MATH SPEAK - TO BE UNDERSTOOD AND MEMORIZED

1) **STATISTICS** = the branch of mathematics used to analyze and interpret ***data***.
2) **DATA** = information, often in the form of numbers, that are gathered, organized, analyzed and interpreted.
3) **DATA SET** = a group of numbers to be statistically analyzed.
4) **POPULATION** = all the members of a group to be statistically analyzed.
5) **POPULATION SIZE** = the number of pieces of data in a ***data set*** and is represented by ***n***.
6) **SAMPLE** = some of the members of a group to be statistically analyzed.
7) **SAMPLE SIZE** = the number of pieces of data in a ***data set*** and is represented by ***n***.
8) **RANGE** = ***range*** $=$ largest number $-$ smallest number . ***Range*** gives a rough indication of the ***dispersion*** of the numbers within a ***data set***.
9) **LINE PLOT** = a number line having dots the representing each of the numbers in a ***data set*** above the corresponding number on the number line.
10) **MODE** = the number in a ***data set*** that appears most often.
11) **MEDIAN** = the number that divides the ***data set*** into two equal upper and lower halves.
12) **MEAN** = the average of all the numbers in a ***data set***. ***Mean*** is represented by these symbols: $\bar{x}$ = ***mean*** for a sample, μ = ***mean*** for an entire population.
13) **OUTLIER** = a ***number*** that is judged to be very far from the other ***data*** in the ***data set***. **i.e.** in the ***data set*** below, 35 is the outlier because it is 20 units away from its closest number.

    e.g. $x = {}^{-}7,\ {}^{-}3,\ 0,\ 1,\ 2,\ 5,\ 9,\ 11,\ 15,\ 35$

INTRODUCING STATISTICS

I) ***Statistics*** is the gathering, organization, analysis and interpretations of ***data***. ***Data*** is the information, often recorded as numbers, that are collected, organized, analyzed, and interpreted. Analysis of ***data*** begins with these four ***statistics***: **RANGE**, **MODE**, **MEDIAN**, and **MEAN**.

A) ***Data*** is often recorded as numbers. Each individual piece of ***data***, each number, collected is represented by the letter $x_{\#}$. The first piece of ***data*** is represented by x_1, the second piece of ***data*** is represented by x_2, the third piece of ***data*** is represented by x_3, etc. ***The number of pieces of data is called the POPULATION SIZE or the SAMPLE SIZE and represented by the letter n***: If there are 5 pieces of ***data*** in a set of ***data***, ***n*** = 5. Study this ***data set***, and the organization that follows.

Data set 1: $x = {}^{-}7,\ {}^{-}3,\ 0,\ 1,\ 2,\ 5,\ 9,\ 11,\ 15$

$$x_1 = {}^{-}7\ ,\ \ x_2 = {}^{-}3\ ,\ \ x_3 = 0\ ,\ \ x_4 = 1\ ,\ \ x_5 = 2\ ,\ \ x_6 = 5\ ,\ \ x_7 = 9\ ,\ \ x_8 = 11\ ,\ \ x_9 = 15\ ;\ \text{ and } n = 9$$

1) Clarifying a ***data set***.
   a) Consider a class of 25 students wrote a Unit Test. The teacher recorded each students resulting percent in a table (see the left hand table found on page 212 of the text). Since the ***data set*** found in the table includes all the students in the class, the ***data set*** represents the entire ***population*** of the class. This means ***n*** represents the number of members of a ***population***: **i.e.** the ***population size*** is n = 25.
   b) Rob's Analytics is a company that conducts surveys of people. The company was asked to survey Canadians about how many hours they spend watching CBC television each week. Since it is not possible to survey every individual Canadian, the company chooses to randomly survey 20 000 people spread across the country. The ***data*** they collected does not include all members of the Canadian ***population***, rather it includes a 20 000 person ***sample*** of the entire ***population*** of Canada. In this situation ***n*** is referred to as the ***sample size***: **i.e.** the ***sample size*** is ***n*** = 25.

B) ***DISPERSION is a measure of the spread of the numbers in a data set***. The ***dispersion*** is analyzed using two analytical tools: ***RANGE*** and a ***LINE PLOT***.

1) ***RANGE is the statistical term that represents one measure of the spread of the numbers in a data set***. The ***range*** gives a rough indication of the ***dispersion***, or numerical spread, of the ***data set***: A small ***range*** indicates the ***data*** is relatively concentrated as the ***data*** falls within a small group of numbers; A large ***range*** indicates the ***data*** exists within a large group of numbers. There are limits to the information a large ***range*** will give. If the ***range*** is large, we know the ***data*** falls within two widely separated numbers, but we don't know if the ***data*** are spread evenly throughout the entire ***range*** or are found mostly closer to the larger or smaller number or exists in several bunched groups between the largest and smallest numbers.

a) The ***range*** of a ***data set*** is calculated using this formula: ***range*** = largest number − smallest number

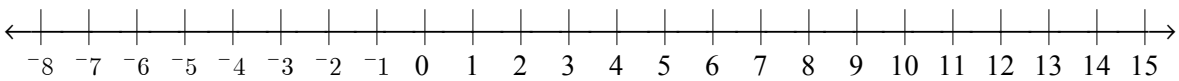
b) ***SAMPLE PROBLEM 1:*** Calculate the ***range*** of ***data set 1*** given above.

range $= 15 - {}^{-}7 = 15 + 7 = 22$ Thus, all the numbers within the ***data*** set fall within a ***range*** of 22.

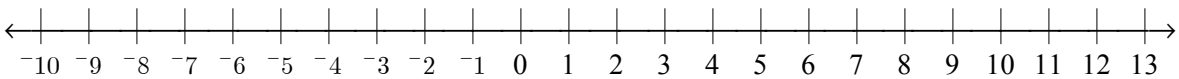
2) The ***dispersion*** of the ***data*** can be visualized by drawing a type of graph called a ***LINE PLOT*** (sometimes called a ***DATA PLOT***). ***A LINE PLOT is a graph composed of a single horizontal number line with one or more dots representing the numbers in a data set placed above its numbers.*** Each dot represents one of the numbers found within the ***data set***.

a) ***SAMPLE PROBLEMS 2:*** Draw a ***line plot*** for each of these ***data sets***.

1. ***Data set 1***: $x = {}^{-}7,\ {}^{-}3,\ 0,\ 1,\ 2,\ 5,\ 9,\ 11,\ 15$



2. ***Data set 2***: $x = {}^{-}10,\ 2,\ 2,\ 3,\ 3,\ 3,\ 7,\ 11,\ 13$



b) ***REQUIRED PRACTICE 2a:*** Remove page 6 & 7 of this notes package. Complete questions 1, 2 & 3 for each person. **{Answers are on page 5 of these notes.}**

c) ***REQUIRED PRACTICE 2b:*** Determine the ***range*** and create a ***line plots*** for each brand of battery using the ***data*** listed in the tables found on page 210 of the text. **{Answers are on page 4 of these notes.}**

C) ***MODE is the number that appears most often in the data set.***

1) ***SAMPLE PROBLEMS 3:*** Determine the ***mode*** of these ***data sets***.

b) ***Data set 3***: $x = 2,\ 3,\ 3,\ 7,\ 11,\ 13$

The number _____ appears _____ while all other numbers appear only once, thus the ***mode*** = ___

a) ***Data set 1***: $x = {}^{-}7,\ {}^{-}3,\ 0,\ 1,\ 2,\ 5,\ 9,\ 11,\ 15$

Each number in this ***data set*** is different, thus it does not have a ***mode***: ***mode*** = _____

2) ***REQUIRED PRACTICE 3:*** Complete question 4 for each person on page 6 & 7. **{Answers are on page 5 of these notes.}**

D) ***MEDIAN is the number that divides the data set into two equal upper and lower halves***; half of the numbers in the ***data set*** are less than the ***median*** while half of the numbers in the ***data set*** are larger than the ***median***. The ***median*** can be a number that is found within the ***data set*** or may be a calculated number that is not a found within the ***data set***.

1) **USE THESE STEPS TO DETERMINE THE *MEDIAN* OF A DATA SET**
   1) List the ***data set*** from smallest to largest value.
   2) If the ***data set*** has an odd number of values, the ***median*** is the middle number.
   3) If the ***data set*** has an even number of values, the ***median*** is the average of the two middle numbers.

2) ***SAMPLE PROBLEMS 4:*** Determine the ***median*** of these ***data sets***.
   a) ***Data set 1***: $x = {}^{-}7,\ {}^{-}3,\ 0,\ 1,\ 2,\ 5,\ 9,\ 11,\ 15$

      This ***data set*** has an _____ number of values, thus, the ***median*** is the _____ number: ***median*** = __

      This means that 2 divides the ***data set*** into two equal sized halves: half of the numbers of the ***data set*** are less than 2 while half the numbers of the ***data set*** are larger than 2.

   b) ***Data set 3***: $x = 2,\ 3,\ 5,\ 7,\ 11,\ 13$

      This ***data set*** has an _____ number of values, thus, the ***median*** is the average of the _____ numbers: ***median*** =

      This means that 6 divides the ***data set*** into two equal sized halves: half of the numbers of the ***data set*** are less than 6 while half the numbers of the ***data set*** are larger than 6.

   c) ***Data set 4***: $x = 22,\ 23,\ 25,\ 25,\ 31,\ 33$

      This ***data set*** has an _____ number of values, thus, the ***median*** is the average of the _____ numbers: ***median*** =

      This means that 25 divides the ***data set*** into two equal sized halves: half of the numbers of the ***data set*** are less than 25 while half the numbers of the ***data set*** are larger than 25.

3) ***REQUIRED PRACTICE 4:*** Complete question 5 for each person on page 6 & 7. **{Answers are on page 5 of these notes.}**

E) ***MEAN is the proper statistical name for the AVERAGE of the numbers found within a data set.*** It is symbolized as $\bar{X}$ (which is read ***x bar***) or μ (which is read ***mew*** as in the sound kittens make). Each version of the ***mean*** is calculated using formulae given below.

$\mu = \frac{\sum x}{n}$ where

$\bar{X} = \frac{\sum x}{n}$ where

1) ***SAMPLE PROBLEM 5:*** Calculate the ***mean*** of ***data set 1*** given above.

   The ***mean***, commonly called ***average***, of the above ***data set*** is:

2) ***REQUIRED PRACTICE 5:*** Complete question 6 for each person on page 6 & 7. **{Answers are on page 5 of these notes.}**

F) ***An OUTLIER is a number located at one of the ends of a data set that is organized from smallest to largest that is very different than the other numbers in the data set in that it is numerically clearly much more distant from the other numbers in the data set.*** A ***data set*** can have no ***outliers***, or one or more ***outliers***.
1) ***SAMPLE PROBLEMS 6:*** Determine the ***outlier(s)*** of these ***data sets***.
   a) ***Data set 5***: $x = {}^{-}7,\ {}^{-}3,\ 0,\ 1,\ 2,\ 5,\ 9,\ 11,\ 15,\ 35$

      The ***outlier*** of this ***data set*** is _____, because it is a number that is located at the right end of the ***data set***, it is extremely large number relative to all other in the ***data set*** and the distance between it

and the nearest number, 20 units, is very large relative to the distances of 1, 2, 3 and 4 units found between other numbers in the ***data set*** providing strong evidence to suggest that 35 is an ***outlier***.

b) ***Data set 6***: $x = {}^{-}12,\ 2,\ 3,\ 3,\ 7,\ 11,\ 23$

The ***outliers*** of this ***data set*** are _____. ${}^{-}12$ is an ***outlier*** because it is a number that is located at the left end of the ***data set***, is very small relative to all the other numbers within the ***data set***, and is 14 units away from its closest number. 23 is an ***outlier*** because it is a number that is located at the right end of the ***data set***, is very large relative to all the other numbers within the ***data set***, and is 12 units away from its closest number. **NOTICE** that all other numbers in the ***data set*** are either 1 or 4 units away from each other. A separation between numbers within this ***data set*** of 12 or 14 units is very large relative to the 1 or 4 units separation of the other numbers within this ***data set*** providing strong evidence to suggest that ${}^{-}12$ and 23 are ***outliers***.

2) ***REQUIRED PRACTICE 6:*** Complete question 7 for each person on page 6 & 7. **{Answers are on page 5 of these notes.}**

G) ***REQUIRED PRACTICE 7:*** For the ***data*** listed in questions **1** & **2** below, complete questions **a)** through **j)** listed below. **{Answers are on page 4, 5 & 6 of these notes.}**

1) Heights of bean plants (cm) 25, 22, 23, 24, 30, 21, 23, 22, 23, 26

2) Masses of soft balls (g) 153.8, 154.2, 155.3, 154.8, 154.5, 154.5, 152.0

a) Arrange the data from smallest to largest. **b)** State the ***sample size***. **c)** Calculate the ***range***.

d) Create a ***line plot***. **e)** Determine the ***mode***. **f)** Determine the ***median***.

g) Calculate the ***mean***. **h)** Identify any ***outliers***

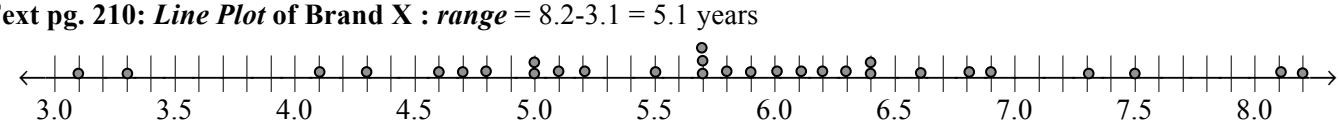
i) Re-calculate the ***sample size***, ***range***, ***mode***, ***median*** and ***mean*** without the ***outlier***.

j) Compare the ***sample size***, ***range***, ***mode***, ***median*** and ***mean*** you calculated with and without the ***outlier***. Describe how the ***outlier*** impacts each statistic.

ANSWERS TO THE REQUIRED PRACTICE

Required Practice 2b from page 2

Text pg. 210: *Line Plot* of Brand X : *range* = 8.2-3.1 = 5.1 years

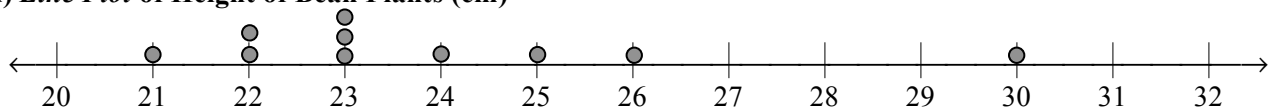


Text pg. 210: *Line Plot* of Brand Y : *range* = 8.2-3.1 = 5.1 years

Required Practice 7 from page 4

1a) $x = 21,\ 22,\ 22,\ 23,\ 23,\ 23,\ 24,\ 25,\ 26,\ 30$ **1b)** $n = 10$ **1c)** ***Range*** = 9 cm

1d) *Line Plot* of Height of Bean Plants (cm)



1e) ***mode*** = 23 cm **1f)** ***median*** = 23 cm **1g)** $\bar{x} = 23.9$ cm **1h)** ***outlier*** = 30 cm

1i) $n = 9$, ***range*** $= 5$, ***mode*** = 23 cm, ***median*** = 23 cm, $\bar{x} = 20.9$ cm

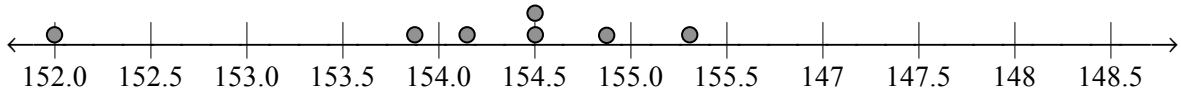
1j)

Stat	With *outlier*	Without *outlier*	Comparison
n =	10	9	The ***sample*** size has decreased significantly by 1.
Range	9 cm	4 cm	The ***range*** has decreased significantly by 5 cm.

Mode =	23 cm	23 cm	The ***mode*** has not changed.
Median =	23 cm	23 cm	The ***median*** has not changed.
Mean: $\bar{x}$ =	23.9 cm	20.9 cm	The ***mean*** has decreased by 3 cm.

2a) $x = 152.0,\ 153.8,\ 154.2,\ 154.5,\ 154.5,\ 154.8,\ 155.3$ **1b)** n = 7 **1c)** *Range* = 3.3 g

2d) *Line Plot* of Height of Soft Balls (g)



2e) *mode* = 154.5 g **2f) *median*** = 154.5 g **2g)** $\bar{x} = 154.16$ g **2h) *outlier*** = 152.0 g

2i) n = 6, ***range*** $= 2.1$ g , ***mode*** = 154.5 g, ***median*** = 154.5 g, $\bar{x} = 154.52$ g

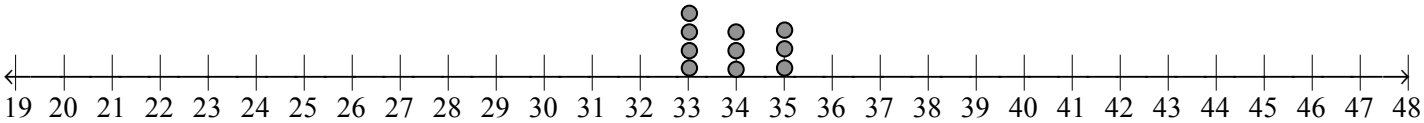
2j)

Stat	**With *outlier***	**Without *outlier***	**Comparison**
n =	7	6	The ***sample size*** has decreased significantly by 1.
Range	3.3 g	2.1 g	The ***range*** has decreased significantly by 1.8 g.
Mode =	154.5 g	154.5 g	The ***mode*** has not changed.
Median =	154.5 g	154.5 g	The ***median*** has increased by 0.3 g.
Mean: $\bar{x}$ =	154.15 g	154.51 g	The ***mean*** has increased by 0.36 g.

Required Practice 2a, 3, 4, 5, 6 & 8 from pages 2, 3, 4 & 5

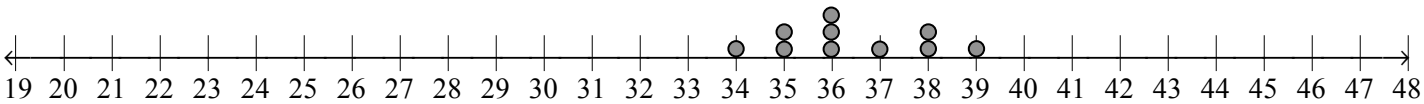
PAIGE:

1. *n* = 10 **2. *range*** = 2% **3.** *The* ***line plot*** *is below.* **4. *mode*** = 33% **5. *median*** = 34%
6. *mean* = $\bar{X}$ = 33.9% **7. *outlier(s)*** = none **8. *standard deviation*** = σ = 0.83%



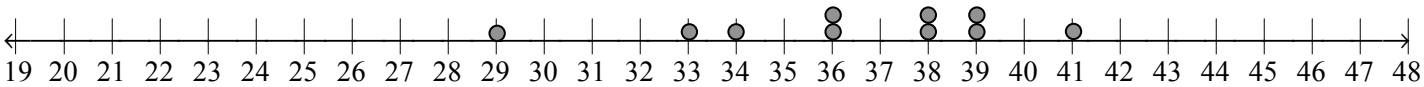
PATRICE:

1. *n* = 10 **2. *range*** = 5% **3.** *The* ***line plot*** *is below.* **4. *mode*** = 36% **5. *median*** = 36%
6. *mean* = $\bar{X}$ = 36% **7. *outlier(s)*** = none **8. *standard deviation*** = σ = 1.50%



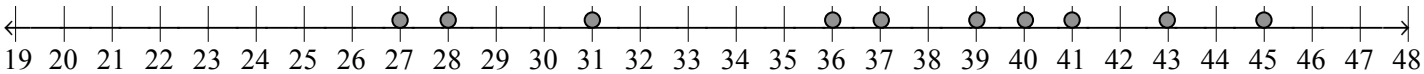
STAR:

1. *n* = 10 **2. *range*** = 12% **3.** *The* ***line plot*** *is below.* **4. *mode*** = 36%, 38% or 39% **5. *median*** = 37%
6. *mean* = $\bar{X}$ = 36.3% **7. *outlier(s)*** = 29% **8. *standard deviation*** = σ = 3.35%



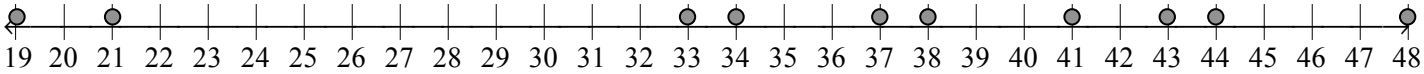
ANNA:

1. *n* = 10 **2. range** = 18% **3.** *The* ***line plot*** *is below.* **4. *mode*** = none **5. *median*** = 38%
6. *mean* = $\bar{X}$ = 36.7% **7. *outlier(s)*** = none **8. *standard deviation*** = σ = 5.88%



MORGAN:

1. *n* = 10 **2. *range*** = 29% **3.** *The* ***line plot*** *is below.* **4. *mode*** = none **5. *median*** = 37.5%
6. *mean* = $\bar{X}$ = 35.8% **7. *outlier(s)*** = 19% & 21% **8. *standard deviation*** = σ = 9.02%

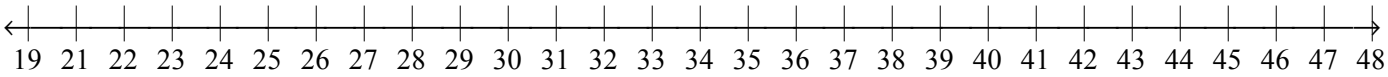


Use the data found in the ***INVESTIGATE the Math*** page 226 to complete the questions below.

1. State the ***sample size***.
2. State the ***range***.
3. Complete the ***line plot***.
4. State the ***mode***.
5. State the ***median***.
6. Complete the ***mean***.
7. State any ***outlier(s)***.
8. State the ***standard deviation***.

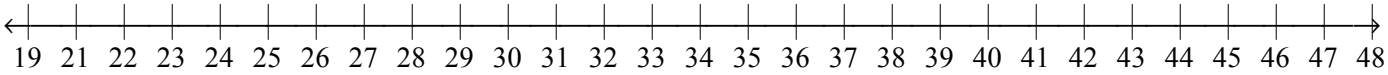
PAIGE:

1. ***n*** = _____
2. ***range*** = _____
3. Complete the ***line plot*** below.
4. ***mode*** = _____
5. ***median*** = _____
6. ***mean*** = $\bar{X}$ = _____
7. ***outlier(s)*** = _____
8. ***standard deviation*** = σ = _____



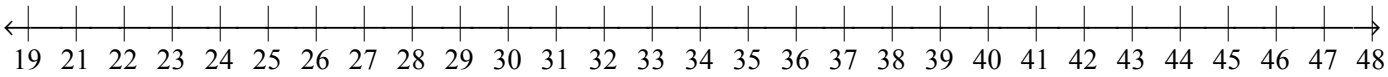
PATRICE:

1. ***n*** = _____
2. ***range*** = _____
3. Complete the ***line plot*** below.
4. ***mode*** = _____
5. ***median*** = _____
6. ***mean*** = $\bar{X}$ = _____
7. ***outlier(s)*** = _____
8. ***standard deviation*** = σ = _____



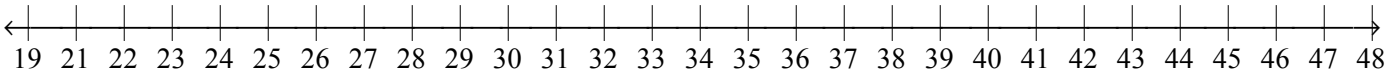
STAR:

1. ***n*** = _____
2. ***range*** = _____
3. Complete the ***line plot*** below.
4. ***mode*** = _____
5. ***median*** = _____
6. ***mean*** = $\bar{X}$ = _____
7. ***outlier(s)*** = _____
8. ***standard deviation*** = σ = _____



ANNA:

1. ***n*** = _____
2. ***range*** = _____
3. Complete the ***line plot*** below.
4. ***mode*** = _____
5. ***median*** = _____
6. ***mean*** = $\bar{X}$ = _____
7. ***outlier(s)*** = _____
8. ***standard deviation*** = σ = _____



MORGAN:

1. ***n*** = _____
2. ***range*** = _____
3. Complete the ***line plot*** below.
4. ***mode*** = _____
5. ***median*** = _____
6. ***mean*** = $\bar{x}$ = _____
7. ***outlier(s)*** = _____
8. ***standard deviation*** = σ = _____