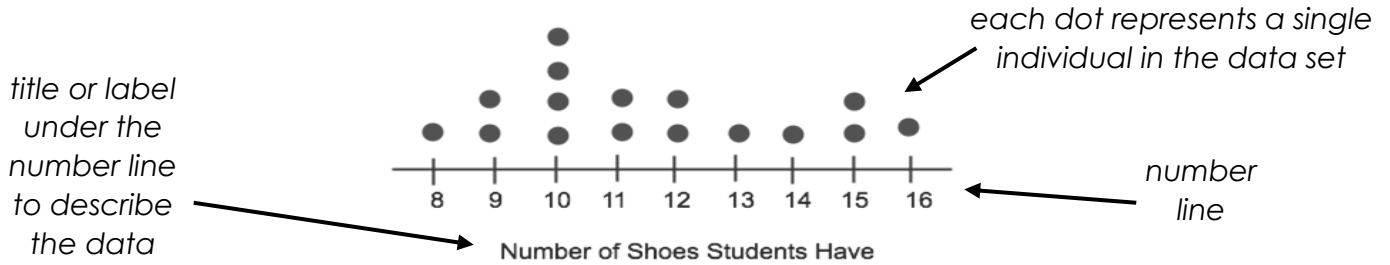
c. Describe the distribution and interpret the meaning of the distribution in terms of this problem situation.

d. Jamie says that more breakfast cereals have over 10 grams of sugar per serving than have under 5 grams of sugar per serving because the whisker connecting Q3 to the maximum is longer than the whisker connecting Q1 to the minimum. Is he correct? Explain why or why not.

Dot Plots

shows all data points

A **dot plot** is a data representation that uses a number line and x's, dots, or other symbols to show frequency. The number of times a value is repeated corresponds to the number of dots above that value. A dot plot also shows the size of the data set.



Advantages of Dot Plots: Simple to make
Shows each individual data point

Disadvantages of Dot Plots: Can be time consuming with lots of data points
Have to count to get exact total
Fractions are hard to display

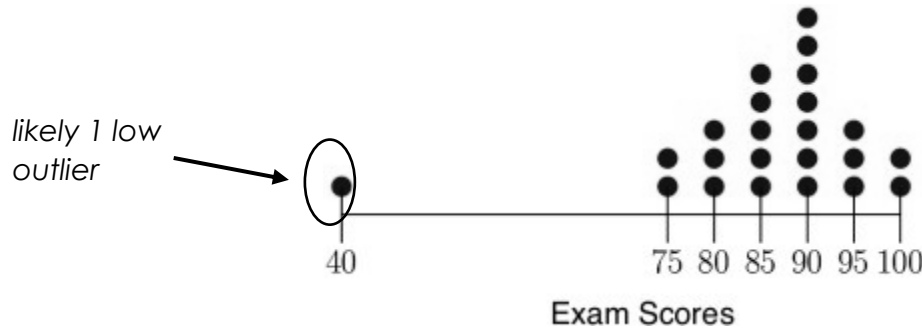
Shapes/Types of Dot Plot Distributions

SHAPE/TYPE	DESCRIPTION	PICTURE
SYMMETRIC	When graphed, a vertical line drawn at the center will form mirror images. This shape is referred to as the bell shaped curve or normal curve The median and mean will be approximately equal.	
UNIFORM	The data is spread equally (or very close to equally) across the range. Uniform distributions are a type of symmetric distributions. The median and mean will be approximately equal.	
SKEWED LEFT (NEGATIVE SKEW)	Fewer data points are found to the left of the graph. The "tail" of the graph is to the left. The mean is less than the median.	
SKEWED RIGHT (POSITIVE SKEW)	Fewer data points are found to the right of the graph. The "tail" of the graph is to the right. The mean is greater than the median.	

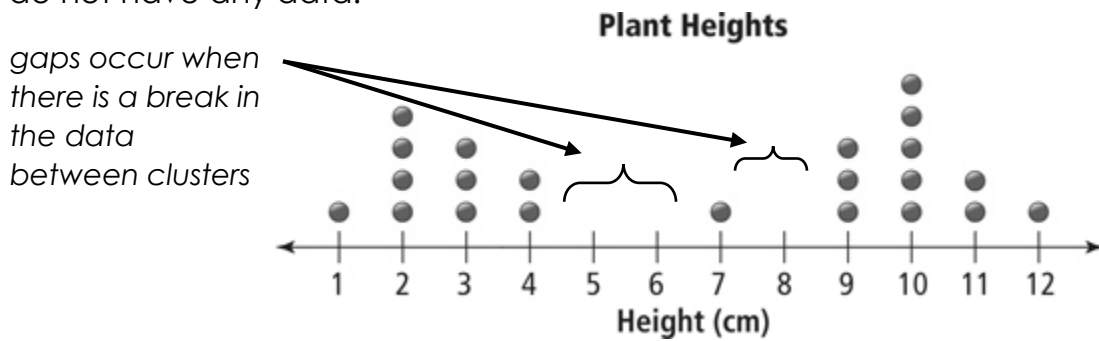
Other Characteristics

Data distributions modeled by dot plots can still have **outliers** – the outliers however are not denoted with any special symbols. Outliers will typically be seen as dots that are far from the majority of the data.

Data sets can have no outliers, one outlier, or multiple outliers. Outliers can be found on the low end of the data, the high end of the data, or both.



Gaps occur in dot plots when there are values on the number line at the bottom that do not have any data.



We can also describe the shapes of dot plots by describing the number of peaks the data has.

Bi-Modal Distributions have ____ peaks.

Uni-Modal Distributions have ____ peak.



Picking the Best Measure of Center

Mean: _____

Median: _____

Mode: _____

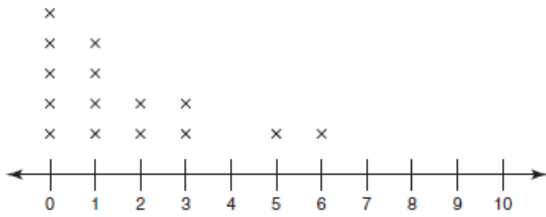
Picking the Best Measure of Spread

Interquartile Range: _____

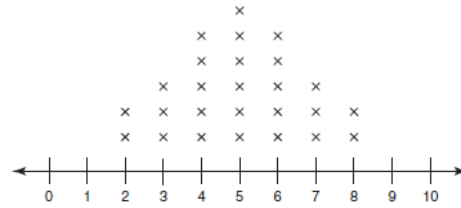
Mean Absolute Deviation: _____

Practice 1: Identify the type of distribution of the following dot plots. Then identify the best measure of center and spread.

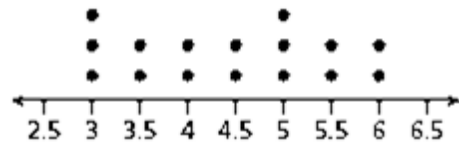
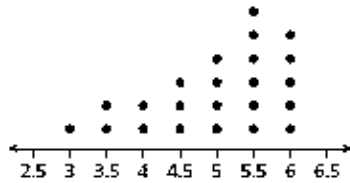
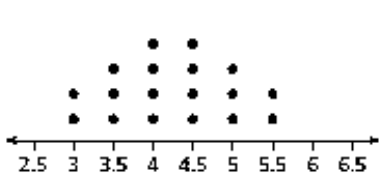
a.



b.



Practice 2: Determine the indicated information for each dot plot below.



Mean:

Mean:

Mean:

Median:

Median:

Median:

Mode:

Mode:

Mode:

Range:

Range:

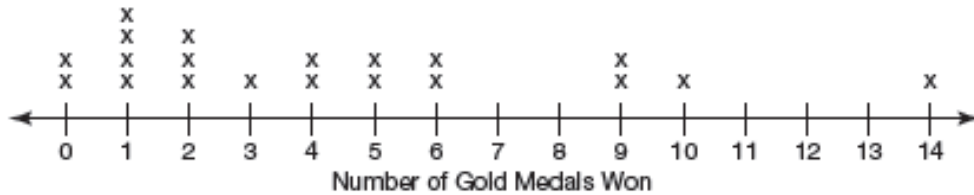
Range:

Distribution shape:

Distribution shape:

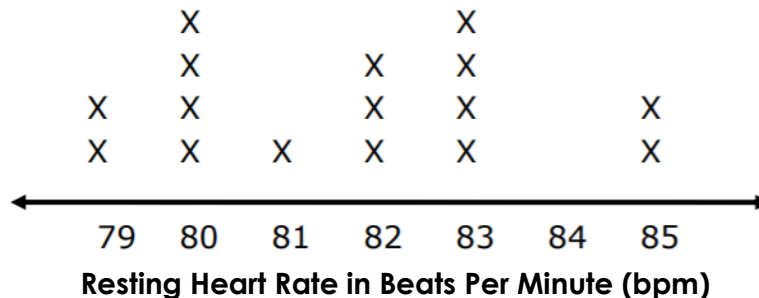
Distribution shape:

Practice 3: The following dot plot represents gold medals won at the Special Olympics:



- How many participants are represented in the dot plot?
- How many participants won 10 or more medals?
- How many participants won less than 4 medals?
- Describe the data distribution and interpret its meaning in terms of this problem situation.

Practice 4: The dot plot below shows the resting heart rate for 16 male students at Harrison. Use this to answer the following questions.

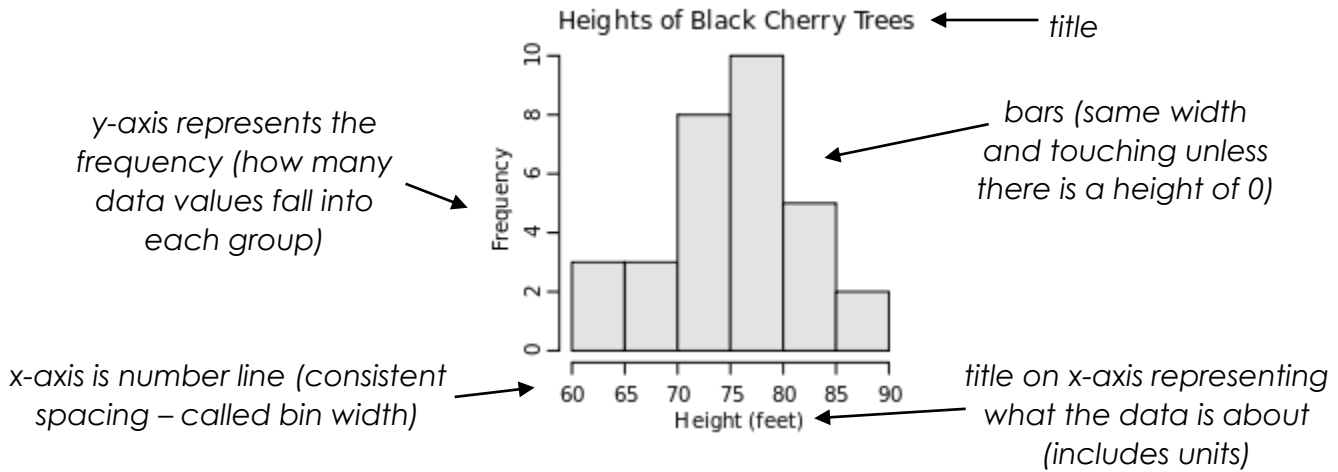


- How many students had a resting heart rate of at least 83 bpm?
- How many students had a rest heart rate of at most 81 bpm?
- Write one sentence drawing a conclusion from the data.

Histograms

displays all the data by grouping

A **histogram** is a bar graph used to display the frequency of data divided into equal intervals, called **bins**. The bars must be of equal width and should touch, but not overlap. The height of each bar gives the frequency of the data.



Advantages of Histograms: Good for determining the shape of data
 Convenient for representing large quantities of data

Disadvantages of Histograms: Cannot read exact values because data is grouped into categories
 More difficult to compare two data sets because measures of center and spread cannot be determined

SHAPE/TYPE	DESCRIPTION	PICTURE
SYMMETRIC	When graphed, a vertical line drawn at the center will form mirror images. This shape is referred to as the bell shaped curve or normal curve The median and mean will be approximately equal.	
UNIFORM	The data is spread equally (or very close to equally) across the range. Uniform distributions are a type of symmetric distributions. The median and mean will be approximately equal.	
SKEWED LEFT (NEGATIVE SKEW)	Fewer data points are found to the left of the graph. The "tail" of the graph is to the left. The mean is less than the median.	
SKEWED RIGHT (POSITIVE SKEW)	Fewer data points are found to the right of the graph. The "tail" of the graph is to the right. The mean is greater than the median.	

Other Characteristics

We can also describe the shapes of dot plots by describing the number of peaks the data has.

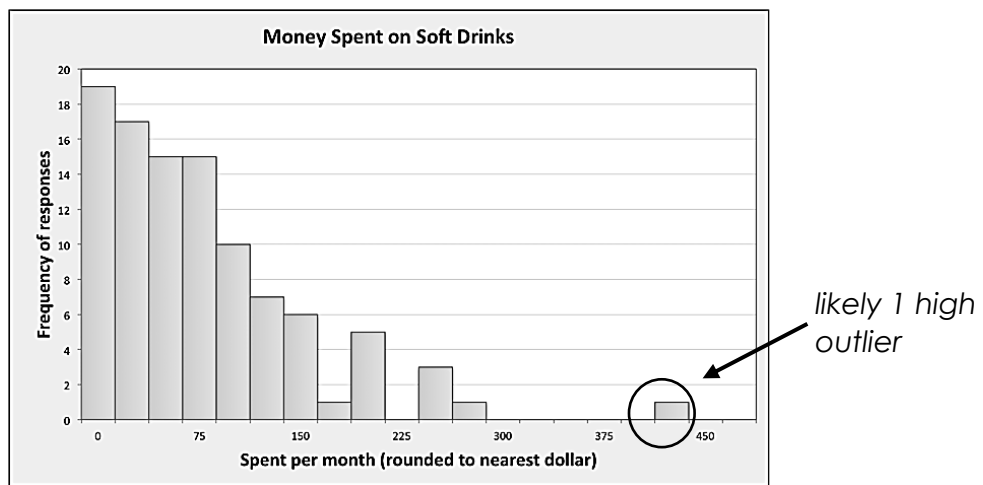
Bi-Modal Distributions have ____ peaks.

Uni-Modal Distributions have ____ peak.

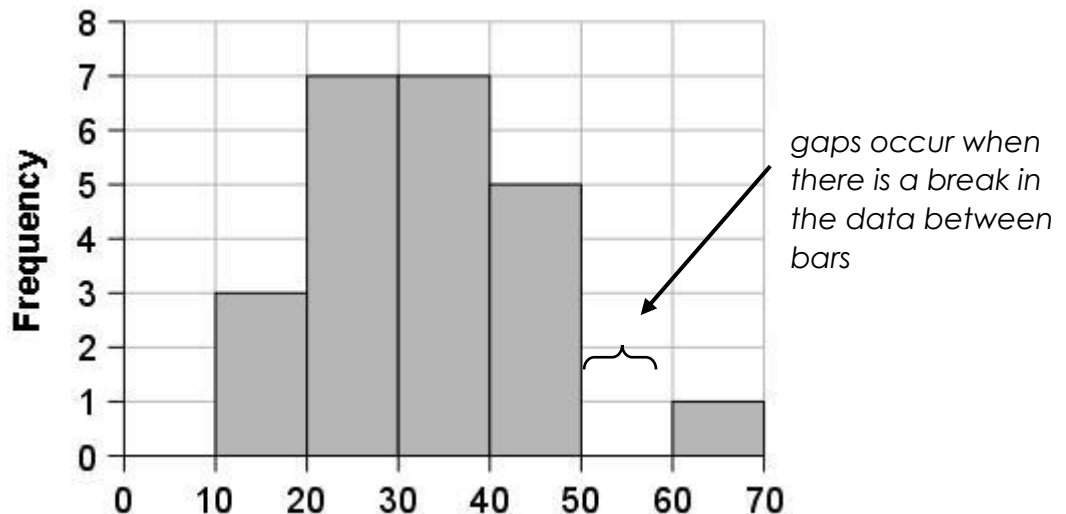


Data distributions modeled by dot plots can still have **outliers** – the outliers however are not denoted with any special symbols. Outliers will typically be seen as dots that are far from the majority of the data.

Data sets can have no outliers, one outlier, or multiple outliers. Outliers can be found on the low end of the data, the high end of the data, or both.



Gaps occur in dot plots when there are values on the number line at the bottom that do not have any data.



Picking the Best Measure of Center

Mean: _____

Median: _____

Picking the Best Measure of Spread

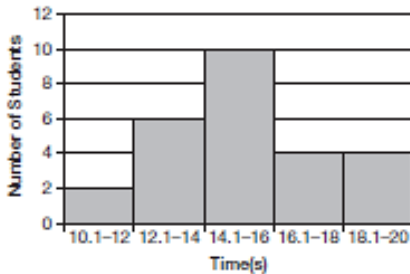
Interquartile Range: _____

Mean Absolute Deviation: _____

Histogram Practice

Practice 1: Describe the distribution of each histogram and if the mean is less, greater, or equal to the median. Then describe which would be a better measure of center; the median or mean.

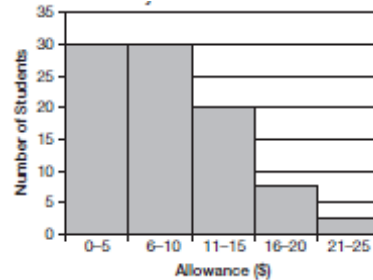
a.



b.

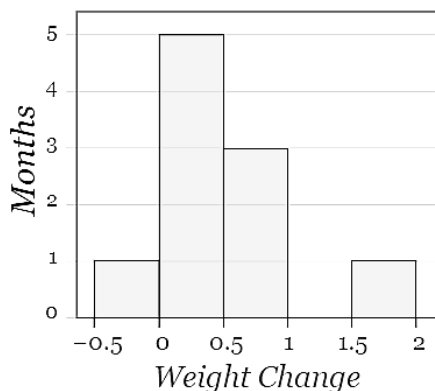


c.



Practice 2: Use the histogram to answer the following questions about how the weight of puppies changes in 5 months.

Puppy Weight Change



a. How many puppies are represented by the histogram?

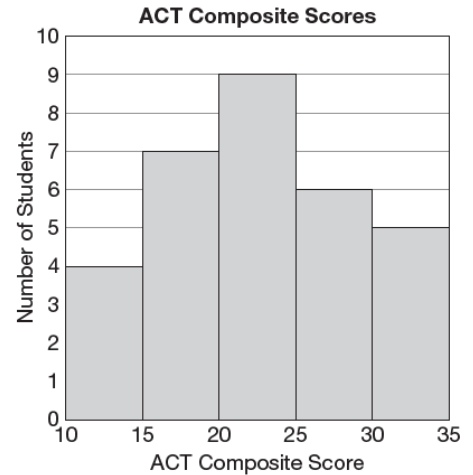
b. How many puppies did not lose weight?

c. Based on the histogram, can you determine how many puppies gained 0.75 pounds?

d. How many puppies gained at least 0.5 pounds?

Practice 3: Analyze the given histogram which displays the ACT composite score of several randomly chosen students.

a. Describe the distribution and explain what it means in terms of the problem situation.



b. How many students had an ACT score of at least 20?

c. How many students had an ACT score less than 30?

d. How many students had an ACT score of exactly 25?

Practice 4: The histogram below shows how much time a sample of freshmen at Harrison High School spend playing video games. Use this graph to answer the following questions.



a. What percentage of students played video games for less than 10 hours?

b. How many students played video games for at most 20 hours?

c. How many students played video games for at least 20 hours?

d. How many students played video games between 5 and 15 hours?

e. What is the maximum amount of time spent playing video games by a single student?

f. Would the mean number of hours spent playing video games be greater or less than the median number of hours spent playing video games? Explain how you know. (Hint: you do not need to do any calculations)

Comparing Data Sets

Scenario: Coach Webb is trying to decide which two of his point guards he wants to start for the first round of play-offs. The data below shows the numbers of points scored by Jace and Tyler from the past six games.

Jace: 11, 11, 6, 26, 6, 12

Tyler: 15, 12, 13, 10, 9, 13

1. Without doing any calculations, who do you think Coach Webb should select as a starting player and why?

2. Identify the mean score for each player.

Jace: _____

Tyler: _____

3. Calculate the deviations for the points scored for each player. Then describe the deviation (ie: are they above the average, below the average, equal to the average)

Jace	
Points Scored	Describe Deviation
11	
11	
6	
26	
6	
12	

Tyler	
Points Scored	Describe Deviation
15	
12	
13	
10	
9	
13	

What do you notice about the deviations for each player?

4. For each player, find the sum of the deviations and then divide by the number of data values. Round your final answer to the nearest hundredth.

Jace Sum of Deviations: _____ Divided by # of values = _____

Tyler Sum of Deviations: _____ Divided by # of values = _____

5. The numbers that you just calculated are the mean absolute deviations (MAD). What does the mean absolute deviation tell you about each player?

6. If you were Coach Webb, which player would you choose to start in the play-off game and why?

Comparing Measures of Center and Spread

Measures of Center

- Mean: _____
- Median: _____
- Mode: _____

If the data is symmetric, the mean is _____ the median.

If the data is skewed left, the mean is _____ the median.

If the data is skewed right, the mean is _____ the median.

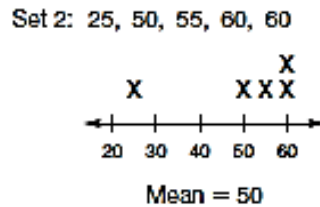
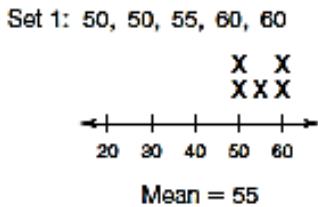
Measures of Spread

- Interquartile Range: _____
- Mean Absolute Deviation: _____

If the data is more spread, than the data values are _____ spread out, which means they are _____ consistent. This results in the MAD being _____.

If the data is less spread, than the data values are _____ spread out, which means they are _____ consistent. This results in the MAD being _____.

Example 1: Which data set will have the greater mean absolute deviation? Explain how you know.



Example 2: The following data represents test scores from Unit 11 test.

Unit 11 Test Scores: 81, 41, 89, 92, 80, 86, 77, 66, 84, 92, 97, 88, 77

a. Compare the mean and median.

b. What type of distribution does the data create? What does this mean?

Continued from previous page – data is copied below.

Unit 11 Test Scores: 81, 41, 89, 92, 80, 86, 77, 66, 84, 92, 97, 88, 77

c. Do you think there are any outliers?

d. Which measure of center best describes the grades and why?

e. Which measure of spread best describes the grades and why?

f. Test scores from the Unit 12 Test are below.

Unit 12 Test Scores: 55, 89, 92, 75, 64, 87, 92, 91, 77, 63, 50, 81

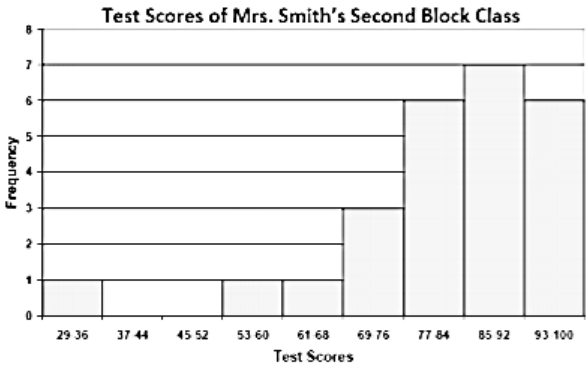
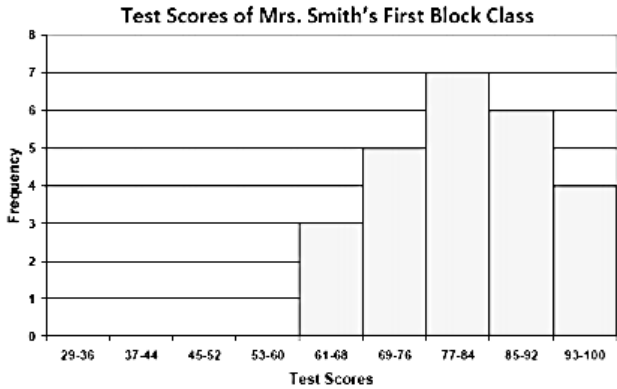
Would the test scores of the Unit 11 Test or the Unit 12 Test have a lower mean absolute deviation (MAD)? Explain how you know without doing calculations.

g. Which test had a larger range of scores – the Unit 11 Test or the Unit 12 Test?

h. Which test had the lowest minimum score – the Unit 11 Test or the Unit 12 Test?

i. Which test had the highest maximum score – the Unit 11 Test or the Unit 12 Test?

Example 3: The histograms below show the scores of Mrs. Smith's first and second block class at Red Rock High School.

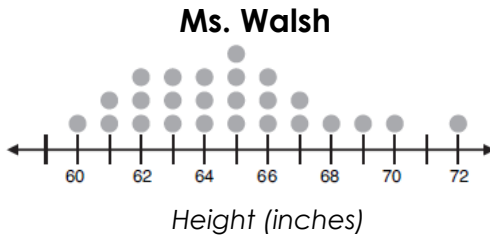
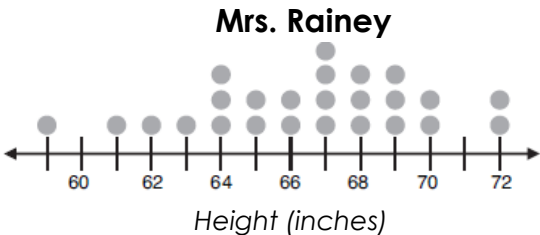


Compare the histograms using the following information:

- 69 and above is passing
- 68 or below is failing

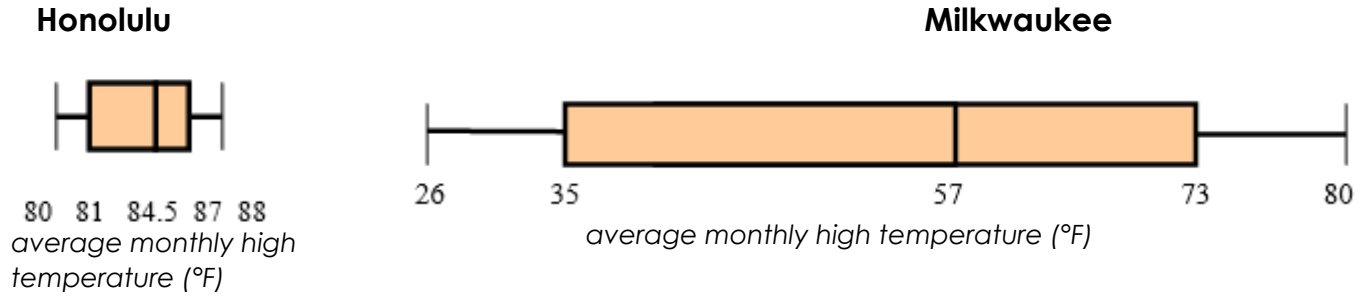
1. How many students are in her 1st and 2nd block class?
2. How many students failed the test in each class?
3. Which measure of center best describes the data and why?
4. Which class seemed to do better overall?

Example 4: Each girl in Mrs. Rainey's class and Ms. Walsh's class measured their own height. The heights were plotted on the dot plots below. Use the dot plots to compare the heights of the girls in the two classes.



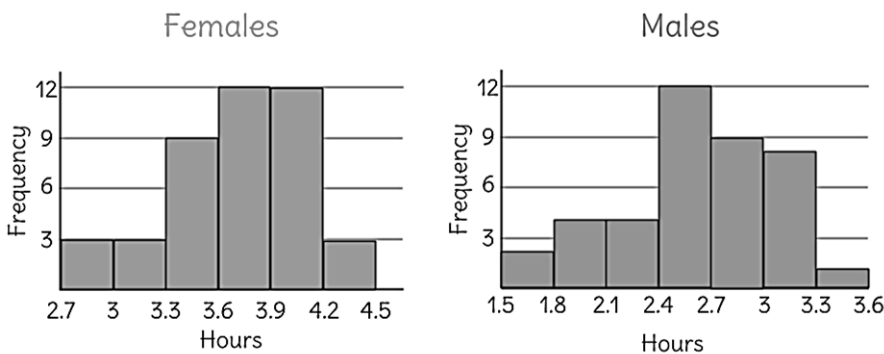
- a. Describe the distribution for each class.
- b. Which teacher has taller female students? How do you know by looking at the dot plots?

Example 5: The following box plots show the average monthly high temperatures for two years in Milwaukee and Honolulu. Use the box plots to answer the following questions.



- What was the median temperature for both cities?
- What was the range for both cities?
- Which city has more spread in its data and why?
- Which city has a larger maximum?
- Which city has a smaller minimum?

Examples 6: The histograms below represent the amount of time students spend on their homework.



- Typically, who spends more time on homework – females or males?
- How many females spend less than 3.3 hours on homework?
- How many males spend less than 3.3 hours on homework?

Two Way Frequency Tables

A **two way table** is a useful way to organize data that can be categorized by two variables.

Below is an example of a two way frequency table.

One variable is shown on the top while the other is on the far left side.

The frequencies (counts) on the bottom and far right are known as **marginal frequencies** – they tell you the total for each category.

	Sport Utility Vehicle (SUV)	Sports Car	Totals
male	21	39	60
female	135	45	180
Totals	156	84	240

The frequencies (counts) in the middle are known as **joint frequencies** – they join the two variables together.

The bottom right gives you the grand total of the frequencies. It is equal to the sum of the totals for each variable.

MathBits.com

The following table shows the results of a poll of randomly selected high school students and their preference for either math or English. Before answering the questions below, calculate the marginal frequencies and grand total.

	9 th Grade	10 th Grade	11 th Grade	12 th Grade	Total
Math	10	12	11	8	
English	12	11	8	8	
Total					

- How many students are in 11th grade?
- How many students are in 9th grade and prefer math?
- How many students prefer English and are in 12th grade?
- How many students are there total?

Two Way Frequency Tables Practice

1) The table below shows the results from a survey given to freshmen at Harrison. Fill in the missing values into the table below and then answer the following questions:

9th Grader's School Transportation Survey

	Male	Female	Total
Walk		46	
Car	28		45
Bus		12	27
Bike		17	69
Total	129	92	

a. How many students are there total?

b. How many 9th boys walk to school?

c. How many 9th girls ride their bike to school?

d. How many males took the survey?

2) The table below represents the favorite meals of 9th and 10th graders. Use the table to answer the following questions.

Favorite Meals of Students

		Burgers	Chicken Nuggets	Pizza	Salad Bar	Total
Grade Level	9th grade	4	1	3	5	13
	10th grade	3	7	3	4	17
	Total	7	8	6	9	30

a. How many 9th graders participated in the survey?

d. Which meal is the least favorite of all students?

b. How many students prefer chicken nuggets?

e. Which meal is the least favorite of 9th graders?

c. How many students prefer burgers?

f. Which meal is most favorite of 10th graders?

Relative Frequencies

A **relative frequency** is the frequency that an event occurs divided by the total number of events.

Example: If your team has won 9 games from a total of 12 games played...

The **frequency** of winning is _____.

The **percent** of games won is _____.

The **relative frequency** of winning is _____.

Below is the two-way frequency table that we initially looked at. It shows the results of a poll of randomly selected high school students and their preference for either math or English.

	9 th Grade	10 th Grade	11 th Grade	12 th Grade	Total
Math	10	12	11	8	41
English	12	11	8	8	39
Total	22	23	19	16	80

On the table below, use the information from the original table to calculate the joint relative frequencies and marginal relative frequencies.

To calculate **joint relative frequencies**, take each joint frequency and divide by the grand total. Round to the nearest thousandth for this example.

To calculate **marginal relative frequencies**, find the sum of the joint relative frequencies for each row and column. Round to the nearest thousandth for this example.

	9 th Grade	10 th Grade	11 th Grade	12 th Grade	Total
Math					
English					
Total					

- What percent of students are in 10th grade and like English?
- What percent of students like math?
- What percent of students like Math and are in 12th grade?
- What percent of those surveyed were seniors?

Practice with Relative Frequencies

1) One hundred people who frequently get migraine headaches were chosen to participate in a study of new anti-headache medicine. Some of the participants were given the medicine; others were not. After one week, the participants were asked if they got a headache during the week. The two way frequency table summarizes the results. Create a table showing the joint relative frequencies and marginal relative frequencies. Round to the nearest hundredth for this problem.

	Took Medicine	Did NOT Take Medicine
Headache	12	15
No Headache	48	25

	Took Medicine	Did NOT Take Medicine
Headache		
No Headache		

- What is the relative frequency of participants that had a headache?
- What is the relative frequency of participants that did NOT take the medicine AND had a headache?

2) Create a relative frequency table to represent the favorite movies of students.

	Comedy	Drama	Horror
Class A	20	8	3
Class B	18	6	9

a. What percent of people prefer to watch comedies?

b. What percent of people prefer to watch horror movies?

	Comedy	Drama	Horror
Class A			
Class B			

c. What percent of people are from class A and prefer to watch drama movies?

d. Which class prefers watching horror movies?

Conditional Frequencies

A **conditional frequency** is restricted to a particular group (or subgroup). Conditional frequencies are typically identified by the words “given that” or “if” or “what percent of (insert condition)”. They do NOT come from the total data, but from a row or column total.

To calculate a conditional frequency, divide the joint frequency by the marginal frequency (does not matter if they are the frequencies or relative frequencies).

	Took Medicine	Did NOT Take Medicine	TOTAL
Headache	12	15	27
No Headache	48	25	73
TOTAL	60	40	100

★ Note: When a question asks you to find the probability, it is asking for the relative frequency. This means that your answer should be in the form of a _____.

1. What is the probability that a participant did not get a headache if they took the medicine?
2. What is the probability that a participant took medicine given they did not have a headache?
3. What is the probability that a participant took medicine given they did have a headache?
4. If we know that a participant did not take medicine, what is the probability that they had a headache?
5. How many participants took the medicine?
6. How many participants took the medicine and did not have a headache?
7. What percent of participants had a headache and took medicine?

Conditional Frequencies Practice

1) Students were surveyed about whether or not they have a pet and if they are allergic or not to animals. The results are below:

	Has a Pet	Does Not Have a Pet
Allergic to Animals	32	192
Not Allergic to Animals	213	63

- a. What percent of those surveyed who are allergic to animals have a pet?
- b. What is the probability that a person who is not allergic to animals has a pet?
- c. Given that someone has a pet, what is the probability that they are allergic to animals?
- d. What percent of those who have a pet are not allergic to animals?

2) The following contains the scores of the latest math project. Use the table to answer the following questions:

Project Scores

	Male	Female
A	9	12
B	18	14
C	8	11
D	2	3
F	1	2

- a. What percentage of males earned a score of an "A"?
- b. What percentage of those who earned an "A" were male?
- c. What percentages of females earned a score of a "B"?
- d. What percentage of those who earned an "F" were female?
- e. Given that a student earned a "C", what is the probability that they are male?

Two Way Frequency Tables – Mixed Practice

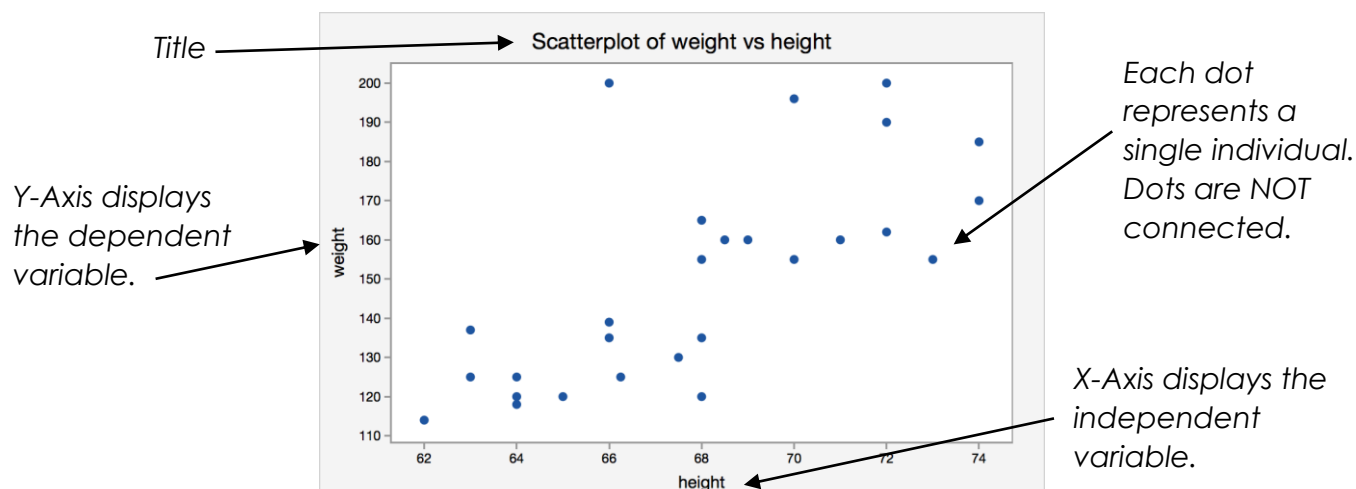
The table below shows concession stand sales at last night's soccer game. Use this information to answer the questions on this page.

	Soda	Water	No Drink
Hot Dog	50	62	46
Pizza	120	58	4
No Food	30	20	10

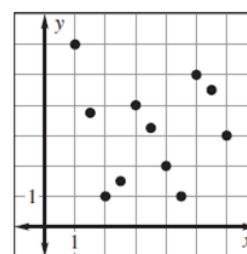
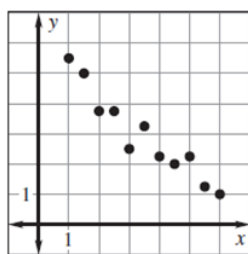
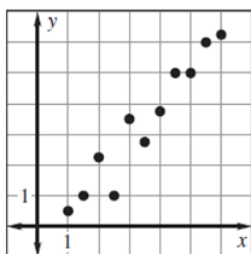
- 1) How many people ordered a soda?
- 2) Given that a person ordered a hot dog, what is the probability that they also ordered a soda?
- 3) What is the relative frequency of people of ordered pizza and water?
- 4) What percent of people ordered no food?
- 5) Find the relative frequency of people who ordered no food and no drink.
- 6) What is the probability that a person ordered soda and pizza?
- 7) What is the probability that a person ordered no drink given that they ordered pizza?
- 8) We know that Sarah ordered a soda. What is the probability that she ordered a hot dog?
- 9) What is the probability that a person ordered no drink?
- 10) How many people ordered only pizza?

Scatterplots and Correlation

A **scatterplot** is a graph of data pairs (x, y) . An example of a scatterplot is below.



Scatterplots are typically used to describe relationships, called **correlations**, between two variables (bi-variate). The **correlation coefficient** describes how well a line fits the data. A **trend line** can be drawn to help determine correlation.

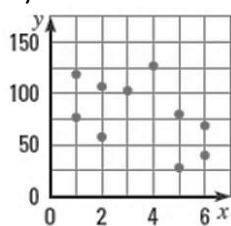


Positive Correlation	Negative Correlation	No Correlation
<ul style="list-style-type: none"> ▫ as x increases, y increases ▫ positive slope ▫ Correlation Coefficient is close to 1 	<ul style="list-style-type: none"> ▫ as x increases, y decreases ▫ negative slope ▫ Correlation Coefficient is close to -1 	<ul style="list-style-type: none"> ▫ no relationship between x and y ▫ Correlation Coefficient is close to 0

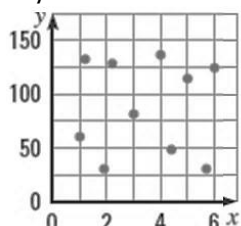
Correlation Coefficient Guidelines			
0.70 to 1.00	Strong Positive	-0.70 to -1.00	Strong Negative
0.30 to 0.69	Moderate Positive	-0.30 to -0.69	Moderate Negative
0.00 to 0.29	None to Weak Positive	0.00 to -0.29	None to Weak Negative

Example: Describe the correlation of each scatterplot below. Be sure to include direction and strength in your description.

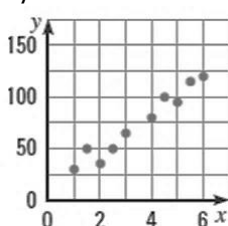
a)



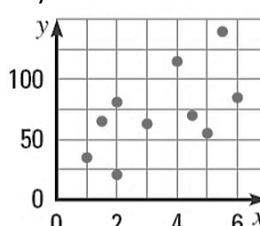
b)



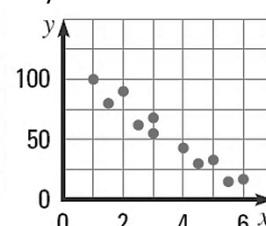
c)



d)

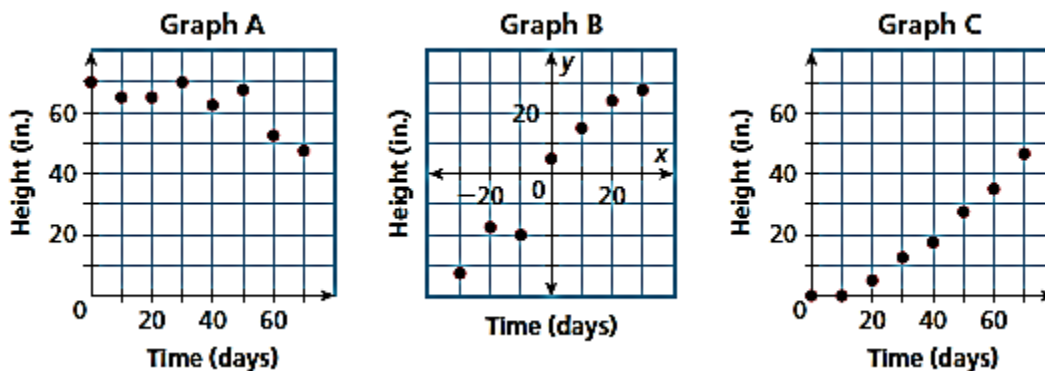


e)



Example 1: Describe the scatterplot that best describes the scenario below and explain why you chose that graph.

Scenario: *The relationship between the number of days since a sunflower seed was planted and the height of the plant.*



I think graph ____ best represents this scenario because:

Example 2: Describe the correlation you would expect to see between each pair of data sets.

Positive Correlation: When the variables go in the _____ direction. (\uparrow, \uparrow or \downarrow, \downarrow)

Negative Correlation: When the variables go in _____ directions. (\downarrow, \uparrow or \uparrow, \downarrow)

a. The number of hours you work vs the amount of money in your bank account:

positive correlation negative correlation no correlation

b. The number of hours workers receive safety training vs the number of accidents on the job:

positive correlation negative correlation no correlation

c. The number of students at Harrison vs the number of dogs in Kennesaw:

positive correlation negative correlation no correlation

d. The number of meals eaten vs the number of cars on I-75 throughout the day:

positive correlation negative correlation no correlation

e. The number of calories burned/lost vs the amount of hours you worked out:

positive correlation negative correlation no correlation

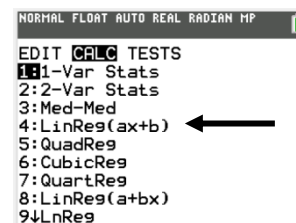
Linear Regression

When given a scatterplot of data, we can use our calculators to run a linear regression model. This linear regression model (also called the least-squares regression), would give us the BEST line of best fit for the data.

Calculating Linear Regression (TI-83/TI-84)

- 1) Press $\boxed{2^{nd}}$, $\boxed{0}$, $\boxed{x^{-1}}$, and scroll down until you see Diagnostic **On**. Press \boxed{ENTER} twice – your screen should then say “Done”
- 2) Press \boxed{STAT} , then \boxed{ENTER} .
- 3) To clear L₁, press $\boxed{\uparrow}$ to go to the very top, \boxed{CLEAR} , $\boxed{\downarrow}$. Then press $\boxed{\rightarrow}$ to go to L₂. To clear L₂, press $\boxed{\uparrow}$ to go to the very top, \boxed{CLEAR} , $\boxed{\downarrow}$. **★NEVER PRESS DELETE★**
- 4) Type your data for the x-variable into L₁. Be sure to hit \boxed{ENTER} after each number. You do NOT need to order your data first.
- 5) Press $\boxed{\rightarrow}$ to go to L₂. Type your data for the y-variable into L₂. Be sure to hit \boxed{ENTER} after each number. You do NOT need to order your data first.

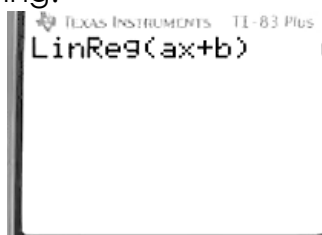
- 4) Once all of your data is entered, press \boxed{STAT} , then $\boxed{\rightarrow}$ and select option 4:LinReg(ax+b). You can do this by either scrolling down to 4 and pressing \boxed{ENTER} or just pressing the $\boxed{4}$ on the keyboard.



- 5) Your screen will then look like one of the following:



Make sure it says Xlist: L₁ and Ylist:L₂. FreqList and Store RegEQ are blank. Press \boxed{ENTER} until screen changes.



Press \boxed{ENTER} . Your calculator automatically assigns L₁ to Xlist and L₂ to Ylist.

- 6) Your screen will then change to this:

LinReg

$y=ax+b$ ← This is the form that you will write your equation.
 (It is slope intercept form just with different letters)

$a=0.0907994319$ ← a represents the slope

$b=16.26492733$ ← b represents the y-intercept

$r^2=0.6315364361$ ← r represents the correlation coefficient

$r=0.7946926677$

★ Be sure to check the signs of a , b , and r . Round to the nearest hundredth for a and b and the nearest thousandth for r unless told otherwise. ★

Linear Regression Practice

1) Using your calculator, find the linear regression model for the data below.

Taylor had guests for dinner at her house eight times and has recorded the number of guests and the total costs for each meal in the table below:

Guests	3	4	4	6	6	7	8	8
Cost (\$)	30	65	88	90	115	160	150	162

$a =$ _____ $b =$ _____ $r =$ _____

Linear Regression Model: _____

Correlation: _____

Now use your linear regression model and information above to answer the following questions.

a) Predict the cost for 2 guests to have a meal.

b) If a meal costs \$89.78, how many guests were there?

c) Interpret what the slope means in context.

When interpreting the slope, follow this general guide:

As (x context) increases by 1 (x-unit), the number of (y context) increases/decreases by (slope # with y-units).

d) Interpret what the y-intercept means in context.

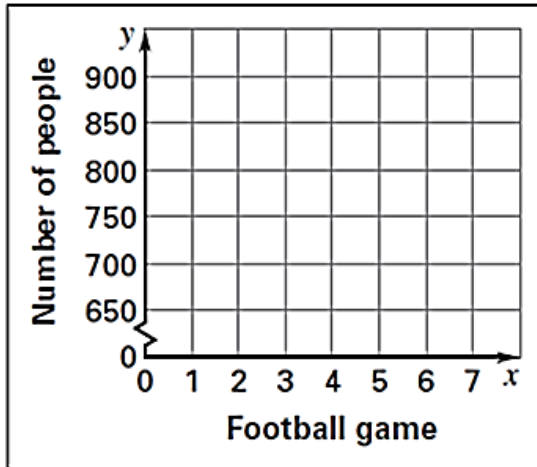
When interpreting the y-intercept, follow this general guide:

For 0 (x context), the number of (y context) is (y-intercept #).

2) The table below gives the number of people, y , who attended each of the first seven football games, x , of the season.

x	1	2	3	4	5	6	7
y	722	763	772	826	815	857	897

a) Create a scatterplot of the data.



b) Calculate the linear regression equation. Round your decimals to the nearest tenth.

$$a = \underline{\hspace{2cm}} \quad b = \underline{\hspace{2cm}}$$

Equation: $\underline{\hspace{10cm}}$

c) Identify the correlation coefficient. What does that tell you about the relationship between the two variables?

d) Interpret the slope. Does it make sense?

e) Interpret the y-intercept. Does it make sense?

f) Using your regression equation, predict the number of people at the last game if there are 9 home games this season.

3) Here is data from a group of students who measured the size of their hand in inches and counted the number of starbursts they could grab at one time.

Size of hand	6	6	7	7	7.5	8	8	9	9	9.5
# of Starbursts	30	26	31	30	31	39	29	40	43	50

a) Calculate the linear regression equation. Round all decimals to the nearest tenth.

b) Interpret the slope. Does it make sense?

c) Interpret the y-intercept. Does it make sense?

4) The table below gives the amount of time students in a class studied for a test and their test scores. Create a scatterplot. Then calculate the linear regression model.

Hours Studied	1	0	3	1.5	2.75	1	0.5	2
Test Score	78	75	90	89	97	85	81	80

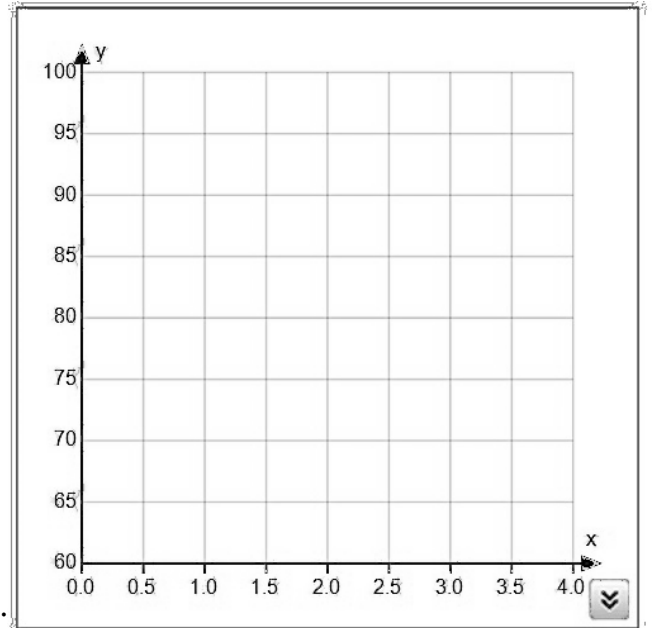
a) Linear Regression Equation:

b) Correlation Coefficient (r): _____

c) Type of Correlation: _____

d) Using the linear regression equation predict a student's test score if they studied for 4 hours.

e) Explain what the y-intercept means in context.



f) Explain what the slope means in context.

5) The table below gives the estimated world population (in billions) for various years.

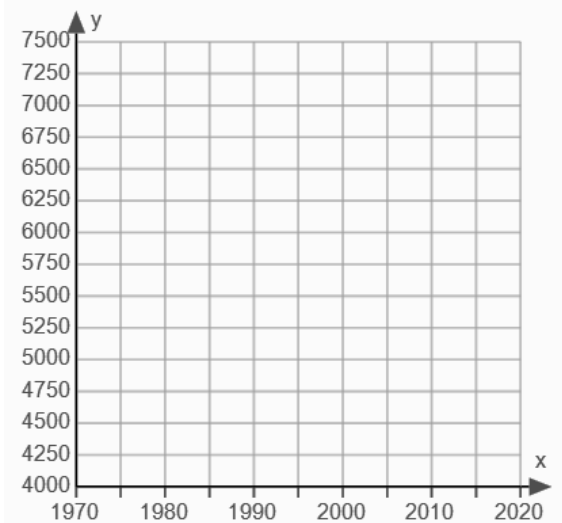
Year	1980	1990	1997	2000	2005	2011
Population	4400	5100	5852	6080	6450	7000

a) Linear Regression Equation:

b) Correlation Coefficient (r): _____

c) Type of Correlation: _____

d) Using the linear regression equation predict the world population in the year 2030.



Extra Practice

Decide whether each scenario has a positive, negative, or no correlation. Remember to think **generally!**

1. Hours studying vs. Grades

positive correlation *negative correlation* *no correlation*

2. Hours shopping vs. Amount of money spent

positive correlation *negative correlation* *no correlation*

3. Person's height vs. Person's age

positive correlation *negative correlation* *no correlation*

4. Person's height vs. How fast they drive

positive correlation *negative correlation* *no correlation*

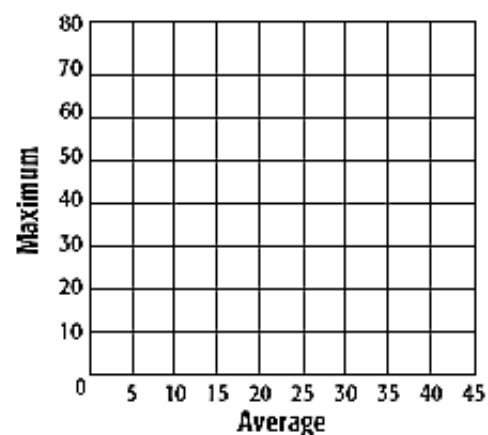
5. Temperature vs. Number of people wearing jackets

positive correlation *negative correlation* *no correlation*

2. The table shows the average and maximum longevity of various animals in captivity.

Longevity (years)								
Avg.	12	25	15	8	35	40	41	20
Max.	47	50	40	20	70	77	61	54

Animal Longevity (Years)



a) Draw a scatterplot and determine, what relationship, if any, exists in the data.

b) Calculate the linear regression equation.

c) Use your equation to predict the maximum life span of an animal that has an average life span of 17 years. *Round to the nearest tenth if needed.*

d) If an animal has a maximum life span of 30 years, what is the average life span of the animal? *Round to the nearest tenth if needed.*