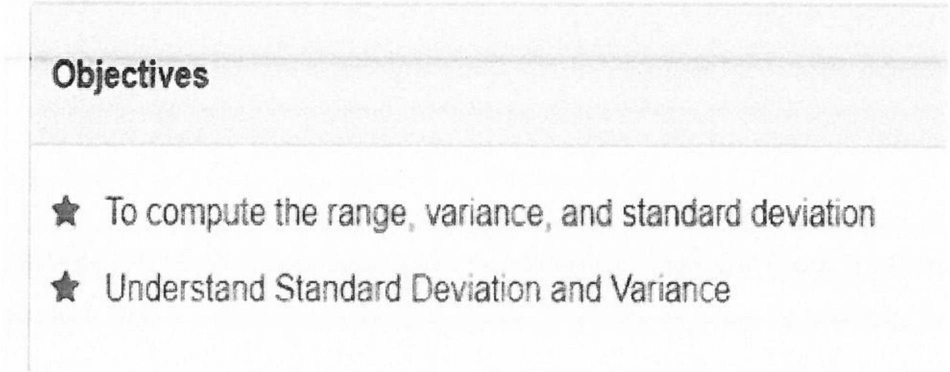


Objectives

- ★ To compute the range, variance, and standard deviation
- ★ Understand Standard Deviation and Variance

Sep 26-7:07 PM

The concept of variability (also referred to as dispersion or spread) is as vague as the concept of central tendency. And vague concepts lead to different measurement ideas. The same issues that are important in evaluating measures of location are meaningful in evaluating measures of dispersion.

Sep 26-7:08 PM

The concept of variability (also referred to as dispersion or spread) is as vague as the concept of central tendency. And vague concepts lead to different measurement ideas. The same issues that are important in evaluating measures of location are meaningful in evaluating measures of dispersion.

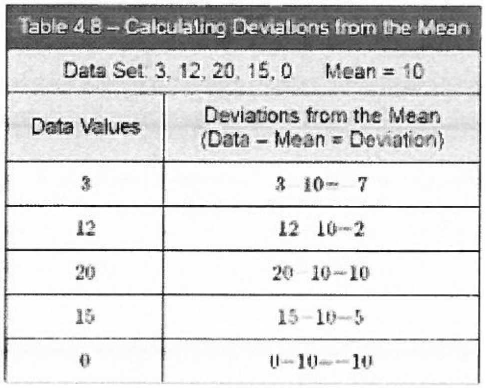


Table 4.8 – Calculating Deviations from the Mean

Data Set: 3, 12, 20, 15, 0 Mean = 10

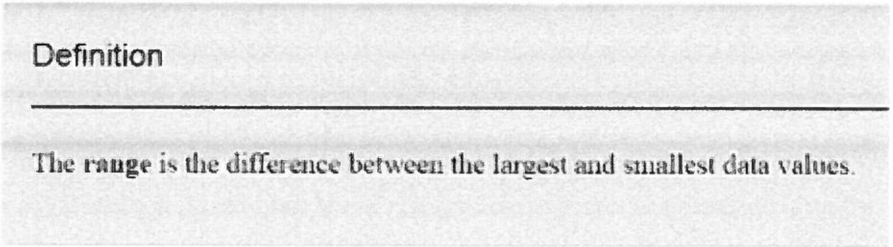
Data Values	Deviations from the Mean (Data – Mean = Deviation)
3	3−10=−7
12	12−10=2
20	20−10=10
15	15−10=5
0	0−10=−10

Because the mean is the point at which the sum of the positive deviations equals the sum of the absolute values of the negative deviations, the deviations will always sum to zero. Many of the variability measures average the deviations in some form.

Sep 26-7:10 PM

Range

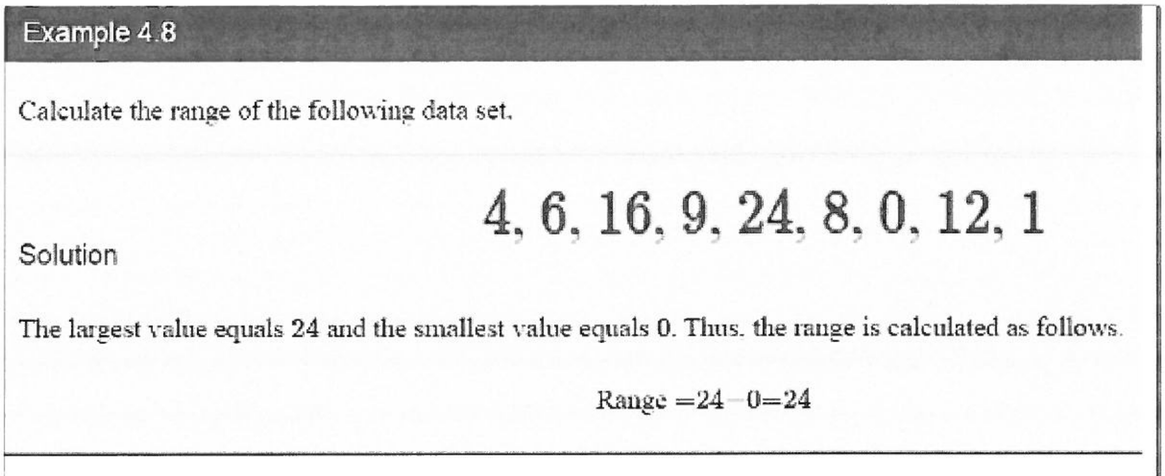
The range is the simplest measure of dispersion. It does not provide much depth or understanding of the measure of spread and does not use the deviation concept.



Definition

The **range** is the difference between the largest and smallest data values.

Sep 26-7:12 PM



Example 4.8

Calculate the range of the following data set.

$$4, 6, 16, 9, 24, 8, 0, 12, 1$$

Solution

The largest value equals 24 and the smallest value equals 0. Thus. the range is calculated as follows.

$$\text{Range} = 24 - 0 = 24$$

Sep 26-7:22 PM

The problem with the range is that it is also affected by outliers, and it does not bring all the information in the data directly to bear on the problem of measuring variation. That is, the range only uses two values (the largest and smallest) to measure spread rather than all of the observations. The other measures of dispersion discussed in this lesson are generally more appropriate measures of spread.

Sep 26-7:23 PM

Mean Absolute Deviation

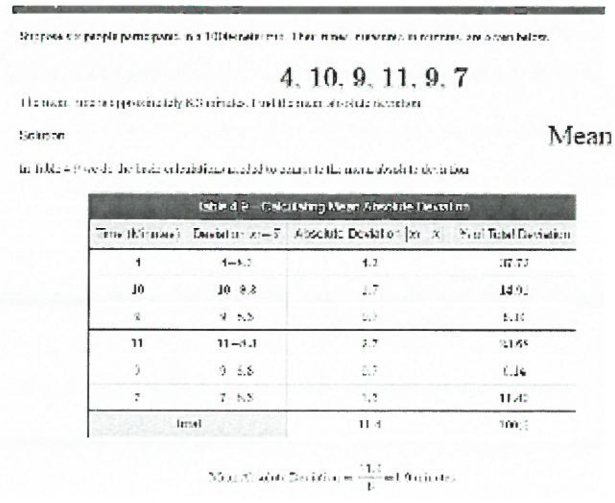
One of the ways of obtaining information about the spread of the data is to analyze the deviations from the mean. Instead of adding the raw deviations, suppose the absolute values of the deviations (which can be interpreted as distance from the mean) are summed and divided by the number of deviations. This new measure computes the average distance from the mean for the data set. This measure is called the mean absolute deviation. If data set A has a larger average deviation than B, then it is reasonable to believe that data set A has more variability than data set B.

Formula: Mean Absolute Deviation

The sample mean absolute deviation (MAD) is given by

$$\text{MAD}=\frac{\sum|x_i-\bar{x}|}{n}$$

Sep 26-7:25 PM



Suppose six people participate in a 100-meter race. Their times, measured in minutes, are given below.

4, 10, 9, 11, 9, 7

The mean time is approximately 8.3 minutes. Find the mean absolute deviation.

Solution

In Table 4.2 we show the basic calculations needed to compute the mean absolute deviation.

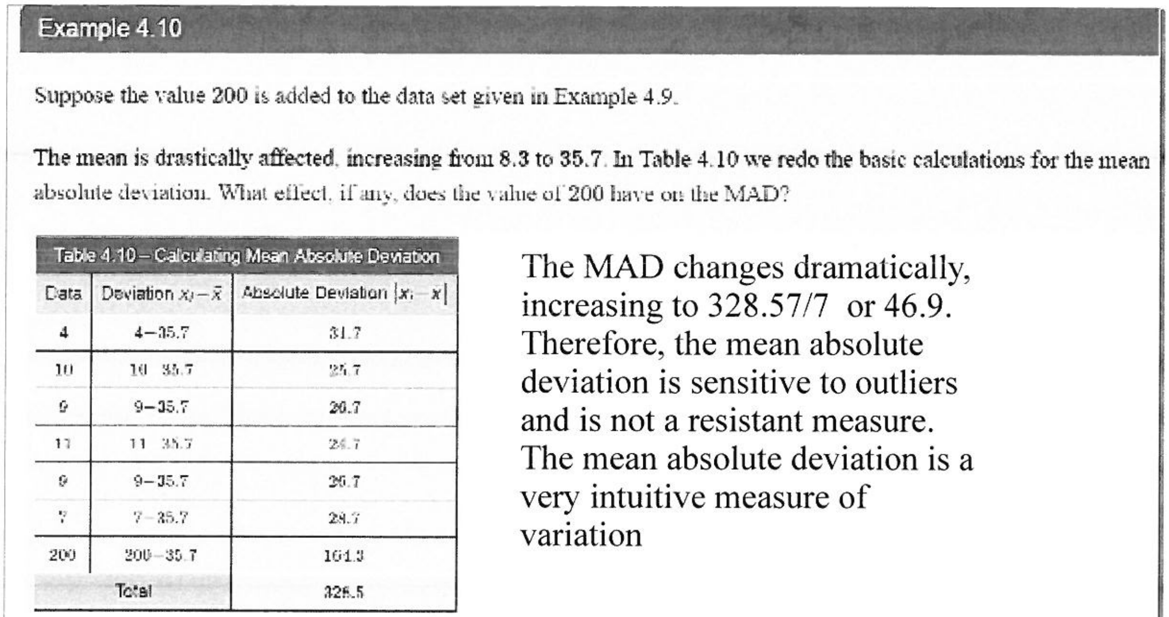
Table 4.2 – Calculating Mean Absolute Deviation

Time (Minutes)	Deviation $x_i-\bar{x}$	Absolute Deviation $\lvert x_i-\bar{x}\rvert$	% of Total Deviation
4	4−8.3	4.3	37.72
10	10−8.3	1.7	14.91
9	9−8.3	0.7	6.14
11	11−8.3	2.7	23.68
9	9−8.3	0.7	6.14
7	7−8.3	1.3	11.40
Total		11.4	100.0

$$\text{Mean Absolute Deviation}=\frac{11.4}{6}\approx1.9\text{ minutes}$$

$$\text{Mean Absolute Deviation}=\frac{11.4}{6}=1.9\text{ minutes}$$

Sep 26-7:27 PM



Example 4.10

Suppose the value 200 is added to the data set given in Example 4.9.

The mean is drastically affected, increasing from 8.3 to 35.7. In Table 4.10 we redo the basic calculations for the mean absolute deviation. What effect, if any, does the value of 200 have on the MAD?

Table 4.10 – Calculating Mean Absolute Deviation

Data	Deviation $x_i - \bar{x}$	Absolute Deviation $\|x_i - x\|$
4	4−35.7	31.7
10	10 35.7	25.7
9	9−35.7	26.7
11	11 35.7	24.7
9	9−35.7	26.7
7	7−35.7	28.7
200	200−35.7	164.3
Total		328.5

The MAD changes dramatically, increasing to 328.57/7 or 46.9. Therefore, the mean absolute deviation is sensitive to outliers and is not a resistant measure. The mean absolute deviation is a very intuitive measure of variation

Sep 26-7:30 PM

Variance and Standard Deviation

The variance and standard deviation are the most common measures of variability.

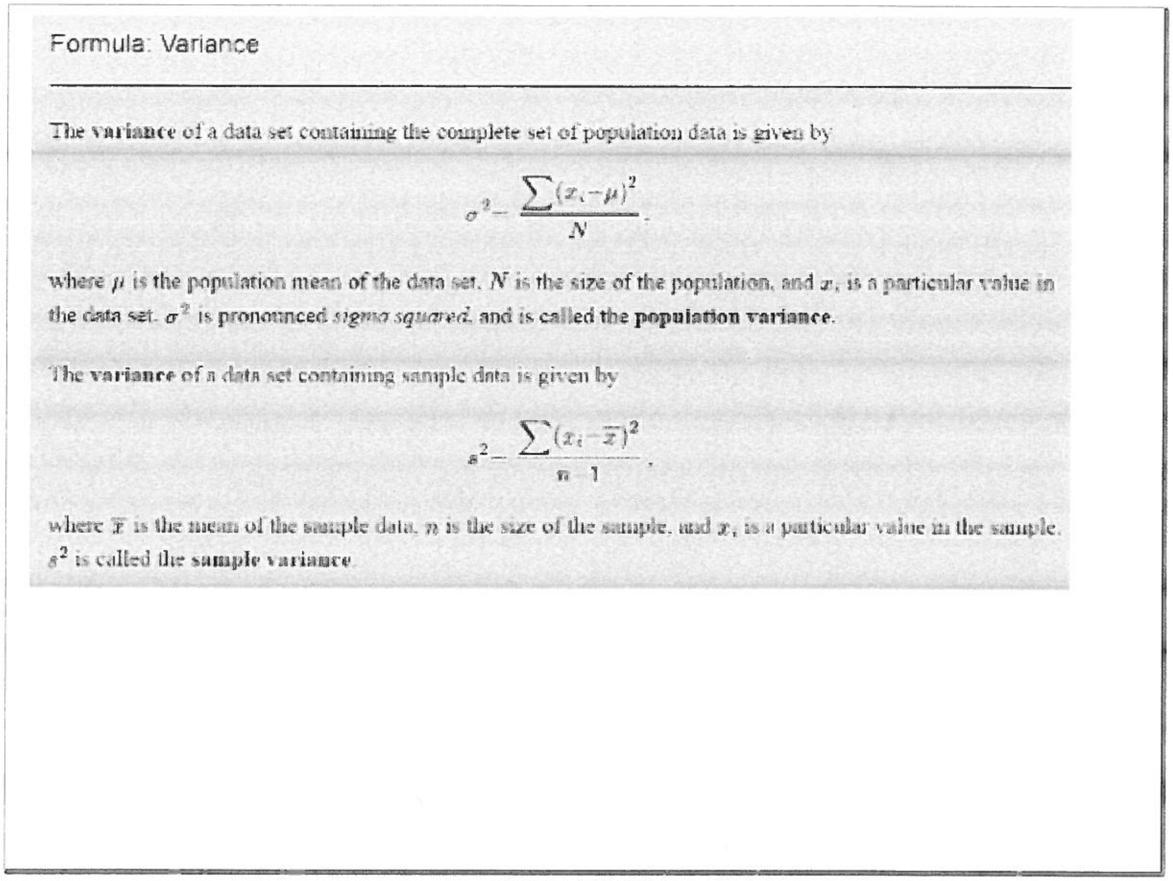
Since the standard deviation is computed directly from the variance, our discussion will center on the variance.

Like the MAD, the variance and standard deviation provide numerical measures of how the data vary around the mean.

If the data are tightly packed around the mean, the variance and standard deviation will be relatively small.

On the other hand, if the data are widely dispersed about the mean, the variance and standard deviation will be relatively large.

Sep 26-7:32 PM



Formula: Variance

The **variance** of a data set containing the complete set of population data is given by

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}.$$

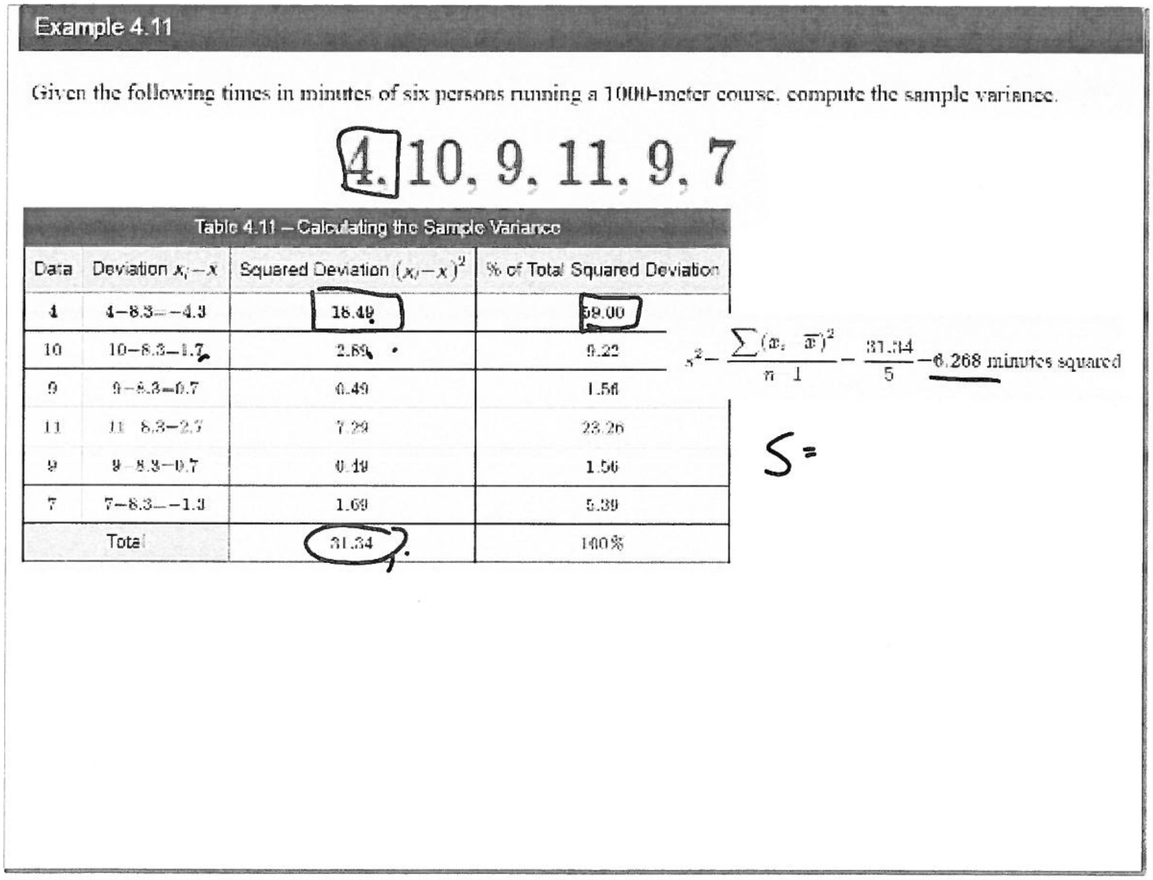
where μ is the population mean of the data set, N is the size of the population, and x_i is a particular value in the data set. σ^2 is pronounced *sigma squared*, and is called the **population variance**.

The **variance** of a data set containing sample data is given by

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1},$$

where $\overline{x}$ is the mean of the sample data, n is the size of the sample, and x_i is a particular value in the sample. s^2 is called the **sample variance**.

Sep 26-7:34 PM



Example 4.11

Given the following times in minutes of six persons running a 1000-meter course, compute the sample variance.

$\boxed{4,}$ 10, 9, 11, 9, 7

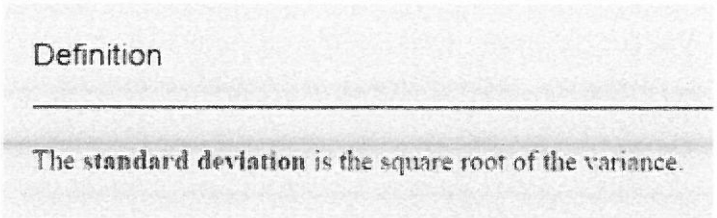
Table 4.11 – Calculating the Sample Variance

Data	Deviation $x_i - \overline{x}$	Squared Deviation $(x_i - \overline{x})^2$	% of Total Squared Deviation
4	4−8.3=−4.3	$\boxed{18.49}$	$\boxed{59.00}$
10	10−8.3=1.7	2.89	9.22
9	9−8.3=0.7	0.49	1.56
11	11−8.3=2.7	7.29	23.26
9	9−8.3=0.7	0.49	1.56
7	7−8.3=−1.3	1.69	5.39
Total		31.34	100%

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1} = \frac{31.34}{5} = \underline{6.268} \text{ minutes squared}$$

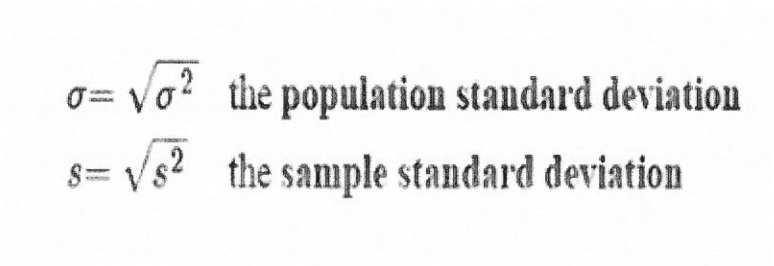
$S =$

Sep 26-7:34 PM



Definition

The **standard deviation** is the square root of the variance.



$\sigma = \sqrt{\sigma^2}$ **the population standard deviation**

$s = \sqrt{s^2}$ **the sample standard deviation**

Sep 26-7:37 PM

It is important to remember these symbols (σ and s) since the standard deviation is a fundamental statistical concept.

The standard deviation can be used to measure how far data values are from their mean. You will often see the majority of data values within one standard deviation of the mean. Further, relatively few data values will be more than two standard deviations from the mean.

The variance and standard deviation both suffer from the same problem as the mean: they are very sensitive to outliers.

Suppose the value 200 was added to the data in Example 4.11. The sample variance would increase from 6.268 to 5253.238.The new sample variance (which includes the outlier 200) then is 838 times as large as the original variance. The standard deviation increases from 2.50 to 72.50. The presence of the outlier tarnishes the interpretation of the standard deviation as a measure of variability.

Sep 26-7:38 PM

Another interesting property of the variance is that values further from the mean contribute a disproportionate amount to the value of the statistic.

In Example 4.11, one data point, 4, which is 4.3 units from the mean, contributes 59% of the variation in the data (see the column labeled "% of Total Squared Deviation" in Table 4.11). Compare this to the data point 7, which is 1.3 units from the mean, yet only contributes 5.39% of the total variation. The reason that 4 contributes so heavily to the total variation is because the deviations are squared. By squaring the deviations, values further from the mean have disproportionate effects on the sum of the squared deviations.

While there are a number of descriptive tools available for summarizing variability, the variance and standard deviation are the most frequently used statistics.

Sep 26-7:42 PM

Sep 26-7:42 PM