**Variance and Standard Deviation**

**Variance ()** – the average of the squared differences from the mean

**Standard Deviation ()** – the square root of the variance

* Low standard deviations indicate data that are clustered near the measures of central tendency (mean, median, mode).
* High standard deviations indicate data that are spread out from the center.

**Practice Problem #1**

The following data represents the number of free throws Mike made after 10 sets of 50 free throws.

35, 40, 38, 43, 38, 36, 46, 47, 40, 37

Find the mean and standard deviation of the data.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data Value x | 35 | 40 | 38 | 43 | 38 | 36 | 46 | 47 | 40 | 37 |
| x – |  |  |  |  |  |  |  |  |  |  |
| (x – )2 |  |  |  |  |  |  |  |  |  |  |

1. What is the mean? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

This value will be represented by .

2. Now take each data value x and subtract the mean from it. Fill in the 2nd row with your answers.

3. Now for the third row, take your answers from the 2nd row and square them.

To find the variance we are going to add up every number in the third row and divide by the total number of trials.

Variance =  = \_\_\_\_\_\_\_\_\_\_\_\_\_\_

Now to find the standard deviation we need to take the square root of the variance.

Standard Deviation =  = \_\_\_\_\_\_\_\_\_\_\_\_\_\_

\*\* This means that most of the data is within ± 3.9 of the mean 40. \*\* Let’s look at another set of data to see if we can’t start to get a better understanding about standard deviation.

**Practice Problem #2**

The following data represents the number of free throws Austin made after 10 sets of 50 free throws.

23, 46, 34, 12, 8, 48, 29, 40, 6, 44

Find the mean and standard deviation of the data.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data Value x |  |  |  |  |  |  |  |  |  |  |
| x – |  |  |  |  |  |  |  |  |  |  |
| (x – )2 |  |  |  |  |  |  |  |  |  |  |

Mean = \_\_\_\_\_\_\_\_\_\_\_

Variance = \_\_\_\_\_\_\_\_\_\_\_

Standard Deviation = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What does the difference in standard deviation from Mike’s data set to Austin’s data set mean? Jot some thoughts down.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Outliers**

Outliers are extreme values that are either much greater or much less than the other data values. Because of this they have a huge effect on the mean and standard deviation. Therefore in many experiments and data sets the outlier is typically removed. But how do we decide what constitutes an outlier? One way is to look at data values that are more than 3 standard deviations away from the mean.

For example, let’s look at this scenario. *mean* = 41.2 and *standard* *deviation* = 3.6

Since we want values that are 3 standard deviations away, 3(3.6) = \_\_\_\_\_\_\_\_\_\_\_\_

Now both subtract and add 10.8 to the mean of 41.2 to find this range.

Any value greater than \_\_\_\_\_\_\_\_\_\_\_\_\_\_ and any value less than \_\_\_\_\_\_\_\_\_\_\_\_\_ would be considered an outlier in the data set.

**Assignment: Problem Solving Worksheet + Practice B Worksheet**