Unit 6 Day 1

Learning Target: Using Measures of Center and Spread to Analyze Data

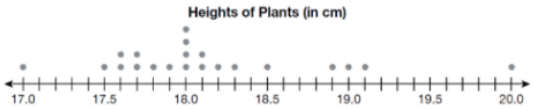
Unit 6 Day 1 Notes - Univariate Statistics: Shape, Center, and Spread

Shape

No matter what types of study you choose, it helps to organize your data in a data display. Here are some types of data displays:

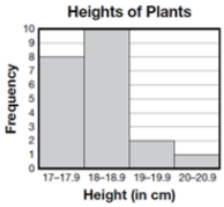
Dot Plot

- Used for numerical data that has relatively few points.
- Dots or x's can be used

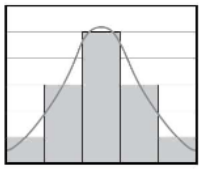


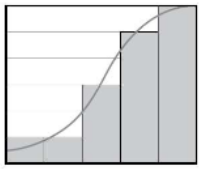
Histogram

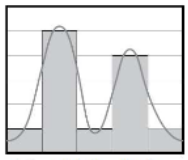
- Groups data points into ranges with **equal** intervals
- Intervals **do not overlap**

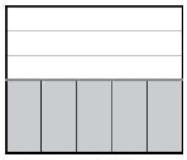


UNDERSTAND You can use the distribution of a data set, or its shape, to interpret it and to compare it to other data sets. Four kinds of distributions are described below.

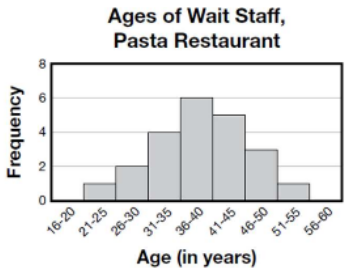

Normal Distribution


Skewed Distribution


Bimodal Distribution


Uniform Distribution

EXAMPLE B The two histograms below show the ages of wait staff at two restaurants.



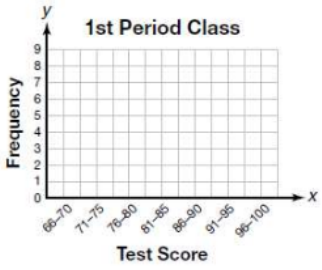
Identify the kind of distribution shown by each histogram. Use the shapes of the data sets to compare them.

For questions 3–5, use the given information. Create a histogram for each data set. Describe the distribution of each data set.

Students in the 1st and 2nd period biology classes took the same test. Their test scores are listed below.

3. 1st period test scores:

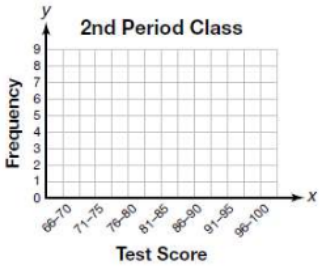
100, 91, 86, 73, 81, 100, 93, 94, 86, 86, 99, 93, 98, 84, 80, 97, 93, 87, 70, 97, 94, 88, 85, 96, 90



Distribution: _____

4. 2nd period test scores:

81, 87, 95, 85, 83, 82, 76, 68, 86, 83, 93, 87, 76, 87, 71, 100, 76, 91, 73, 80, 80, 84, 87, 88, 73



Distribution: _____

5. Compare and contrast the histograms for the biology classes in questions 4 and 5.

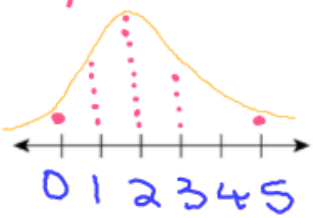
_____

_____

Create a dot plot for the given data. Describe the shape of the data.

6. Nathaniel opened 20 peanut shells and recorded the number of peanuts he found in each shell.

3, 2, 0, 1, 5, 2, 1, 2, 3, 1, 2, 2, 1, 2, 2, 3, 2, 3, 1, 2



This is a normal distribution

Fill in each blank with an appropriate word or phrase.

7. A dot plot shows data points as dots above a number line.

8. A histogram shows how frequently data occur within certain ranges or intervals.

9. Ranges or Intervals used in a histogram must be equal.

10. A Normal distribution is symmetric and resembles a bell curve.

11. A Bimodal distribution has two distinct peaks.

12. A Skewed distribution has a "tail" that extends more to one side of the graph than the other.

Measures of Center

Measures of Central Tendency are used to generalize data sets and identify common values.

Mean	**Definition:** Average of a numerical data set, $\bar{x}$ **Calculation:** Add up all the data values and divide by the number of data values.
Mode	**Definition:** Value that occurs most frequently. There can be no, one, or several modes.
Outlier	Data value that is much greater than or much less than the rest of the data in a data set If an outlier is present, you would use the median to describe the data, NOT the mean!

Example: Below are the scores that Justin earned on his last 8 homework assignments.

80, 95, 0, 90, 95, 80, 85, 90

1. What is his mean/ average homework score?
2. What is his median homework score?
3. Are there any outliers?

0, 80, 80, 85, 90, 90, 95, 95

$\frac{85+90}{2} = 87.5$

Find the mean and median for each data set.

3. 5, 25, 10, 15, 20

median: _____

mean: _____

4. 1, 7, 3, 2, 6

median: _____

mean: _____

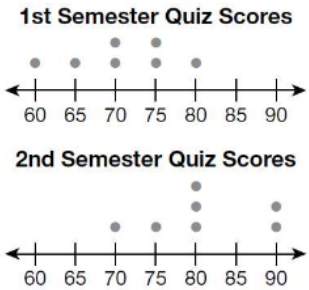
5. 10, 90, 10, 60, 40, 30

median: _____

mean: _____

Find the median for each data set or determine the interval in which the median must fall. Then compare the medians.

10. The dot plots show Kyla's Spanish quiz scores during the 1st and 2nd semesters.

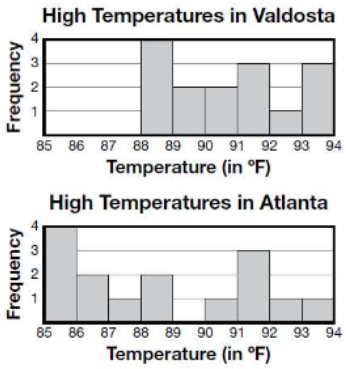


median score, 1st semester: _____

median score, 2nd semester: _____

Comparison: _____

11. The histograms show the daily high temperatures in two cities.



median high temperature, Valdosta: _____

median high temperature, Atlanta: _____

Calculator

1. Data: put in the numbers, press enter after each number

2. 2nd Data

3. Enter 4 times

1: n
2: $\bar{x}$ — mean
7: min
* Q1
9: med
A: Q3
B: max

Find the mean for each data set. Then compare the means.

12. The tables show the number of ads that were sold by the actors and stage-crew members working on a school play.

Actor	Rajiv	Amy	Penny	Leonard	Adriel
Ads Sold	4	4	5	6	7

Crew Member	Tina	Ben	Ronny	Irene	Cris
Ads Sold	6	7	8	9	9

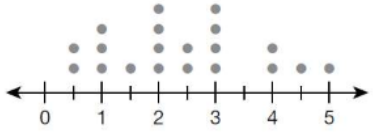
mean number sold, actors: _____

mean number sold, crew members: _____

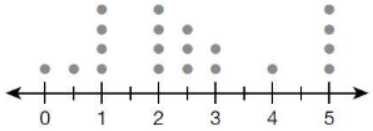
Comparison: _____

13. The dot plots show the number of hours of television watched yesterday by students from two homerooms.

Hours of Television Watched, Room 101



Hours of Television Watched, Room 102



mean number of hours, Room 101: _____

mean number of hours, Room 102: _____

Comparison: _____

14. Which statement accurately compares the average weight of a puppy from the 2nd litter to the average weight of a puppy from the 1st litter?

Weights of Puppies (in ounces)

1st Litter	$3\frac{1}{2}$, 4, 4, $4\frac{1}{2}$
2nd Litter	$4\frac{1}{2}$, 5, 7, $7\frac{1}{2}$

A. The average weight is about the same for both litters.

B. The average weight of a puppy from the 2nd litter is about $\frac{1}{2}$ as great.

C. The average weight of a puppy from the 2nd litter is about $1\frac{1}{2}$ times as great.

D. The average weight of a puppy from the 2nd litter is about $2\frac{1}{2}$ times as great.

15. To compare two shipments, five packages from each shipment were chosen at random and weighed. Which measure or measures of center would be best to use if you wanted to compare the weight of a typical package from each shipment?

Weights of Packages (in pounds)

1st Shipment	2, 4, 6, 8, 10
2nd Shipment	3, 3, 5, 8, 50

A. Median would be the best measure of center.

B. Mean would be the best measure of center.

C. Median and mean would both be equally good measures of center.

D. Neither the mean nor the median would be a good measure of center.

Measures of Spread

Measures of Spread describe the "diversity" of the values in a data set. Measures of spread are used to help explain whether data values are very similar or very different.

Range	• Range = Biggest # - Smallest #
Mean Absolute Deviation (MAD)	• Indicates how spread out or variable data are. • Measures how the data points in a set vary from the mean, $\bar{x}$ The formula for mean absolute deviation is: $\frac{\sum_{i=1}^{N} \left\| x_i - \bar{x} \right\|}{N}$ X_1 = data value $\bar{x}$ = mean $\sum$ = sum N = number of data values **Calculation:** - Find the mean of the set of numbers - Subtract each number in the set by the mean and take the absolute value of each new number (new number will be positive) - Find the sum of the new numbers and divide by the number of data values

Example: Calculate the MAD of this data set: 5, 8, 9, 11, 12.

Calculate the mean, $\bar{x}$.

$$\bar{x} = \frac{5+8+9+11+12}{5} = \frac{45}{5} = 9$$

Find the absolute deviation of each data point from the mean. Use a table to organize your work.

Data Point, x	Deviation from Mean, $x - \bar{x}$	Absolute Deviation from Mean, $\lvert x - \bar{x} \rvert$
5	$5 - 9 = -4$	$\lvert -4 \rvert = 4$
8	$8 - 9 = -1$	$\lvert -1 \rvert = 1$
9	$9 - 9 = 0$	$\lvert 0 \rvert = 0$
11	$11 - 9 = 2$	$\lvert 2 \rvert = 2$
12	$12 - 9 = 3$	$\lvert 3 \rvert = 3$

Calculate the mean of the absolute deviations.

$$\text{MAD} = \frac{4+1+0+2+3}{5} = \frac{10}{5} = 2$$

The mean absolute deviation is 2.

Use the data below for questions 1–4.

Heights (in inches) of Starting Players, Girls' Basketball Team	Heights (in inches) of Starting Players, Boys' Basketball Team
64, 66, 66, 68, 71	67, 67, 69, 70, 72

1. Calculate the mean and MAD of the heights of starting players for the girls' team. Use the table. Show your work.

$\bar{x}$ = _____

x	$x - \bar{x}$	$\lvert x - \bar{x} \rvert$
64		
66		
66		
68		
71		

MAD = _____

2. Calculate the mean and MAD of the heights of starting players for the boys' team. Use the table. Show your work.

$\bar{x}$ = _____

x	$x - \bar{x}$	$\lvert x - \bar{x} \rvert$
67		
67		
69		
70		
72		

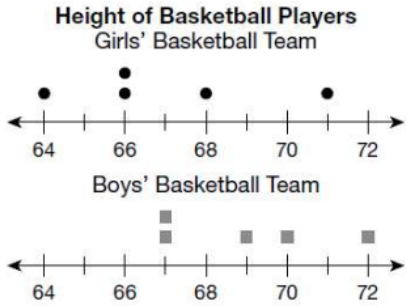
MAD = _____

3. On average, which team has taller starting players? Use the means you calculated above and the dot plots on the right.

_____

4. On which team are the heights of the starting players more variable? Use the MADs you calculated above and the dot plots on the right.

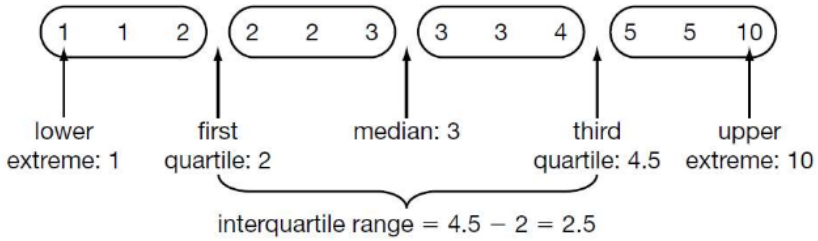
_____



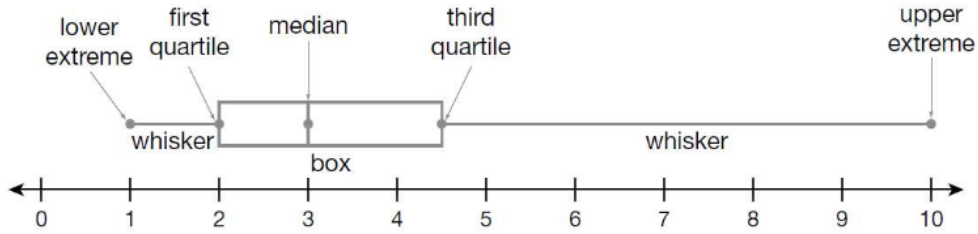
Day 2 - Box (and Whisker) Plots

UNDERSTAND Besides mean absolute deviation, another measure of spread is the **interquartile range (IQR)**. The interquartile range measures the variability of the middle 50% of the data, which is bounded by the **first quartile (Q_1)** and the **third quartile (Q_3)**. Recall that the median, *M*, divides a set of data into two halves. The first quartile is the median of the lower half of the data set. The third quartile is the median of the upper half of the data set.

The diagram below shows the **lower extreme, upper extreme,** quartiles, and median of a data set, as well as the interquartile range. It also helps illustrate how the median and quartiles divide a data set into four discrete sets of data.



A **box plot** (sometimes called a box-and-whisker plot) is an excellent way to display the extremes, quartiles, and median of a data set.



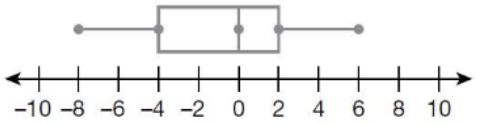
The box contains the middle 50% of the data, bounded by the first and third quartiles. The whiskers are on either end of the box. The left whisker contains the lower 25% of the data, and the right whisker contains the upper 25% of the data.

Outliers have much less effect on median and IQR than they do on mean and mean absolute deviation. Consider the example above and imagine that the upper extreme is 110, instead of 10. The first quartile, median, third quartile, and IQR would remain the same. The mean and mean absolute deviation, however, would be much greater.

In general, mean absolute deviation is a better measure of spread for data with a symmetric distribution and without outliers. For data with a skewed, or nonsymmetric, distribution and with outliers, IQR is usually a better measure of variability.

Practice

Use the box plot for questions 1–5.



1. What is the median? _____

HINT The median is the middle value in the set.

2. What is the lower extreme? _____

3. What is the upper extreme? _____

4. What is the first quartile? _____

5. What is the third quartile? _____

Find the median (M), first quartile (Q_1), and third quartile (Q_3) of the data.

6. 1, 2, 3, 5, 7, 9, 10

M = 5 min = 1

Q_1 = 2 max = 10

Q_3 = 9

REMEMBER The median divides the data set into two halves.

7. 10, 12, 12, 15, 17, 19, 21, 25

M = 16 $\frac{15+17}{2}$

Q_1 = 12 $\frac{19+21}{2}$

Q_3 = 20

min = 10 max = 25

8. −2, −1, 2, 3, 4, 6, 7, 7, 9

M = _____

Q_1 = _____

Q_3 = _____

9. 25, 35, 40, 45, 45, 50, 60, 65, 75, 95

M = _____

Q_1 = _____

Q_3 = _____

10. 15, 12, 18, 25, 36, 48, 28, 15

M = _____

Q_1 = _____

Q_3 = _____

11. 1.5, 2.5, 4.5, 8.5, 3.5, 0.5, 0.75, 2.25, 3.25

M = _____

Q_1 = _____

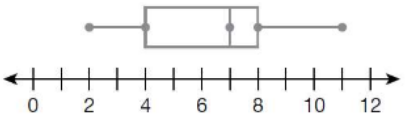
Q_3 = _____

IQR = $Q_3 - Q_1$

$IQR = Q_3 - Q_1$

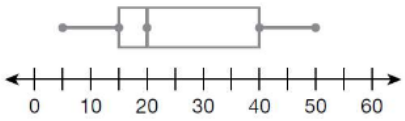
Calculate the interquartile range of the data.

16.



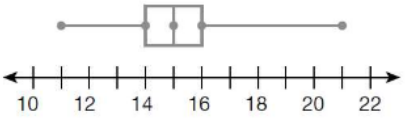
IQR = _____

17.



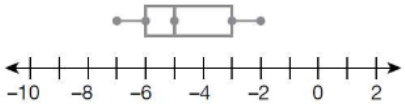
IQR = _____

18.



IQR = _____

19.

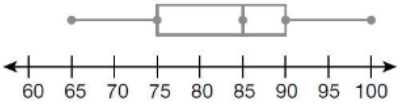


IQR = _____

Choose the best answer.

20. The box plot shows the test scores earned by students in a biology class. Which statement about the test scores is **not** true?

Biology Test Scores



A. The scores ranged from 65 to 100.

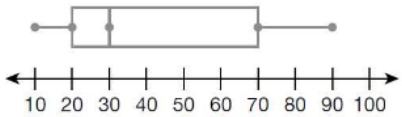
B. The median score earned was an 85.

C. 25% of students scored less than 75 points on the test.

D. 50% of students had scores that ranged from 75 to 85 points.

21. The box plot shows the prices of 20 skirts for sale at a boutique. Which statement about the prices is true?

Skirt Prices (in dollars)



A. The highest-priced skirt costs $100.

B. The median price of a skirt is $70.

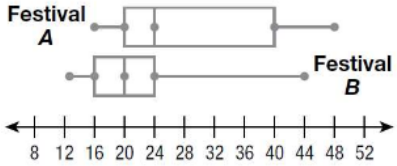
C. Half the skirts have prices that range from $20 to $70.

D. The prices of the skirts are close to the median and not very variable.

Use the box plots and information below for questions 22 and 23.

Music festival *A* and music festival *B* each had 100 volunteers. The box plots show the ages of the volunteers at each festival.

Ages of Music Festival Volunteers



22. Compare the median ages of volunteers at each festival.

_____

23. Which festival has more variability in the ages of its volunteers? Explain your answer.

_____

_____