# Lesson 13-3 Using Series in Statistics Formulas

**Vocabulary**

mean

measure of center, measure of central tendency

absolute deviation

mean absolute deviation, *m.a.d.*

standard deviation, *s.d.*

**Mental Math**

**Roderick has averaged 87 on two advanced algebra tests. What is the minimum score he can receive on the final test if he wants to finish the course with an average test score of at least**

**a.** 90?

**b.** 70?

**c.** 95?

**BIG IDEA** Formulas for certain statistics, such as the *mean*, *mean absolute deviation*, and *standard deviation*, involve sums.

When you used Σ-notation in the previous two lessons, there was a formula for the numbers being added. In many situations, and particularly in statistics, data may be represented with an index variable and no formula. You can think of the data as terms of a sequence. We also say that each datum is an element of a data set. In data sets, different elements can have the same value. For instance, suppose that you have to read 10 short stories this semester in your English class and they have the following numbers of pages:

$$6, 14, 3, 8, 8, 9, 4, 11, 10, 23.$$

Let $L_i$ be the length of the $i$th story, so $L_1 = 6$, $L_2 = 14$, $L_3 = 3$, and so on. The sum of these 10 numbers can be represented as $\sum_{i=1}^{10} L_i$. In this case, $\sum_{i=1}^{10} L_i = 96$, and so the mean of the lengths $L_i$ is $\frac{\sum_{i=1}^{10} L_i}{10} = \frac{96}{10} = 9.6$. That is, the mean length of a story is 9.6 pages.

In general, when $S$ is a data set of $n$ numbers $x_1, x_2, x_3, \ldots, x_n$, then the **mean** of $S = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

The Greek letter $\mu$ (mu, pronounced "mew") is customarily used to represent the mean of a data set.

## Deviations and Absolute Deviations

Recall from Lesson 3-5 that the difference between an element of a data set and the mean of the set is called the element's deviation from the mean. If the mean is $\mu$ and the element is $x_i$, then the deviation is

$$x_i - \mu.$$

For instance, in the data set of story lengths, the deviation of the 6-page story from the mean of 9.6 pages is 6 – 9.6, or -3.6. That is, the story is 3.6 pages shorter than the mean length. In general, deviation is positive if an element is larger than the mean and negative if an element is smaller than the mean.

The sum of the deviations of the elements of a set from the mean of the set is 0. That is,

$$\sum_{i=1}^{n}(x_i - \mu) = 0.$$

This suggests that the mean is a number on which the data set "balances." For this reason, the mean is a called a **measure of center,** or a **measure of central tendency,** of a data set.

STOP **QY1**

Suppose you are estimating the number of beans in a jar and you want to know how close your estimate $E$ is to the actual number $A$. Then you do not care whether $E$ is larger or smaller than $A$. You want to know the *absolute deviation* of your estimate from $A$, or $|E - A|$. The **absolute deviation** of an element of a data set from the mean of the set is

$$|x_i - \mu|.$$

For instance, the absolute deviation of the element 6 from the mean 9.6 is $|6 - 9.6| = |-3.6| = 3.6$.

STOP **QY2**

The mean of the absolute deviations is a statistic called the **mean absolute deviation**, or ***m.a.d***. The *m.a.d.* is a measure of the *spread* or *dispersion* of a data set. In Σ-notation,

$$m.a.d. = \frac{1}{n}\sum_{i=1}^{n}|x_i - \mu|.$$

**▶ QY1**

**a.** Find the deviations of all 10 elements of the data set of story lengths from the mean of the lengths.

**b.** Find the mean of the deviations.

**▶ QY2**

**a.** Find the absolute deviations of all 10 elements of the data set of story lengths from the mean of the lengths.

**b.** Find the mean of these absolute deviations.

## Example

You are offered the chance to play one of two games, Game A or Game B. In each game you reach into a jar and pull out a slip of paper. You win the dollar amount written on the slip. In both games there are 10 slips of paper in the jar with a mean value of $50.

Calculate the *m.a.d.* of the values on the slips for each game. Which game gives more spread out results?

| Game A Slips | | | | |
|---|---|---|---|---|
| $49 | $49 | $49 | $49 | $50 |
| $50 | $51 | $51 | $51 | $51 |

| Game B Slips | | | | |
|---|---|---|---|---|
| $0 | $0 | $0 | $0 | $0 |
| $0 | $0 | $0 | $0 | $500 |

**Solution** For each game, the mean dollar amount $\mu$ is \$50. For Game A, there are 8 slips whose amounts deviate by \$1 from the mean, and the other two slips have no deviation from the mean. So,

$$\text{m.a.d. for Game A} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \mu| = \frac{1}{10}(8 \cdot 1 + 2 \cdot 0) = 0.8.$$

In Game B, there are 9 amounts that deviate by \$50 from the mean, and 1 amount that deviates by \$450. So,

$$\text{m.a.d. for Game B} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \mu| = \frac{1}{10}(9 \cdot 50 + 1 \cdot 450) = 90.$$

There is quite a difference in the spread of dollar amounts in the games. The more the values are spread out from the mean, the less likely you are to win an amount of money close to the mean amount. Since 90 is greater than 0.8, the results of Game B are more spread out.

## The Standard Deviation

The *m.a.d.* of a data set is relatively easy to calculate, but it does not have as many useful properties as a second measure of spread, the *standard deviation*. The lower-case Greek letter sigma ($\sigma$) is often used to denote the standard deviation, as in the following definition.

### Definition of Standard Deviation

Let $S$ be a data set of $n$ numbers $\{x_1, x_2, ..., x_n\}$. Let $\mu$ be the mean of $S$. Then the **standard deviation,** or ***s.d.***, of $S$ is given by

$$s.d. = \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}.$$

This formula looks complicated, so it may help to describe it in words. To use the formula, find the mean of the data set and then square each element's deviation from the mean. Then find the mean of these squared deviations. The square root of this mean is the standard deviation. The mean of the squared deviations is also called the *variance* of the data set. So, the standard deviation of a set is the square root of the variance of the set.

Standard deviation is used in a wide variety of statistical analyses. For example, a readability index is a measure of how difficult written text is to understand. To find one index, a computer program analyzes random paragraphs to find the mean and standard deviation of the number of words per sentence.

## Activity

**MATERIALS** Paper and the first two paragraphs under the Big Idea of Lesson 13-5

**Step 1** Make a table like the one at the right and record the number of words in each sentence of each paragraph. When counting words, note that a "word" is any character or group of characters with a space before and after it.

| Number of words | | |
|---|---|---|
| Sentence Number | Paragraph 1 | Paragraph 2 |
| 1 | ? | ? |
| 2 | ? | ? |
| 3 | ? | ? |
| 4 | ? | ? |
| 5 | ? | ? |

**Step 2** Calculate the mean number of words per sentence in each paragraph.

**Step 3** Calculate the square of the deviation of each element in the Paragraph 1 data set from its mean. Repeat for the Paragraph 2 data set.

**Step 4** Find the mean of the squared deviations for each data set in Step 3. Then calculate the standard deviation for each.

**Step 5** Which paragraph from Lesson 13-5 do you think is easier to read and understand? Support your conclusion with your statistics from Steps 2–4.

Most spreadsheet programs and calculators calculate standard deviations. One CAS shows the results at the right for the standard deviations of the data sets in the Activity.

| | |
|---|---|
| stDevPop({12,11,10,11,9}) | 1.0198 |
| stDevPop({27,16,19,31,14}) | 6.52993 |

## Questions

### COVERING THE IDEAS

1. Find the mean absolute deviation of the data set {2, 3, 3, 4, 5, 6, 6, 6, 6, 7}.

2. In 2006, the Miami Heat salaries, in millions of dollars, were approximately 0.4, 0.74, 2.88, 0.15, 0.07, 5.53, 0.41, 0.075, 1.19, 2.5, 20, 0.74, 6.39, 0.41, 0.93, 3.84, 7.61, 8.25, and 1.33. Find the *m.a.d.* of these salaries.

3. Calculate the standard deviation of the data set of story lengths on the first page of this lesson.

4. A person bowls games of 158, 201, 175, and 134. For these scores, calculate
   a. the *m.a.d.*
   b. the *s.d.*

**In 5–9, suppose that 100 scores are identified as $s_1, s_2, \dots, s_{100}$. What does each expression represent?**

5. $\sum_{i=1}^{100} s_i$

6. $\frac{1}{100}\sum_{i=1}^{100} s_i$

7. $\frac{1}{100}\sum_{i=1}^{100} (s_i - \mu)$

8. $\frac{1}{100}\sum_{i=1}^{100} |s_i - \mu|$

9. $\sqrt{\frac{1}{100}\sum_{i=1}^{100} (s_i - \mu)^2}$

10. Refer to the two games described in the Example.
    a. Calculate the standard deviation of the amounts in each game
    b. Why would someone want to play Game B?

11. Fundraisers sell 10,000 raffle tickets for \$5 each. The raffle officials need to decide whether to have several winners of small amounts or just a few winners of large amounts. One option is to have two prizes worth \$10,000 each and two prizes worth \$5000 each. A second option is to have two prizes worth \$15,000 apiece. In which option are the results more spread out? (Consider any non-winning ticket as a \$0 prize.)

## APPLYING THE MATHEMATICS

12. **Multiple Choice** A store has two managers and nine employees. Each manager earns \$40,000 a year, six employees earn \$25,000 a year, and three employees earn \$15,000 a year. If each person gets a \$1000 raise next year, the standard deviation of the salaries
    A will increase by \$1000.
    B will increase by \$3000.
    C will not change.
    D will increase by about 3%.

13. a. Let $x_i = 2i$, for $i = 1, 2, 3, \dots, 10$. Find the mean and standard deviation of the $x_i$ values.
    b. Let $y_i = 2i + 1$, for $i = 1, 2, 3, \dots, 10$. Find the mean and standard deviation of the $y_i$ values.

14. Below are the ages of the Democratic and Republican United States Presidents when they were first inaugurated into office, as of 2008.

    Democrats: 43, 46, 47, 48, 49, 51, 52, 54, 55, 55, 56, 60, 61, 65

    Republicans: 42, 46, 49, 50, 51, 51, 51, 52, 54, 54, 54, 55, 55, 56, 61, 62, 64, 69

    Compare the ages at inauguration of Democrats and Republicans by calculating means and standard deviations.

15. Give an example, different from the one in the lesson, of two different data sets that have the same mean but different standard deviations.

While John F. Kennedy was the youngest person elected president, Theodore Roosevelt was the youngest person to become president when William McKinley was assassinated.

16. What would a data set with standard deviation equal to 0 look like?

## REVIEW

17. Lotta Moola invests \$350 on the first day of every month in an account that earns an annual interest rate of 6% compounded monthly. Assume no other deposits or withdrawals are made.
    a. How much interest will the first \$350 deposit earn in 6 months?
    b. How much will be in Lotta's account just after she makes her 7th deposit? **(Lessons 13-2, 11-1, 7-4)**

**In 18 and 19, suppose a tennis ball is released from a height of 1 meter above the floor. Each time it hits the floor it bounces to 40% of its previous height. (Lessons 13-2, 7-5)**

18. Suppose the ball has hit the floor four times. How high will it get on the next bounce?
19. If the ball hits the floor eight times, find the vertical distance it will have traveled.
20. Beginning with 1, how many consecutive positive integers do you have to add in order to total 2701? **(Lesson 13-1)**

**In 21–23, evaluate and write your answer in $a + bi$ form. (Lessons 6-9, 6-8)**

21. $(1 + i)^2$
22. $\frac{-8 + 2i}{i}$
23. $i^4 + i^5 + i^6 + i^7$

24. If 4 thingies and 3 somethings weigh 190 lb, and 6 thingies and 7 somethings weigh 350 lb, what will 2 thingies and 4 somethings weigh? **(Lesson 5-4)**
25. Ivan has test scores of 80, 97, 90, and 88. What must Ivan score on the next test to have
    a. a mean of 90 for the five tests?
    b. a median of 90 for the five tests?
    c. a mode of 90 for the five tests? **(Previous Course)**

## EXPLORATION

26. You have used your calculator's statistical regression functions to find equations to model sets of data points. To do this, the regression procedure minimizes the sum of the squares of the deviations of the points from the curve. Search the Internet (for example, search for "sum of squares applet") to find interactive websites that allow you to graphically explore how to minimize a sum of squares.

**QY ANSWERS**

1. a. –3.6, 4.4, –6.6, –1.6, –1.6, –0.6, –5.6, 1.4, 0.4, 13.4
   b. 0
2. a. 3.6, 4.4, 6.6, 1.6, 1.6, 0.6, 5.6, 1.4, 0.4, 13.4
   b. 3.92