

Describing Data Variations (Spread)

Name: _____

A company is shooting a commercial and asked two different modeling agencies to send them a group of 5 models with an average age 17.



The company “**Modeling Marvels Agency**” sent 5 models with the following ages: **16, 16, 15, 17, and 21.**

The company “**Acting Up Models Inc.**” sent 5 models with the following ages: **3, 5, 6, 31, and 40.**

1. Did each company correctly send a group of 5 models with an average age of 17?

MMA

$$\text{MEAN}(\bar{x}) = \frac{16+16+15+17+21}{5} = \frac{85}{5} = 17$$

AUM

$$\text{MEAN}(\bar{x}) = \frac{(3+5+6+31+40)}{5} = \frac{85}{5} = 17$$

2. What would you describe as being the most different between the two groups and how might you quantify this?

RANGE = HIGHEST DATA VALUE - LOWEST DATA VALUE

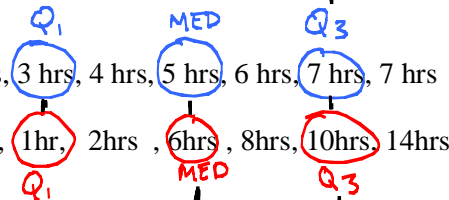
$$\text{MMA (RANGE)} = \overset{\text{HIGH}}{21} - \overset{\text{LOW}}{15} = 6$$

$$\text{AUM (RANGE)} = \overset{\text{HIGH}}{40} - \overset{\text{LOW}}{3} = 37$$

Two competing companies design similar android phones. A magazine is writing a review on the two companies and sampled 7 phones from each company to determine their battery life to the nearest hour while watching streaming videos:

Samsung's Universe 4 Android Phone Battery life sample: 3 hrs, 3 hrs, 4 hrs, 5 hrs, 6 hrs, 7 hrs, 7 hrs

Motovola's Void 5 Phone Battery life sample: 1hr, 2hrs, 6hrs, 8hrs, 10hrs, 14hrs



3. What is the **mean** of each sample of phones?

$$\text{UNIV}(\bar{x}) = \frac{3+3+4+5+6+7+7}{7} = \frac{35}{7} = 5 \text{ hrs}$$

$$\text{VOID}(\bar{x}) = \frac{1+1+2+6+8+10+14}{7} = \frac{42}{7} = 6 \text{ hrs}$$

4. What is the **median** of each sample of phones?

$$\text{UNIV (MEDIAN)} = 5 \text{ hrs}$$

$$\text{VOID (MEDIAN)} = 6 \text{ hrs}$$

5. What is the **range** of each sample of phones?

$$\text{UNIV (RANGE)} = 7 - 3 = 4 \text{ hrs}$$

$$\text{VOID (RANGE)} = 14 - 1 = 13 \text{ hrs}$$

6. What is the **lower quartile (Q1)** of each sample of phones?

$$\text{UNIV (Q}_1) = 3 \text{ hrs}$$

$$\text{VOID (Q}_1) = 1$$

7. What is the **upper quartile (Q3)** of each sample of phones?

$$\text{UNIV (Q}_3) = 7 \text{ hrs}$$

$$\text{VOID (Q}_3) = 10$$

8. What is the **inner quartile range (IQR)** of each sample of phones?

$$\text{UNIV (IQR)} = Q_3 - Q_1 = 7 - 3 = 4 \text{ hrs.}$$

$$\text{VOID (IQR)} = 10 - 1 = 9 \text{ hrs}$$

80, 81, 82, 90, 90, 98, 99, 100
MED

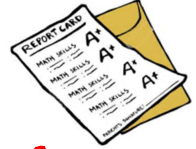
89, 89, 90, 90, 90, 91, 91
MED 90.5

A math teacher must make a recommendation for a \$2000 scholarship to a local chamber of commerce. The teacher has two students in mind **Alan** and **Brianna**. The teacher decides to let their grades be the determining factor. Here are their test scores for the semester:



Alan: 90, 90, 80, 100, 99, 81, 98, 82
~~80, 81, 82, 90, 90, 98, 99, 100~~

Brianna: 90, 90, 91, 89, 91, 89, 90, 90
~~89, 89, 90, 90, 90, 91, 91~~



1. Which student has the higher arithmetic **mean**, (average (\bar{x}))?

ALAN (\bar{x}): $\frac{90+90+80+100+99+81+98+82}{8} = 90$

BR1 (\bar{x}): $\frac{90+90+91+89+91+89+90+90}{8} = 90$

2. Which student has the higher **median**?

ALAN (MED): $\frac{90+90}{2} = 90$

BR1 (MED): $\frac{90+90}{2} = 90$

3. What might be the problem of using these measures of central tendency?

BOTH MEASURES OF CENTER ARE THE SAME FOR BOTH STUDENTS

4. What is **RANGE** of the data set?

ALAN (RANGE) = $100 - 80 = 20$

BR1 (RANGE) = $91 - 89 = 2$

5. What is the **IQR** of each data set?

ALAN (Q_1) = $\frac{81+82}{2} = 81.5$

BR1 (Q_1) = $\frac{89+90}{2} = 89.5$

ALAN (Q_3) = $\frac{98+99}{2} = 98.5$

BR1 (Q_3) = $\frac{90+91}{2} = 90.5$

6. Consider using the measures of variability (or measures of spread) as a possible determining factor for the scholarship recipient.

ALAN (IQR) = $98.5 - 81.5 = 17$

BR1 (IQR) = $90.5 - 89.5 = 1$

a. Find Mean Deviation

b. Find Variance, σ^2 .

c. Find Standard Deviation.

$$\left(\frac{\sum_{i=1}^n |X_i - \bar{X}|}{n} \right)$$

$$\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \right)$$

$$\left(\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} \right)$$

Alan's Data (X_i)	$X_i - \bar{X}$ DEVIATION	$ X_i - \bar{X} $ DEVIATION	$(X_i - \bar{X})^2$ (DEVIATION) ²
90	-90 = 0	0	0 ² = 0
90	-90 = 0	0	0 ² = 0
80	-90 = -10	10	(-10) ² = 100
100	-90 = 10	10	(10) ² = 100
99	-90 = 9	9	(9) ² = 81
81	-90 = -9	9	(-9) ² = 81
98	-90 = 8	8	(8) ² = 64
82	-90 = -8	8	(-8) ² = 64

Brianna's Data (X_i)	$X_i - \bar{X}$	$ X_i - \bar{X} $	$(X_i - \bar{X})^2$
90	-90 = 0	0	0 ² = 0
90	-90 = 0	0	0 ² = 0
91	-90 = 1	1	(1) ² = 1
89	-90 = -1	1	(-1) ² = 1
91	-90 = 1	1	(1) ² = 1
89	-90 = -1	1	(-1) ² = 1
90	-90 = 0	0	0 ² = 0
90	-90 = 0	0	0 ² = 0

MEANS: $\frac{0}{8} = 0$ $\frac{54}{8} = 6.75$ $\frac{490}{8} = 61.25$
Should be Zero Mean Deviation Variation

MEANS: $\frac{0}{8} = 0$ $\frac{4}{8} = 0.5$ $\frac{4}{8} = 0.5$
Should be Zero Mean Deviation Variation

$\sqrt{61.25} \approx 7.83$
Standard Deviation

$\sqrt{0.5} \approx 0.71$
Standard Deviation

7. What does the difference in the measures of variability (spread) suggest?

THE DATA SUGGESTS THAT BRIANNA IS MUCH MORE CONSISTENT BUT HAS NEVER SCORED ABOVE A 91.

8. Using your measures, explain which student you think the teacher should choose and why. **BRIANNA**



Alan



OR

Brianna

9. Matching:

- D Has a standard deviation of 0.
- B Has a standard deviation of 1.
- E Has a standard deviation of 2.
- C, A Has a standard deviation of 3.

- $\sigma = 3$ A. 1, 7, 1, 7, 7, 1
- $\sigma = 1$ B. 1, 3, 1, 3, 1, 3
- $\sigma = 3$ C. 4, -2, 8, 2, 4, 2
- $\sigma = 0$ D. 1, 1, 1, 1, 1, 1
- $\sigma = 2$ E. 5, 5, 5, 1, 1, 1

$$4 = \frac{24}{6} = \frac{1+7+1+7+7+1}{6}$$

X	\bar{X}	DEV	DEV ²
1	4	-3	9
7	4	3	9
1	4	-3	9
7	4	3	9
7	4	3	9
1	4	-3	9
			<u>54</u>
			$\frac{54}{6} = 9$
			$\sqrt{9} = 3$

10. Can you make a data set of 6 elements that has a standard deviation of 4? **RANGE = 8**
 IT SEEMS THAT THE STANDARD DEVIATION IS JUST HALF OF THE RANGE WHEN YOUR DATA POINTS ALTERNATE BACK AND FORTH BETWEEN 2 VALUES.
 SO, WE CAN CREATE A DATASET LIKE THIS WITH A RANGE OF 8. $\rightarrow 0, 8, 0, 8, 0, 8$

11. The table below shows the scores of the last 6 based ball games.

Winning Score	5	2	6	9	5	3
Losing Score	3	1	2	2	4	0

WINNING MARGIN: 2 1 4 7 1 3

The winning margin for each game is the difference between the winning score and the losing score. What is the standard deviation of the winning margins for these data?

$$\sigma \approx 2.08$$

L1	L2	L3	2
Winning Margin	█	-----	
L2(1)=			

```
1-Var Stats
x̄=3
Σx=18
Σx²=80
Sx=2.0075005
σx=2.081665999
n=6
```

12. The following shows the shoes sizes of the students in a class

Shoe Size	7	8	9	10	11	12
Frequency	1	2	4	3	2	2

DATA SET: 7, 8, 8, 9, 9, 9, 9, 10, 10, 10, 11, 11, 12, 12

What is the standard deviation of this data set? $\sigma \approx 1.44$

What is the range of the data set? **RANGE: 12 - 7 = 5**

L1	L2	L3	2
Shoe Size	█	-----	
L2(1)=			

```
1-Var Stats
x̄=9.642857143
Σx=135
Σx²=1331
Sx=1.499083969
σx=1.444553458
n=14
```

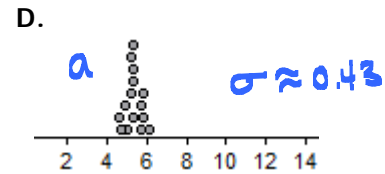
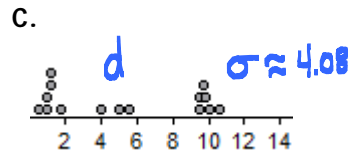
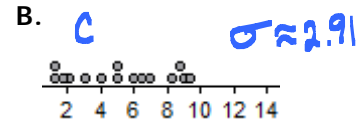
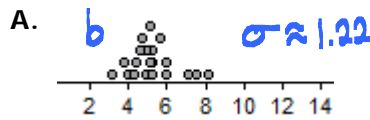
13. Matching: Use the following dot plots to estimate which of each of the following distributions corresponds to which given standard deviation?

a. $\sigma \approx 0.43$ **D**

b. $\sigma \approx 1.22$ **A**

c. $\sigma \approx 2.91$ **B**

d. $\sigma \approx 4.08$ **C**



14. The following represents the grades of each student on a test
31 79 97 70 70 79

~~31, 79, 70, 79, 79, 97~~

Find the MEAN: $\frac{426}{6} = 71$

Find the MEDIAN: $\frac{70+79}{2} = 74.5$

Find the MODE: **70 AND 79**

Which the most appropriate central tendency to use to describe the data set? and Why?

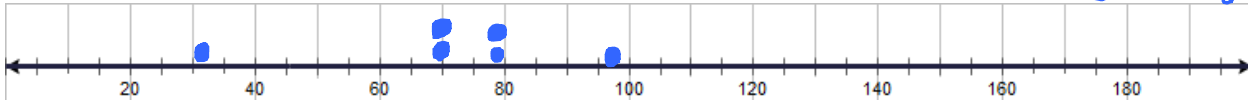
Find the RANGE: $97 - 31 = 66$

Find the ABS. MEAN DEV.: 14

Find the STANDARD DEV.: $\sqrt{401} \approx 20.02$

Data	Mean	Deviation	Deviation	Deviation ²
31	71	-40	40	1600
79	71	8	8	64
97	71	26	26	676
70	71	-1	1	1
70	71	-1	1	1
79	71	8	8	64
		$\frac{0}{6} = 0$	$\frac{84}{6} = 14$	$\frac{2406}{6} = 401$

Create a Dot Plot of the data:



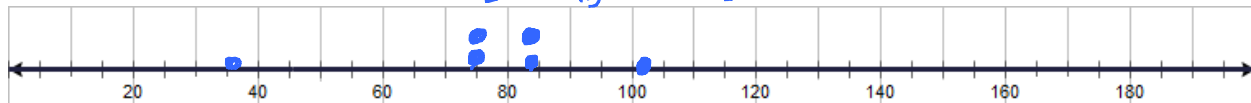
The teacher thought the class average was too low and decided to curve the tests 5 points. Add 5 points to everyone's grade and re-evaluate the following:

MEAN: **76** MEDIAN: **79.5** MODE: **75 AND 84**

RANGE: **66** ABS. MEAN DEV.: **14** STANDARD DEV.: ≈ 20.02
 $102 - 36 = 66$ $\sqrt{401}$

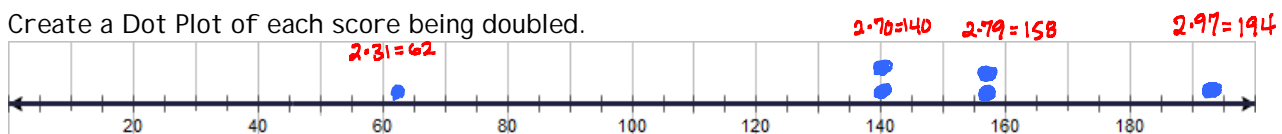
Data	Mean	Deviation	Deviation	Deviation ²
36	76	-40	40	1600
84	76	8	8	64
102	76	26	26	676
75	76	-1	1	1
75	76	-1	1	1
84	76	8	8	64
		$\frac{0}{6} = 0$	$\frac{84}{6} = 14$	$\frac{2406}{6} = 401$

Describe how each statistic changed. **CENTERS WENT UP BY 5**
SPREAD STAYED THE SAME



What do you think would happen to each of the statistics if the teacher decided to double each student's score?

Create a Dot Plot of each score being doubled.



MEASURES OF CENTER ALL DOUBLED (MEAN, MEDIAN, MODE).
MEASURES OF SPREAD ALSO ALL DOUBLED (RANGE, IQR, ABS. MEAN DEV., STD DEV.)