Topic 14.1 Statistical Questions

<u>Statistical Question</u>: a question that anticipates there will be different answers in the data.

1. Mr. Borden wants to ask his class a question that will give several different answers. Which question should he ask?
    a. What is the area of one sheet of notebook paper?
    b. How many sheets of paper do you use each week?

The correct answer is b. Mr. Borden should ask, "How many sheets of paper do you use each week?" because this question has more than one right answer. He could display the answers to this question using a bar graph. The question, "What is the area of one sheet of notebook paper?" has only one correct answer.

Examples: Statistical vs. Not Statistical

1. How many siblings do you students in your class have? *Statistical*
2. In which month is your birthday? *Statistical*
3. How many states are in the United States? *Not Statistical*
4. How tall are the students in Grade 6? *Statistical*


Creating a Statistical Question

1. Numbers of pets classmates own: *How many pets do you own?*
2. Weights of pumpkins for sale: *How many pounds is each pumpkin?*
3. Heights of different plants growing: *How many inches tall is each plant?*
4. Visitors' ages at a park on a Saturday: *How old are you?*

Topic 14.2 Looking at Data Sets

Data Distribution: how data values are arranged by looking at its overall shape, its center, and the least and greatest values.

Outlier: an extreme value with few data points located near it.

Temperatures in May (°F)
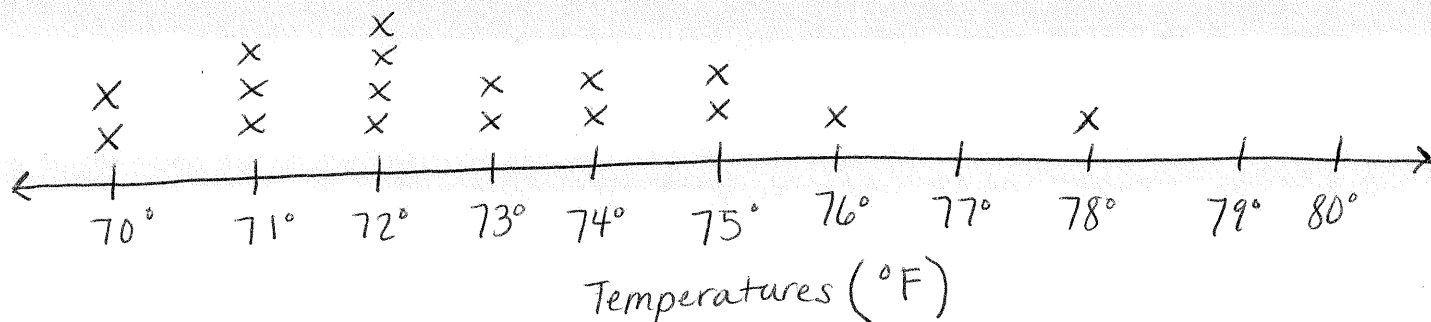
Temperatures (°F)

1. What is the lowest and highest temperature?

   70°F; 78°F

2. About where is the center of the data?

   Around 72°F and 73°F.

3. Are there any outliers? How do you know?

   There are no outliers because even though 78°F is separated from the rest of the data, it is not separated enough to be considered an outlier.

Mean or Average: the sum of all the values in a set of data divided by the total number of values in the set.

How to find the Mean/Average:

Step 1 – Add the values of the set of data to find the sum.

Step 2 – Divide the sum by the total number of values in the set.

Examples:

1. 5, 4, 9, 8

    $5 + 4 + 9 + 8 = 26$

    $26 \div 4 = \boxed{6.5}$

2. 101, 105, 103

    $101 + 105 + 103 = 309$

    $309 \div 3 = \boxed{103}$

3. 19, 55, 34, 16

    $19 + 55 + 34 + 16 = 124$

    $124 \div 4 = \boxed{31}$

4. 2, 5, 4, 5

    $2 + 5 + 4 + 5 = 16$

    $16 \div 4 = \boxed{4}$

5. 35, 45, 70

    $35 + 45 + 70 = 150$

    $150 \div 3 = \boxed{50}$

Topic 14.4 Median, Mode, and Range

<u>Median:</u> the <u>middle</u> data value. In order to find the median, list the data from least to greatest.

<u>Mode:</u> the data value that occurs <u>most</u> often in the data set.

<u>Range:</u> the *difference between the greatest and least values* in a data set.

Examples:

1. 5, 7, 5, 4, 6, 3, 5   3, 4, 5, 5, 5, 6, 7

Median: 5

Mode: 5

Range: 7 - 3 = 4

2. 21, 21, 23, 32, 43   21, 21, 23, 32, 43

Median: 23

Mode: 21

Range: 43 - 21 = 22

3. 13, 14, 14, 16, 17, 19   13, 14, 14, 16, 17, 19

Median: 14 + 16 = 30 ÷ 2 = 15

Mode: 14

Range: 19 - 13 = 6

Topic 14.5 Frequency Tables and Histograms

Frequency Table: shows the number of times a data value or range of values occurs in the data set.

Age Range	Tallies	Frequency
3 - 5	III	3
6 - 8	HHt I	6
9-11	IIII	4
12-14	III	3

Ages: 12, 8, 3, 5, 5, 10, 13, 11, 7, 6, 9, 6, 10, 12, 7, 6

Histogram: a graph that uses bars to show the frequency of equal ranges or groups of data.

Children Shoppers

Topic 14.6 Box Plots

Box Plots: shows a distribution of data values on a number line.

Quartiles: values that divide a data set into four equal parts.

1. Find the median, minimum, and maximum values.

(7,) 7, 9, 10, 10, 11, 13, (13,) 13, 14, 15, 15, 17, 18, (22)

↑ minimum   ↑ median   ↑ maximum

2. Find the median for each half.

7, 7, 9, (10) 10, 11, 13, (13) 13, 14, 15, (15) 17, 18, 22

↑ First quartile   ↑ Second quartile   ↑ Third quartile

3. Draw the box plot. Show a number line with an appropriate scale, a box between the first and third quartiles, and a verticle segment that shows the median. Draw segments that extend from the box to the minimum and maximum values.

6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

↑ minimum   ↑ 1st Quartile   ↑ Median 2nd Quartile   ↑ 3rd Quartile   ↑ maximum

Topic 14.7 Measures of Variability

Absolute Deviation: the distance between each data value and the mean.

Mean Absolute Deviation: the mean of the absolute deviations of a set of data. Add up all of the absolute deviations an divide by the total number of data values in the set.

Interquartile Range (IQR): the difference between the third quartile and the first quartile.

★ use the following data set to answer these questions

$$4, 5, 5, 6, 7, 8, 8, 10, 10$$

1. Absolute Deviation:

★ Find the Mean: $4+5+5+6+7+8+8+10+10 = 63 \div 9 = 7$

$7-4 = ③$ $7-7 = ⓪$

$7-5 = ②②$ $8-7 = ①①$

$7-6 = ①$ $10-7 = ③③$

2. Mean Absolute Deviation:

Take all Absolute Deviations and Divide by total.

$$3+2+2+1+0+1+1+3+3 = 16 \div 9 = 1.\overline{7}$$

3. Interquartile Range: $\left(1^{st}\ \text{Quartile} - 3^{rd}\ \text{Quartile}\right)$

Take the Average:

1. $5+5 = 10 \div 2 = 5$
2. $8+10 = 18 \div 2 = 9$

4   5   ⑤   5   6   7   8   8   ⑨ 10   10

1st Quartile   median 2nd Quartile   3rd Quartile
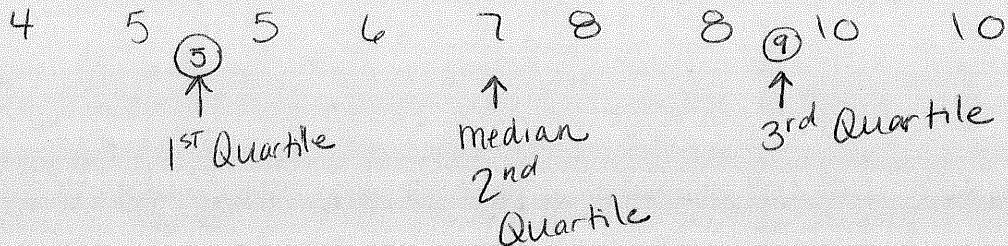
$$9 - 5 = \underline{\underline{4}}$$

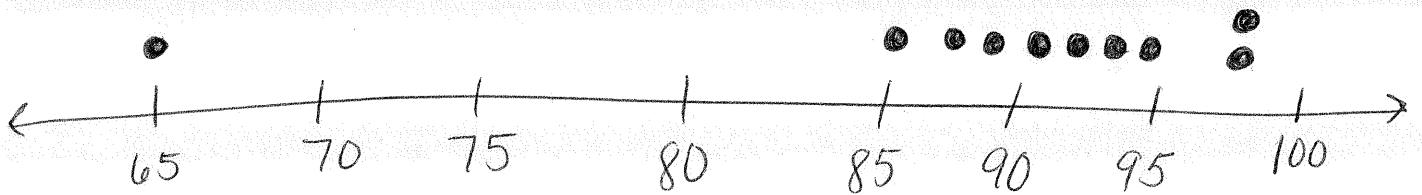## Topic 14.8 Approximate Use of Statistical Measures

Which measure of centers is most useful to describe a given situation?

Step 1: Organize your data using a dot plot and find the mean, median, mode, and outliers.

Step 2: Look for clusters, gaps, and outliers to help you reason what the measures of center tell you.

Step 3: Draw conclusions.

88, 87, 91, 65, 90, 85, 98, 98, 92, 93

Mean: $88 + 87 + 91 + 65 + 90 + 85 + 98 + 98 + 92 + 93 = 887$

$887 \div 10 = \underline{88.7}$

Median: 65, 85, 87, 88, 90, 91, 92, 93, 98, 98

$90 + 91 = 181 \div 2 = \underline{90.5}$

Mode: 98

Outlier: 65

★ The outlier 65 affects the mean and brings down the data

★ Most of the data is clustered around 90.5 which is the median.

★ 90.5 would be the most useful measure of center.

How can a data set be summarized?

✶ Find mean, median, mode.

✶ Draw a box plot.

✶ Find interquartile Range to describe variability.

✶ Make generalizations about data.

✶ Find the "center" of the data.
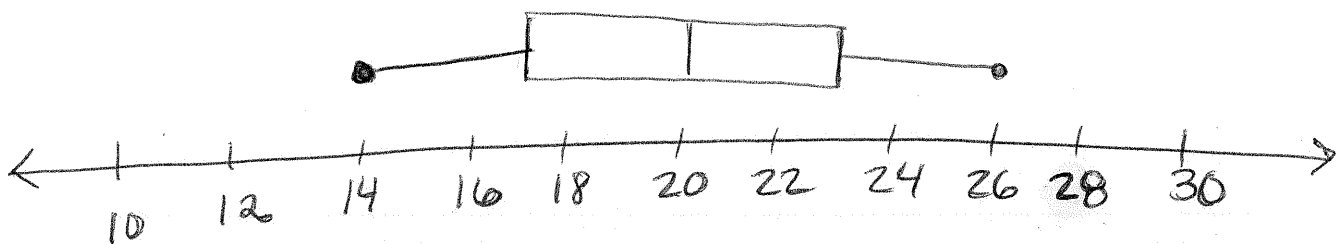
① What is the median?

→ <u>20</u>

② How would you describe the shape of the data?

→ <u>symmetric</u>

③ What are the greatest and least values?

→ <u>26</u> and <u>14</u>

④ What are the 1st and 3rd Quartiles?

→ <u>17</u> and <u>23</u>

⑤ What is the Interquartile Range?

→ 23-17 = <u>6</u>

Topic 14.10 Try, Check, and Revise

A store sells 5 different kinds of milk in gallon jugs. The mean price of a gallon of milk is $3.29. No two prices are the same. List 5 possible prices for the gallons of milk. Make sure that each price is a reasonable for a gallon of milk.

Use reasoning. Find 5 possible prices having a mean of $3.29.

Try these possible prices: $2.89, $2.99, $3.09, $3.29, and $3.49.

Check to see if the mean is equal to $3.29.

2.89 + 2.99 + 3.09 + 3.29 + 3.49 = 15.75

15.75/5 = $3.15

The mean is too low, so you need to use higher prices.

Revise the prices using what you know. The mean was 14 cents too low.

Try adding 14 cents to each price: $3.03, $3.13, $3.23, $3.43, and $3.63.

Check to see if the mean equals $3.29.

3.03 + 3.13 + 3.23 + 3.43 + 3. 63 = 16.45

16.45/5 = $3.29

5 possible prices for the milk are $3.03, $3.13, $3.23, $3.43, and $3.63.