Measures of Spread: Standard Deviation

So far in our study of numerical measures used to describe data sets, we have focused on the mean and the median. These measures of center tell us the most typical value of the data set. If we were asked to make a prediction about a member of a data set, we would use a measure of center to predict that value. However, measures of center do not give us the complete picture.

Consider the following test scores:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Student  | Test 1  | Test 2  | Test 3  | Test 4  |
| Johnny  | 65 | 82 | 93 | 100 |
| Will  | 82 | 86 | 88 | 84 |
| Anna  | 80 | 99 | 73 | 88 |

Who is the best student? How do you know?

# Thinking about the Situation

**Discuss the following with your partner or group. Write your answers on your own paper. Be prepared to share your answers with the class.**

1. What is the mean test score for each student?
2. Based on the mean, who is the best student?
3. If asked to select one student, who would you pick as the best student? Explain.

# Investigation 1: Deviation from the Mean

**Discuss the following with your partner or group. Write your answers on your own paper. Be prepared to share your answers with the class.**

Usually we calculate the mean, or average, test score to describe how a student is doing. Johnny, Will, and Anna all have the same average. However, these three students do not seem to be “equal” in their test performance. We need more information than just the typical test score to describe how they are doing. One thing we can look at is how consistent each student is with their test performance. Does each student tend to do about the same on each test, or does it vary a lot from test to test? Measures of spread will give us that information. In statistics, deviation is the amount that a single data value differs from the mean.

1. Complete the table below by finding the deviation from the mean for each test score for each student.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Score**$$x$$ | **Mean**$$\overbar{x}$$ | **Deviation from the Mean**$$x-\overbar{x}$$ |
| **Johnny** |
| **Test 1**  |  |  |  |
| **Test 2**  |  |  |  |
| **Test 3**  |  |  |  |
| **Test 4**  |  |  |  |
| **Sum of Deviations** |  |
|  | **Score**$$x$$ | **Mean**$$\overbar{x}$$ | **Deviation from** **the Mean**$$x-\overbar{x}$$ |
| **Will** |
| **Test 1**  |  |  |  |
| **Test 2**  |  |  |  |
| **Test 3**  |  |  |  |
| **Test 4**  |  |  |  |
| **Sum of Deviations** |  |

1. How does the sum of the deviations from the mean relate to the mean being the balance point for a set of data?

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Score**$$x$$ | **Mean**$$\overbar{x}$$ | **Deviation from** **the Mean**$$x-\overbar{x}$$ |
| **Anna** |
| **Test 1**  |  |  |  |
| **Test 2**  |  |  |  |
| **Test 3**  |  |  |  |
| **Test 4**  |  |  |  |
| **Sum of Deviations** |  |

# Investigation 2: Mean Absolute Deviation

**Discuss the following with your partner or group. Write your answers on your own paper. Be prepared to share your answers with the class.**

One way to measure consistency is to find the average deviation from the mean. In other words, how far do most values in a data set fall from the mean? One way to answer this question would be to find the average deviation, or distance, that the data values fall from the mean. So we would add up the deviations to find the total deviation and then divide by the number of data values to find the mean deviation. However, the fact that the deviations from the mean always add up to zero is a problem. No matter what we divide zero by, we always get zero! When talking about spread, a value of zero indicates that there is no spread, or variability. One way to fix this problem is to look at only the distances from the mean, and not their directions as indicated by the sign of the deviation (positive or negative). We can take the absolute value of the distances and then find the average distance.

1. Complete the table below by filling in the deviation from the mean for each test score for each student that you calculated in Investigation 1. Then find the absolute value of each deviation.

|  | **Deviation from the Mean**$$x-\overbar{x}$$ | **Absolute Deviation from the Mean**$$x-\overbar{x}$$ |
| --- | --- | --- |
| **Johnny** |
| **Test 1**  |  |  |
| **Test 2**  |  |  |
| **Test 3**  |  |  |
| **Test 4**  |  |  |
| **Sum** |  |
| **Mean of Absolute Deviation** |  |

|  | **Deviation from the Mean**$$x-\overbar{x}$$ | **Absolute Deviation from the Mean**$$x-\overbar{x}$$ |
| --- | --- | --- |
| **Will** |
| **Test 1**  |  |  |
| **Test 2**  |  |  |
| **Test 3**  |  |  |
| **Test 4**  |  |  |
| **Sum** |  |
| **Mean of Absolute Deviation** |  |

|  | **Deviation from the Mean**$$x-\overbar{x}$$ | **Absolute Deviation from the Mean**$$x-\overbar{x}$$ |
| --- | --- | --- |
| **Anna** |
| **Test 1**  |  |  |
| **Test 2**  |  |  |
| **Test 3**  |  |  |
| **Test 4**  |  |  |
| **Sum** |  |
| **Mean of Absolute Deviation** |  |

The average of the absolute deviations from the mean is called the **Mean Absolute Deviation (MAD)**.

1. What does the Mean Absolute Deviation (MAD) tell you about each student? Is there one student who seems to be more consistent than the others?
2. Interpret the MAD for Johnny in context.

# Investigation 3: Calculating the Standard Deviation

Below is the formula for calculating the standard deviation. It looks pretty complicated, doesn’t it? Let’s break it down step by step so we can see how it is finding the “average deviation from the mean.”

Let’s start with Johnny’s data.

**Step 1:** In the table below, record the data values (test scores) in the second column labeled “Value.” The *x* is used to denote a value from the data set. *The first value is written for you.*

**Step 2:** Find the mean of the test scores (we did this in the “Thinking About the Situation”) and record at the bottom of the second column next to the symbol μ. Mu (pronounced “mew”) is the lowercase Greek letter that later became our letter “m.” μ is another symbol that we use for mean (in addition to $\overbar{x}).$

**Step 3:** In the third column, find the deviation from the mean for each test score by taking each test score and subtracting the mean. *The first difference has been done for you.*

**Step 4:** Add the values in the third column to find the sum of the deviations from the mean. If you have done everything correctly so far, the sum should be zero. The capital Greek letter Σ, called sigma, is a symbol that is used to indicate the sum.

**Step 5:** Square each deviation to make it positive and record these values in the last column of the table. *The first value is done for you.*

**Step 6:** Find the sum of the squared deviations by adding up the values in the fourth column and putting the sum at the bottom of the column. This is the sum of the squared deviations from the mean.

**Step 7:** Find the average of the squared deviations from the mean by dividing the sum of column four by the number of data values, *n*, (the number of test scores).

**Step 8:** “Un-do” the squaring by taking the square root. Now you have found the standard deviation! The symbol for standard deviation is the lower-case letter sigma, σ.

**Johnny’s Data**

|  |  |  |  |
| --- | --- | --- | --- |
| Test | Value $(x)$ | Deviations from the Mean Value – Mean $(x-μ$) | Squared Deviations from the Mean(Value – Mean)2 $(x-μ)^{2}$ |
| 1 | 65 | 65-85 = -20 | (-20)2 = 400 |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
|  | **Mean** $μ=$ | **Sum** $\sum\_{}^{}(x-μ)=$ | **Sum** $\sum\_{}^{}(x-μ)^{2}=$ |
|  |  | **Average of Squared Deviations** | $$\frac{\sum\_{}^{}(x-μ)^{2}}{n}=$$ |
|  |  | **Square Root of Average of Squared Deviations** | $$\sqrt{\frac{\sum\_{}^{}(x-μ)^{2}}{n}}=$$ |

Now repeat the process with Will and Anna’s data.

**Will’s Data**

|  |  |  |  |
| --- | --- | --- | --- |
| Test | Value $(x)$ | Deviations from the Mean Value – Mean $(x-μ$) | Squared Deviations from the Mean(Value – Mean)2 $(x-μ)^{2}$ |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
|  | **Mean** $μ=$ | **Sum** $\sum\_{}^{}(x-μ)=$ | **Sum** $\sum\_{}^{}(x-μ)^{2}=$ |
|  |  | **Average of Squared Deviations** | $$\frac{\sum\_{}^{}(x-μ)^{2}}{n}=$$ |
|  |  | **Square Root of Average of Squared Deviations** | $$\sqrt{\frac{\sum\_{}^{}(x-μ)^{2}}{n}}=$$ |

**Anna’s Data**

|  |  |  |  |
| --- | --- | --- | --- |
| Test | Value $(x)$ | Deviations from the Mean Value – Mean $(x-μ$) | Squared Deviations from the Mean(Value – Mean)2 $(x-μ)^{2}$ |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
|  | **Mean** $μ=$ | **Sum** $\sum\_{}^{}(x-μ)=$ | **Sum** $\sum\_{}^{}(x-μ)^{2}=$ |
|  |  | **Average of Squared Deviations** | $$\frac{\sum\_{}^{}(x-μ)^{2}}{n}=$$ |
|  |  | **Square Root of Average of Squared Deviations** | $$\sqrt{\frac{\sum\_{}^{}(x-μ)^{2}}{n}}=$$ |

Discussion Questions:

1. Translate into words: $\sum\_{}^{}(x-μ)^{2}$.
2. Interpret Anna’s standard deviation in context.
3. Who is the best student? How do you know?

**Finding the mean, median and standard deviation on the calculator.**

Enter the following data into L1 in the calculator.

28, 48, 53, 25, 38, 23, 49, 32











(Sample)

* Notations: σx (Population)
* We will be using Sx (Sample Standard Deviation) for standard deviation, this is used more frequently because in order to use σx (Population Standard Deviation) we need to have the data for the entire population and that is not often the case. Sample Deviation is the larger of the two values.