Describing Data
Variations (Spread)

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

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

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

   **MMA (RANGE):** 21 - 15 = 6
   
   **AUM (RANGE):** 40 - 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:

Simsong’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: 1 hr, 1 hr, 2 hrs, 6 hrs, 8 hrs, 10 hrs, 14 hrs

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

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

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

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

   **UNIV (MEdIAN)** = 5 hrs

   **VOID (MEdIAN)** = 6 hrs

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

   **UNIV (RANGE)** = 7 - 3 = 4 hrs

   **VOID (RANGE)** = 14 - 1 = 13 hrs

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

   **UNIV (Q1)** = 3 hrs

   **VOID (Q1)** = 1

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

   **UNIV (Q3)** = 7 hrs

   **VOID (Q3)** = 10

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

   **UNIV (IQR)** = Q₃ - Q₁ = 7 - 3 = 4 hrs

   **VOID (IQR)** = 10 - 1 = 9 hrs
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  
**Brianna**: 90, 90, 91, 89, 91, 89, 90, 90

1. Which student has the higher arithmetic mean, (average $\bar{x}$)?
   - **Alan** $(\bar{x}) = \frac{90+90+80+100+99+81+98+82}{8} = 90$
   - **Brianna** $(\bar{x}) = \frac{90+90+91+89+91+89+90+90}{8} = 90$

2. Which student has the higher median?
   - **Alan** $(\text{MED}) = 90$
   - **Brianna** $(\text{MED}) = \frac{90+90+89+90}{2} = 90$

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** $(\text{RANGE}) = 100 - 80 = 20$
   - **Brianna** $(\text{RANGE}) = 91 - 89 = 2$

5. What is the IQR of each data set?
   - **Alan** $(\text{IQR}) = 99.5 - 89.5 = 10$
   - **Brianna** $(\text{IQR}) = 90.5 - 89.5 = 1$

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

   a. Find Mean Deviation
   
      $\left( \frac{\sum |X_i - \bar{x}|}{n} \right)$
   
   b. Find Variance, $\sigma^2$.
      
      $\left( \frac{\sum (X_i - \bar{x})^2}{n} \right)$
   
   c. Find Standard Deviation.
      
      $\sqrt{\left( \frac{\sum (X_i - \bar{x})^2}{n} \right)}$

### Alan’s Data (X_i) | X_i – $\bar{x}$ | $|X_i - \bar{x}|$ | $(X_i - \bar{x})^2$
---|---|---|---
90 | -90 | 0 | 0
90 | -90 | 0 | 0
80 | -90 | -10 | 100
100 | -90 | 10 | 100
99 | -90 | 9 | 81
81 | -90 | -9 | 81
98 | -90 | 8 | 64
82 | -90 | -8 | 64

**MEANS:**

$\frac{0}{8} = 0$ | $\frac{84}{8} = 10.5$ | $\frac{810}{8} = 101.25$

Should be Zero | Mean Deviation | Variation

**Standard Deviation** $= 7.83$

### Brianna’s Data (X_i) | X_i – $\bar{x}$ | $|X_i - \bar{x}|$ | $(X_i - \bar{x})^2$
---|---|---|---
90 | -90 | 0 | 0
90 | -90 | 0 | 0
91 | -90 | 1 | 1
89 | -90 | 1 | 1
91 | -90 | 1 | 1
89 | -90 | 1 | 1
90 | -90 | 0 | 0
90 | -90 | 0 | 0

**MEANS:**

$\frac{0}{8} = 0$ | $\frac{4}{8} = 0.5$ | $\frac{4}{8} = 0.5$

Should be Zero | Mean Deviation | Variation

**Standard Deviation** $= 0.71$
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**

9. Matching:

<table>
<thead>
<tr>
<th>A. 1, 7, 1, 7, 7, 1</th>
<th>B. 1, 3, 1, 3, 1, 3</th>
<th>C. 4, –2, 8, 2, 4, 2</th>
<th>D. 1, 1, 1, 1, 1, 1</th>
<th>E. 5, 5, 5, 1, 1, 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ = 3</td>
<td>σ = 1</td>
<td>σ = 3</td>
<td>σ = 0</td>
<td>σ = 2</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
</tbody>
</table>

10. Can you make a data set of 6 elements that has a standard deviation of 4?

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: **0, 0, 0, 0, 0, 0**

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

<table>
<thead>
<tr>
<th>Winning Score</th>
<th>5</th>
<th>2</th>
<th>6</th>
<th>9</th>
<th>5</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Losing Score</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

**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? **σ ≈ 2.08**

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

<table>
<thead>
<tr>
<th>Shoe Size</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

**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? **σ ≈ 1.44**

What is the range of the data set? **RANGE: 12 - 7 = 5**
13. Matching: Use the following dot plots to estimate which of each of the following distributions corresponds to which given standard deviation?

a. $\sigma \approx 0.43$  
   ![Dot Plot A](image1.png)

b. $\sigma \approx 1.22$  
   ![Dot Plot B](image2.png)

c. $\sigma \approx 2.91$  
   ![Dot Plot C](image3.png)

d. $\sigma \approx 4.08$  
   ![Dot Plot D](image4.png)

14. The following represents the grades of each student on a test

<table>
<thead>
<tr>
<th>Score</th>
<th>31</th>
<th>79</th>
<th>97</th>
<th>70</th>
<th>70</th>
<th>79</th>
</tr>
</thead>
</table>

Find the MEAN: $\frac{31 + 79 + 97 + 70 + 70 + 79}{6} = 71$

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.: $\sqrt{\frac{\sum (x - \mu)^2}{n}} \approx 20.02$

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:

<table>
<thead>
<tr>
<th>Data</th>
<th>Mean</th>
<th>Deviation</th>
<th></th>
<th>Deviation</th>
<th>Deviation$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>76</td>
<td>-40</td>
<td>40</td>
<td>1600</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>76</td>
<td>8</td>
<td>8</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>76</td>
<td>20</td>
<td>20</td>
<td>676</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>76</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>76</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>76</td>
<td>8</td>
<td>8</td>
<td>64</td>
<td></td>
</tr>
</tbody>
</table>

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

Create Dot Plot

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).