Data Analysis & Statistics

Distributions

Measures of Center

Measures of Spread

Mean Absolute Deviation (MAD)

Line Plots

Histograms

Box-and-Whisker Plots

DISTRIBUTIONS

Distribution: the arrangement of values in a data set

Gap: The break or opening

Cluster: a group of things of persons close together

Peak: being at the point of maximum frequency, intensity, use, etc.

Data can be CATEGORICAL or NUMERICAL.

CATEGORICAL: when the categories for the data collected are words; examples-name of current book you are reading, favorite movie, etc. The students will help generate more examples

NUMERICAL (QUANTITATIVE): when the data collected is numbers; examples-how many books did you read last year, how many movies do you go to in a month, etc. Students should help generate others

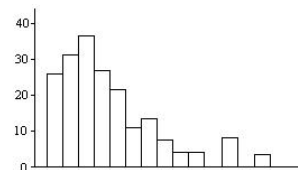**Example:** Identify each of the following as categorical or numerical (quantitative):

1) height of 6th graders:

2) favorite brand of cereal:

3) birth county for a person:

4) cost of dvd:

5) number of students in class who made A's on an exam:

6) how many books read in the past month:

7) gender of the next baby born in a hospital:

8) actual weight of chips in a bag:

Data can be described based on its shape:
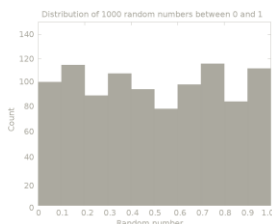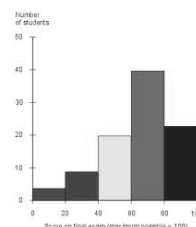
- MOUND SHAPE- bell shape curve (normal)
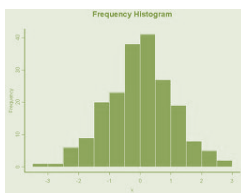
- SKEW RIGHT – extremes in high values

- UNIFORM – even distribution (flat)

- SKEW LEFT – extremes in low value

MEASURES OF CENTER

Mean when no outlier	*The sum of the values divided by the number of values.* ("average") Think of…my math teacher is really **mean** to make me do all this math.
Median When there is an outlier	*The middle value of numbers* **Numbers must be written in order from least to greatest. If there are two numbers in the middle, add the 2 numbers and divide by two. Think of…the median in the center of a highway.
Mode Categorical data	*The value that occurs most often* A data set can have no mode, one mode, or more than one mode. Think of…brownie a la mode (ice cream with brownies) the dessert you want the most!

Find the MMMR of the following data set:

Data set:	Mean	Median	Mode	Range
4, 7, 8, 2, 1, 2, 4, 2, 5, 6, 2				
28, 29, 25, 26, 29, 27, 28, 26, 29, 28				

MEASURES OF SPREAD

DESCRIBES HOW MUCH VALUES TYPICALLY VARY FROM THE CENTER

Range	*The difference (subtract) between the greatest value and lowest value.* Describes how spread out the data is. Think of… driving range in golf. How far does the ball travel from beginning to end??
IQR	*Inter-quartile range* Describes the middle 50% of the data; Q3- Q1 gives you IQR
MAD	*Mean Absolute Deviation* Describes the variability of the data from the mean (high MAD- data is more varied, low MAD- data is less varied)

Find the mean, median, and mode of the following data set with AND without the outlier:

88, 90, 55, 89, 90, 86, 87, 86, 87, 88, 89

What is the outlier? _____	Mean	Median	Mode
WITH THE OUTLIER			
WITHOUT THE OUTLIER			

MEAN ABSOLUTE DEVIATION
MAD

MEAN ABSOLUTE DEVIATION: The average distance of all data values from the mean of the set.

VARIABILITY: degree to which data are spread out around a center value.

TO GET THE MAD: 1) Find your mean

2) Find the difference between the mean and EACH data point (absolute value)

3) Find the mean of all of the differences

Find the MAD of the following:

1)   3, 2, 4, 2, 9, 8, 2, 11, 4

2)   85, 90, 68, 75, 79

2) The top five salaries and the bottom five salaries for the 2010 New York Yankees are shown in the table below. Salaries are in the millions of dollars and are rounded to the nearest hundredth. Find the mean absolute deviation for each set of data. Round to the nearest hundredth.

2010 New York Yankees Salaries (millions of $)	
Top Five Salaries	Bottom Five Salaries
33.00 24.29 22.60 20.63 16.50	0.45 0.44 0.43 0.41 0.41

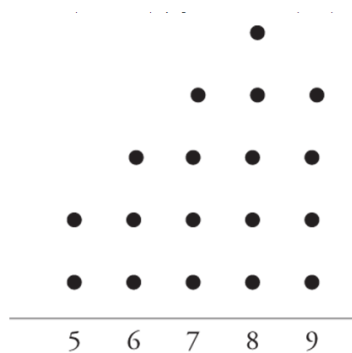MAD: MAD:

LINE PLOTS/DOT PLOTS

How to make:
1. Make a horizontal line.
2. Put the numbers for each category under the line.
3. Mark an x above each number listed.

The DOT PLOT displays scores on a 10-points quiz for 18 students.

1) What is the most common score?

2) What is the median score?

3) How many students scored a 9?

4) Which score has a frequency of 3?

5) What is the range of the scores?

6) How many students took the quiz?

The LINE PLOT displays students test scores on a recent science test.
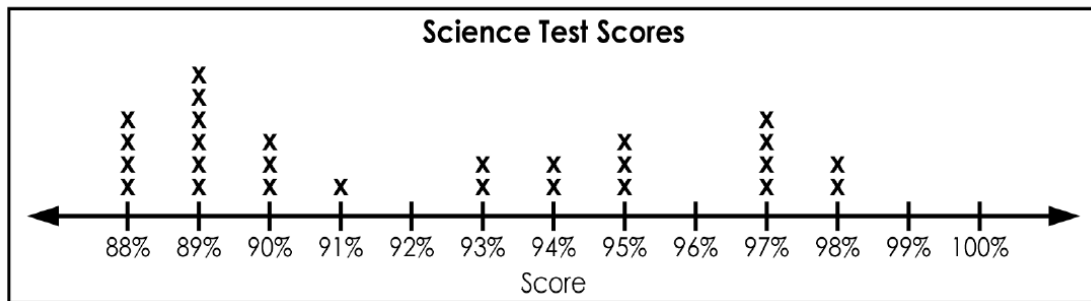
1) How many students received a score of 94%?

2) What was the highest score in the class?

3) What was the lowest score in the class?

4) How many students received a score in the 80s?

5) How many students received a score in the 90s?

6) How many students scored 93% or less?

7) How many students are in Mr. Bradley's science class?

FREQUENCY TABLE

How to make:

1. Make 3 columns. One for the category, one for the tally, one for the frequency.
2. Fill in the categories, make ranges if necessary.
3. Make tally marks. Make 4 marks, and then cross them off for the 5th.
4. The frequency is the number of tally marks you made.

Pairs of Shoes	Frequency
1–3	11
4–6	4
7–9	0
10–12	3
13–15	6

Answer the following using the frequency table:

1. How many pairs of shoes are represented within each interval?

2. How many total people were surveyed?

HISTOGRAMS

How to make:

1. Made from a frequency table.
2. Looks like a bar graph.
3. Bars touch and are same width.
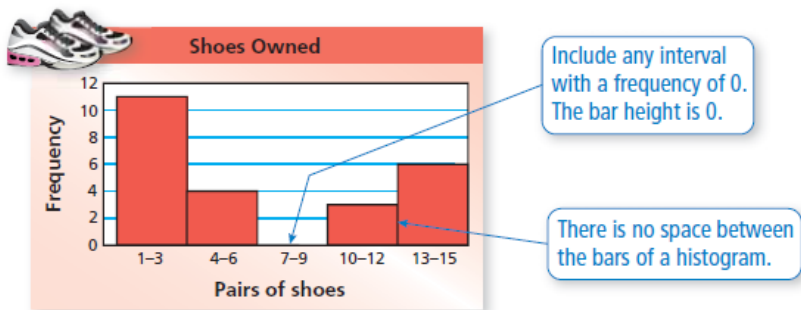
Use the histogram above to answer the following questions:

1. Describe the distribution of the histogram.

2. Where is the median?

3. What is the mode?

BOX-AND-WHISKER PLOTS

A Box-and-whisker plot consists of FIVE numbers:
- ❖ LOWER EXTREME: smallest number in data set.
- ❖ LOWER QUARTILE: median of lower half of data set.
- ❖ MEDIAN: median of the data set.
- ❖ UPPER QUARTILE: median of upper half of data set.
- ❖ UPPER EXTREME: largest number in data set.

BOX-AND-WHISKER PLOTS ARE BROKEN INTO 4 EQUAL PARTS, EACH REPRESENTING 25% OF THE DATA

THE INTER-QUARTILE RANGE is the range of the middle half of the data and contains 50% of the data set. To get this, subtract the lower quartile from the upper quartile.

Use the box-and-whisker plot to answer the questions.

Weekly Mileage Totals, 24 Runners

1.) Identify the 1st quartile :   Identify the 3rd quartile:

2.) What is the highest weekly total?   Lowest?

3.) What is the median weekly total?

4.) What percent of runners run less than 40 miles a week?

5.) How many runners run less than 20 miles a week?

6.) What percent of runners run less than 35 miles a week?

7.) What percent of runners run between 10 and 40 miles a week?

8.) How many runners run between 20-40 miles a week?

EOG Style Questions

1. What is the median of the following data set?   **9, 10.2, 10.4, 11**

    A) 10.3   B) 10.2
    C) 9   D) 10.25

2. Find the mean absolute deviation for the set below. **S = {65, 90, 85, 70, 70, 95, 55}**

    A) 12.24   B) 85.7
    C) 75.7   D) 40

3. The amounts of money that a local library collected for fines during the first eight months of last year are shown in the table below.

    Library Fines Collected

Month	Amount of Money	Month	Amount of Money
January	$45.00	May	$23.50
February	$21.50	June	$28.00
March	$48.00	July	$20.00
April	$40.50	August	$21.50

    What is the median of the amounts of money listed in the table?

    A) $18.00   B) $21.50
    C) $25.75   D) $31.00

4. The list below shows the number of miles Sophia hiked on each of 7 days.

    1.6, 3.1, 1.5, 2.0, 1.1, 1.8, 1.5

    What was the mean number of miles she hiked each day?

    A) 1.5 miles   B) 1.6 miles
    C) 1.8 miles   D) 2.0 miles

5. Which is NOT a measure of center?

    A) mean   B) median
    C) range   D) mode

6. Sarah's goal is to exercise an average of 45 minutes per day for one week. For the first 6 days of the week, she exercised 35, 40, 37, 42, 45, and 50 minutes. What is the number of minutes Sarah must exercise on the 7th day of the week to reach her goal exactly?
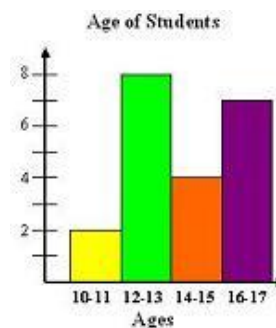
    A) 21 minutes   B) 42 minutes
    C) 49 minutes   D) 66 minutes

7. Use the frequency table to determine how many students received a score of 60 of better on an English exam.

Score	Frequency
50–59	1
60–69	8
70–79	3
80–89	7
90–100	4

    A) 8 students
    B) 9 students
    C) 14 students
    D) 22 students

8. From the histogram, determine the group where the median falls.
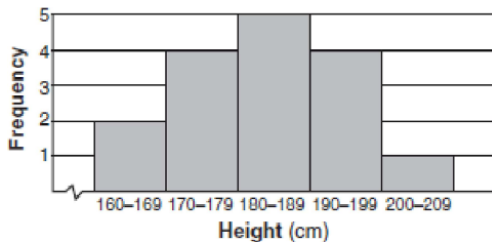

    A) 10-11   B) 12-13
    C) 14-15   D) 16-17

9. If the outlier in the set of data below was <u>dropped</u>, which measure of central tendency would change? **32, 44, 37, 53, 98, 52, 37, 50, 37, 44**

    A) Mean   B) Median
    C) Mode   D) all change

The following histogram shows the heights of students in Kayla's class. Use it for 10-13.

10. What is the total number of students in her class?

    A) 5   B) 15

    C) 16   D) 20

11. What is the median height of the students?

    A) 170-179   B) 180-189

    C) 190-199   D) 200-209

12. What was the most common height range?

    A) 160-169   B) 170-179

    C) 180-189   D) 200-209

13. What type of distribution is this histogram?

    A) skew right   B) skew left

    C) uniform   D) mound

14. What percentile is equivalent to the 1st quartile?

    A) 25th   B) 50th

    C) 75th   D) 100th

15. A group of students listed the number of children in their family. The range for the number of children was 5. What list of numbers could represent this data?
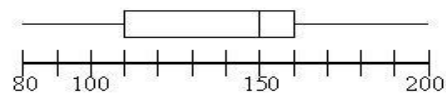
    A) 1, 1, 2, 3, 4   B) 1, 2, 3, 3, 4, 6

    C) 5, 5, 5   D) 2, 3

16. Which measure of central tendency would best represent the average weight of students with the following weights?
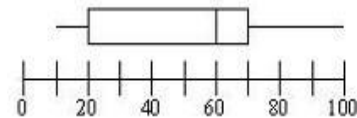
      120, 130, 120, 190, 135, 180

    A) Mean   B) Median

    C) Mode   D) Range

17. The box plot below represents the weights of the wrestlers on the varsity team. What percent of the wrestlers weigh between 150 and 160 pounds?

    A) 10%   B) 15%

    C) 25%   D) 50%

18. The box plot below represents a set of 100 numbers. How many scores are between 60 and 70?

    A) 10%   B) 15%

    C) 25%   D) 50%

19. The scores on Mr. Landrio's seventh grade math test are below. What is the inter-quartile range for this set of data?

    76, 90, 82, 85, 72, 96, 85, 91, 97, 80

    A) 11   B) 23

    C) 85   D) 86