Session 5

Variation About the Mean

Key Terms for This Session

Previously Introduced

- line plot
- median
- variation

New in This Session

- allocation
- deviation from the mean
- fair allocation (equal-shares allocation)
- mean
- mean absolute deviation (MAD)
- standard deviation
- variance

Introduction

In Session 4, we explored the Five-Number Summary and its graphical representation, the box plot. We also explored the median, a common numerical summary for a data set.

In this session, we'll investigate another common numerical summary, the mean. We'll also learn several ways to describe the degree of variation in data, based on how much the data values vary from the mean. **[See Note 1]**

Learning Objectives

In this session, you will do the following:

- Understand the mean as an indicator of fair allocation
- Explore deviations of data values from the mean
- Understand the mean as the "balancing point" of a data set
- Learn how to measure variation about the mean

Materials Needed

You will need to have these materials in hand before you begin:

- 45 coins
- a calculator
- several pieces of blank paper or 22″ by 28″ poster board
- self-adhesive colored dots or 2″ by 1.5″ adhesive notepaper squares

Note 1. In this session, the concepts of the arithmetic mean and deviation from the mean are first introduced through a physical representation: stacks of coins. The transition is then made to a line plot as the graphical representation of these ideas.

Part A: Fair Allocations (25 min.)

The Mean

The term *average* is a popular one, often used, and often incorrectly.

Although there are different types of averages, the typical definition of the word "average" when talking about a list of numbers is "what you get when you add all the numbers and then divide by how many numbers you have." This statement describes how you calculate the arithmetic mean, or average. But knowing how to calculate a mean doesn't necessarily tell you what it *represents*.

Let's begin our exploration of the mean: Using your 45 coins, create 9 stacks of several sizes. You must use all 45 coins, and at least 1 coin must be in each of the 9 stacks. It's fine to have the same number of coins in multiple stacks.

Here is one possible arrangement, or allocation, of the 45 coins:

Problem A1. Record the number of coins in each of your 9 stacks. What is the mean number of coins in the 9 stacks? **[See Tip A1, page 155]**

Problem A2. Create a second allocation of the 45 coins into 9 stacks.

a. Record the number of coins in each of your 9 stacks, and determine the mean for this new allocation.

b. Why is the mean of this allocation equal to the mean of the first allocation?

c. Describe two things that you could do to this allocation that would change the mean number of coins in the stacks.

Problem A3. Create a third allocation of the 45 coins into 9 stacks in a special way:

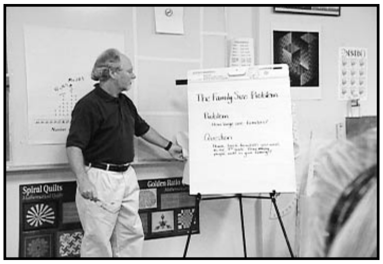
- First take a coin from the pile of 45 and put it in the first stack.
- Then take another coin from the pile and put it in the second stack.
- Continue in this way until you have 9 stacks with 1 coin each, and 36 coins remaining.
- Take a coin from the pile of 36 and put it in the first stack.
- Then take another coin from the pile and put it in the second stack.
- Continue in this fashion until all of the remaining coins have been used.

  a. Now record the number of coins in each of your 9 stacks, and determine the mean for this new allocation.

  b. What observations can you make about the mean in this special allocation?

This method produces what is called a fair or equal-shares allocation. Each stack, in fact, contains the *average* number of coins. You might think of this as a fair allocation of the 45 coins among 9 people: Each person gets the same number of coins.

Part A, cont'd.



Video Segment (approximate times: 2:37-3:20 and 8:20-10:42): You can find the first part of this segment on the session video approximately 2 minutes and 37 seconds after the Annenberg/CPB logo. Use the video image to locate where to begin viewing. The second part of this segment begins approximately 8 minutes and 20 seconds after the Annenberg/CPB logo.

In this video segment, Professor Kader asks participants to create snap-cube representations of the number of people in their families. He then asks them to find a way of finding the mean without using calculation. Watch this segment for an exploration of the mean.

How does the mean relate to the fair allocation of the data?

The Mean and the Median

Let's now contrast the mean and the median as summary measures. You previously found that when 45 coins were allocated to 9 stacks, the mean stack size was always 5 because the number of stacks and the sum of the stack sizes was a constant, 9 and 45 respectively. **[See Note 2]**

Write and Reflect

Problem A4. Do you expect that the median stack size for the 9 stacks will always be the same for any allocation? Why or why not? You might want to look back at the allocations you created for Problems A1-A3.

Problem A5. Put your 45 coins into this allocation:

a. Why is the median *not* the fifth stack in the allocation at right?

b. Arrange your stacks in ascending order from left to right. Then find the median stack size using one of the methods you learned in Session 4, Part B.

[See Tip A5, page 155]

Problem A6. Create a new allocation of the 45 coins into 9 stacks so that the median is equal to 5. (Do not use the allocation with 5 coins in each stack.)

Problem A7.

a. Create a new allocation of the 45 coins into 9 stacks so that the median is *not* equal to 5.

b. What is the mean for your new allocation?

[See Tip A7, page 155]

Problem A8. Find a third allocation that has a median different from the ones in Problems A6 and A7.

Take It Further

Problem A9. What is the smallest possible value for the median? What is the largest possible value for the median? Remember that there must be 9 stacks for the 45 coins, and each stack must contain at least 1 coin.

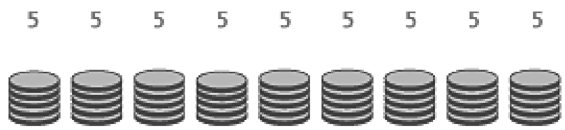
Note 2. In Problems A4-A8, you examine the differences between various allocations of values. In particular, you focus on the median in each data set and its relation to the rest of the data. While the median is not the main focus of this session, it is important to emphasize the difference between the mean and the median, a distinction that is not always clear.

Part B: Unfair Allocations (25 min.)

Fair and Unfair Allocations

The average for a set of data corresponds to the fair allocation or equal-shares allocation of the data. For example, suppose that each of 9 people has several dollars and altogether they have $45. The mean of $5 represents the number of dollars each of the 9 people would get if they combined all their money and then redistributed it fairly (i.e., equally).

As seen in Problem A3, the fair allocation of 45 coins into 9 stacks is to place 5 coins in each stack, as follows:



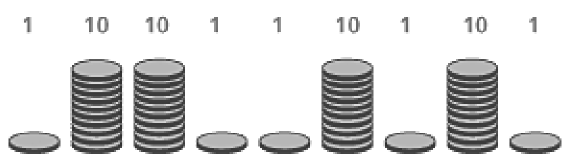
Here is a second allocation of the 45 coins, which is almost fair:

The above allocation is almost fair because most stacks have 5 coins, and the others have close to 5. But the following allocation of 45 coins doesn't seem fair at all:

Problem B1. Look at these five allocations:

Allocation A

Allocation B



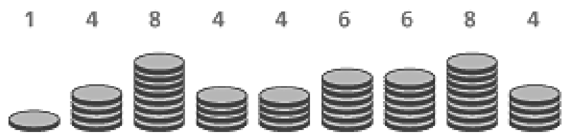
Allocation C

Part B, cont'd.

Problem B1, cont'd.

Allocation D

Allocation E



Rank the five allocations on their "fairness," from most fair to least fair. Also, explain how you decided on the level of "unfairness" in an allocation. **[See Tip B1, page 155]**

Measuring the Degree of Fairness

In Problem B1, it's easy to identify the most and least fair allocations. It's more difficult to decide the degree of fairness among the remaining three allocations. **[See Note 3]**

We need a more objective way to measure how close an allocation is to the fair allocation. Here's one method:

- Arrange the stacks in the allocation in increasing order.
- Move 1 coin from an above-average stack to a below-average stack. (Here, "average" refers to the mean.) This move will create a new allocation of the coins with the same mean.
- Continue moving coins, one at a time, from an above-average stack to a below-average stack until (if possible) each stack contains the same number of coins.
- One measure of the degree of fairness for an allocation is the number of coins you had to move. The smaller this number is, the closer the allocation is to the fair allocation.

Problem B2. Below are three allocations of 45 coins in 9 stacks. For each allocation, find the minimum number of moves required to change the allocation into a fair allocation (i.e., one with 5 coins in each stack).

Allocation A

Note 3. In this part, you consider the fairness of different allocations: Which allocation is the most fair? Which allocation is the least fair?

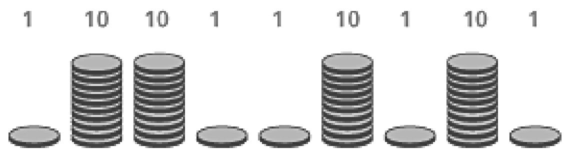
These questions ask for more than a distinction between fair and unfair—they suggest that there are degrees of unfairness. Come up with your own method for measuring degrees of fairness. Part of the point of this lesson is to realize that statistical measures are invented by people.

If you are working in a group, discuss several methods for measuring fairness.

Part B, cont'd.

Problem B2, cont'd.

Allocation B



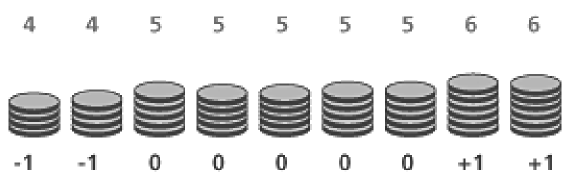
Allocation C

Problem B3. Based on the minimum number of moves, which of the allocations in Problem B2 is the most fair? Which is the least fair?

Looking at Excesses and Deficits

We've seen how to judge the relative fairness of an allocation. A related question asks how you might determine the number of moves required to make an unfair allocation fair, without making the actual moves. **[See Note 4]**

Below is Allocation A from Problem B2, arranged in ascending order. We've already determined that only two moves are required to make this allocation fair. Let's look more closely at why this is true:



The notations below each stack indicate the number of coins that each stack is above or below the mean of 5. In other words:

a. Two stacks are above the mean. Each of these has 6 coins, an excess of 1 coin (+1) above the average. (These excesses are noted in the figure above.) The total excess of coins above the average is 2.

b. Two stacks are below the mean. Each of these has 4 coins, a deficit of 1 coin (-1) below the average. (These deficits are also noted in the figure above.) The total deficit of coins below the average is 2.

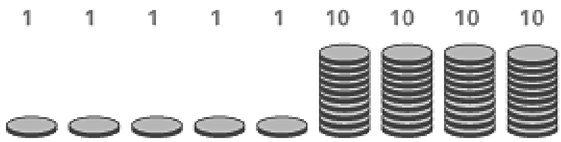
c. Five stacks are exactly average and have no excess or deficit (0).

Note 4. Excesses and deficits show how far a stack is above or below the mean. Excesses and deficits also show how many moves are required to transform an allocation to a fair allocation. Keep in mind that the total excesses must equal the total deficits. This is the defining characteristic of the mean: It is the balance point for all the deviations from the mean.

In Problems B4 and B5, you investigate an algorithm to determine the number of moves required to transform an unfair allocation to a fair one. Then you think about how the number of moves required corresponds to the fairness of the allocation. In other words, how does the number of moves measure the degree of unfairness of an allocation?

Part B, cont'd.

Problem B4. Perform the calculations on the previous page to show that 20 moves are required to obtain a fair allocation for Allocation B, which is arranged in ascending order below:



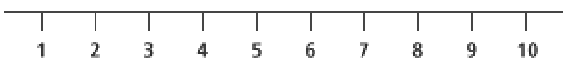
Problem B5. Perform the same calculations to find the number of moves required to obtain a fair allocation for Allocation C, which is shown in ascending order below:

Part C: Using Line Plots (30 min.)

Creating a Line Plot

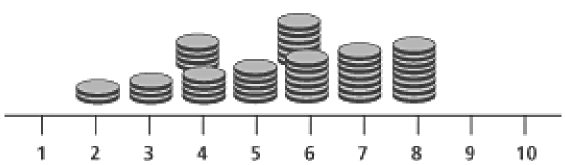
Remember the line plots you worked with in previous sessions? If you reshuffle your stacks of coins just a bit, you can create a line plot representation that corresponds to the number of coins in each of the nine stacks, which will allow us to explore other interpretations of the mean. **[See Note 5]**

To do this yourself, create a line plot on your paper or poster board. Across the bottom of the page, draw a horizontal line with 10 vertical tick marks numbered from 1 to 10 (placed far enough apart for an adhesive dot or note to fit between each). Your number line should look like this:



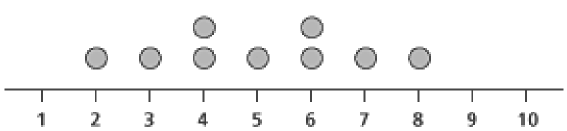
Set up your 45 coins in this ordered allocation:

Now arrange the stacks of coins on the paper above the number that corresponds to the height of each stack, like this:



Note that the 2 stacks of size 4 and the 2 stacks of size 6 are placed above the same number.

To form your line plot, replace each of the stacks with an adhesive dot or note. You should now have a line plot that looks like this:



Each dot in the line plot corresponds to a stack of that specified size.

Try It Online! www.learner.org

This problem can be explored online as an Interactive Illustration. Go to the *Data Analysis, Statistics, and Probability* Web site at www.learner.org/learningmath and find Session 5, Part C.

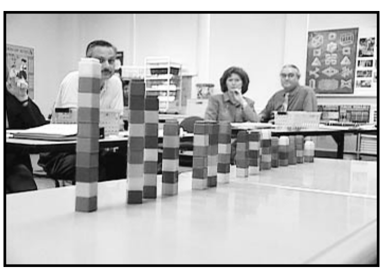
Note 5. The transition from the physical representation of coins to a graphical representation (the line plot) can be challenging. It is easy to confuse the stacks of coins with the stacks of dots in the line plot, but in fact they are very different representations of the data. Each stack of coins corresponds to one dot on the line plot, and each dot on the line plot represents a stack of coins, rather than an individual coin (which it more closely resembles).

Part C, cont'd.

Problem C1. Use the method described above to create a line plot for the following ordered allocation of 45 coins:

Problem C2. Create a line plot for this allocation of 45 coins:

Problem C3. Create a line plot for this equal-shares allocation of 45 coins:



Video Segment (approximate times: 5:44-7:09 and 8:01-8:18): You can find the first part of this segment on the session video approximately 5 minutes and 44 seconds after the Annenberg/CPB logo. Use the video image to locate where to begin viewing. The second part of this segment begins approximately 8 minutes and 1 second after the Annenberg/CPB logo.

In this video segment, participants compare their ordered stacks of snap-cubes to a line plot of the same data. Watch this segment after completing Problems C1-C3 to observe the transition from a physical representation of the data to a graphical representation.

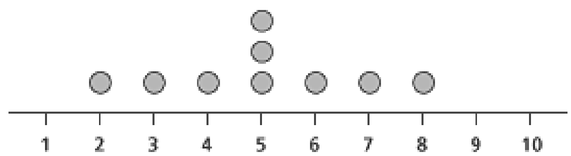
What is one stack of snap-cubes equivalent to on the line plot of the data?

Part C, cont'd.

Means From the Line Plots

In the previous examples, you explored line plot representations for sets of 45 coins, each in 9 stacks. For each allocation, the mean was 5 coins. Now let's use these line plot representations to explore another way to interpret the mean.

Problem C4. Here is a line plot corresponding to an allocation of 45 coins in 9 stacks:



From this line plot, we can see that there are 3 stacks containing exactly 5 coins each, and 1 stack containing 6 coins. The maximum number of coins in a stack is 8, and the minimum is 2.

Rearrange the nine dots to form a line plot with each of these requirements:

a. Form a different line plot with a mean equal to 5.

b. Form a line plot with a mean equal to 5 that has exactly 2 stacks of 5 coins.

c. Form a line plot with a mean equal to 5 but a median not equal to 5.

d. Form a line plot with a mean equal to 5 that has no 5-coin stacks.

e. Form a line plot with a mean equal to 5 that has two 5-coin stacks, 4 stacks with more than 5 coins, and 3 stacks with fewer than 5 coins.

f. Form a line plot with a mean equal to 5 that has two 5-coin stacks, 5 stacks with more than 5 coins, and 2 stacks with fewer than 5 coins.

g. Form a line plot with a mean equal to 5 that has two 5-coin stacks, two 10-coin stacks, and 5 stacks with fewer than 5 coins. **[See Tip C4, page 155]**

Balancing Excesses and Deficits

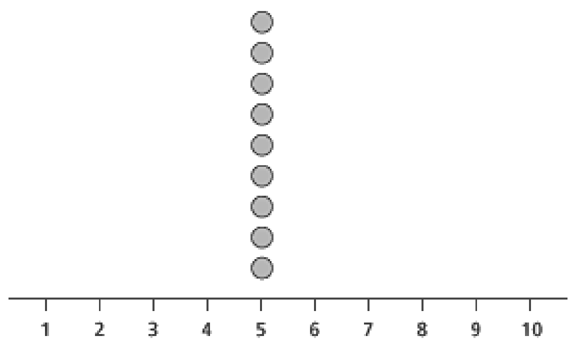
Regardless of the strategy you used in Problem C4, you must end up with an arrangement in which the sum of the 9 values is equal to 45. Let's look at one possible strategy more closely. **[See Note 6]**

Note 6. A typical method for solving Problem C4 is to keep the sum of the 9 values equal to 45. Often, people don't think of this in terms of balancing deviations. The purpose of Problems C5-C8 is to demonstrate that balancing deviations from the mean is equivalent to keeping the sum of the 9 values equal to 45.

Again, making the connections between the physical, graphical, and numerical representations can be challenging. If you are working in a group, take time to discuss how the process of making changes to the line plot corresponds to the idea of equality of excesses and deficits explored in Part C. If you are working alone, take some time to think through this concept.

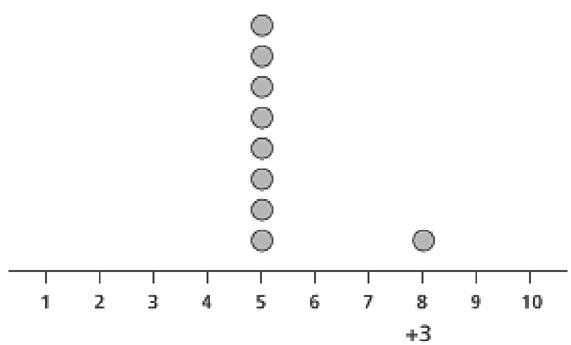
Part C, cont'd.

For the sake of simplicity, we will begin with the line plot that corresponds to the fair allocation, 9 stacks of 5 coins each:



For this line plot, the sum is 45 and the mean is 5.

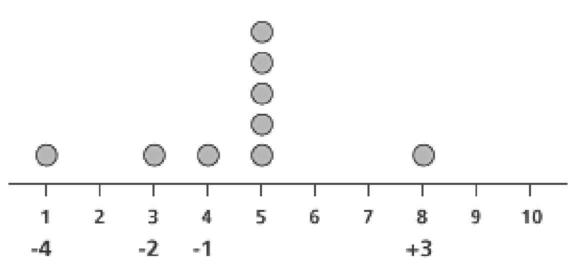
If we change one of the stacks of 5 coins to a stack of 8, the sum will increase by +3 to 48 and the mean will increase by +3/9 to 5 3/9. The line plot now looks like this:



Problem C5. How could you change another stack of 5 coins to reset the mean to 5?

Problem C6. If you could change the value of more than one stack, could you solve Problem C5 another way?

Problem C7. Now suppose that we change one of the stacks of 5 to a stack of 1, which reduces the total by 4. Here is the resulting line plot:



Describe at least three different ways to return the mean to 5.

Part C, cont'd.

Problem C8. Apply the strategy you developed in Problems C5-C7 to revisit the type of allocations you worked with in Problem C4. You should begin with the fair allocation of the 45 coins; that is, 9 dots at the mean of 5. Try to come up with answers for the questions below that are different from the ones you found in Problem C4.

a. Form a line plot with a mean equal to 5 that has exactly 2 stacks of 5 coins.

b. Form a line plot with a mean equal to 5 but a median not equal to 5.

c. Form a line plot with a mean equal to 5 that has no 5-coin stacks.

d. Form a line plot with a mean equal to 5 that has two 5-coin stacks, 4 stacks with more than 5 coins, and 3 stacks with fewer than 5 coins.

e. Form a line plot with a mean equal to 5 that has two 5-coin stacks, 5 stacks with more than 5 coins, and 2 stacks with fewer than 5 coins.

f. Form a line plot with a mean equal to 5 that has two 5-coin stacks, two 10-coin stacks, and 5 stacks with fewer than 5 coins.

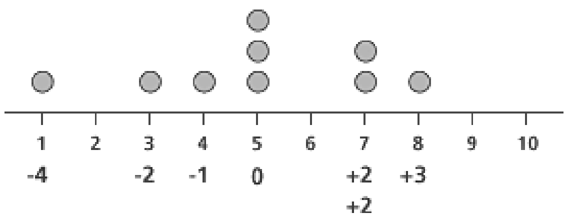
Try It Online!	**www.learner.org**
This problem can be explored online as an Interactive Activity. Go to the *Data Analysis, Statistics, and Probability* Web site at www.learner.org/learningmath and find Session 5, Part C, Problem C8.	

For a non-interactive version of this activity, use your dots and coins on paper or poster board.

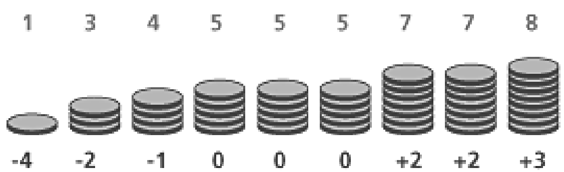
Part D: Deviations From the Mean (30 min.)

Tallying Excesses and Deficits

In Part B, we looked at excesses and deficits when we moved coins in the stacks to obtain the fair allocation. In Part C, we used a line plot to represent these excesses and deficits. We are now going to explore a new way to consider excesses and deficits. Let's look at another line plot:



For this line plot, here is the corresponding allocation of our 45 coins:



Remember that the total of the excesses above the mean must equal the total of the deficits below the mean. In this case, each adds up to 7.

If you denote the values of excesses as positive numbers and deficits as negative numbers, then the total of the excesses is:

$(+2) + (+2) + (+3) = +7$

The total of the deficits is:

$(-4) + (-2) + (-1) = -7$

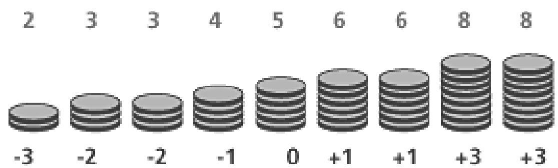
Statisticians refer to these excesses and deficits as deviations from the mean. For this allocation, the deviations from the mean are recorded in the table below.

	Number of Coins in Stack	**Deviation From the Mean**
	1	-4
	3	-2
	4	-1
	5	0
	5	0
	5	0
	7	+2
	7	+2
+	8	+3
	45	0

Note that the deviations always sum to 0 because the total excesses (positive deviations) must be the same as the total deficits (negative deviations).

Part D, cont'd.

Problem D1. Here is another allocation of our 45 coins divided into 9 stacks:



a. Draw the corresponding line plot. Indicate the deviation for each dot as was done in the line plot that opened Part D.

b. Complete the following table of deviations:

	Number of Coins in Stack	Deviation From the Mean
	2	-3
	3	
	3	
	4	
	5	
	6	
	6	
	8	
+	8	+3
	45	0

Line Plot Representations

Problem D2. Create a line plot with these deviations from the mean = 5:

(-4), (-3), (-2), (-1), (0), (+1), (+2), (+3), (+4)

Try It Online! **www.learner.org**

This problem can be explored online as an Interactive Activity. Go to the *Data Analysis, Statistics, and Probability* Web site at www.learner.org/learningmath and find Session 5, Part D, Problem D2.

Problem D3. Create a line plot with these deviations from the mean = 5:

(-4), (-2), (-2), (-1), (0), (+1), (+2), (+2), (+4)

Problem D4. Create a line plot with these deviations from the mean = 5, and specify a set of four remaining values:

(-4), (-3), (-3), (-1), (-1)

Problem D5. How would the line plots you created in Problems D2-D4 change if you were told that the mean was 6 instead of 5? Would this change the degree of fairness of these allocations (as described in Problem B2)?

When the positive and negative deviations are added together, the total is always 0. This property illustrates another way to interpret the mean: The mean is the balance point of the distribution when represented in a line plot, since the total deviation above the mean must equal the total deviation below the mean.

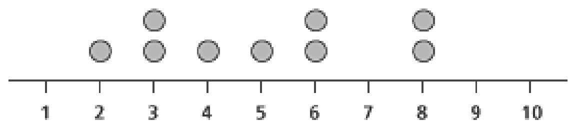
Part E: Measuring Variation (45 min.)

Mean Absolute Deviation (MAD)

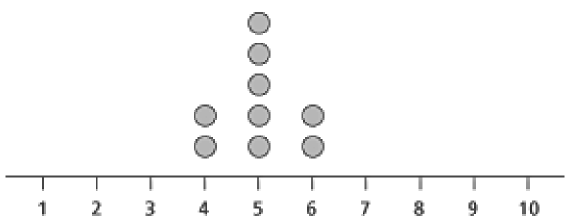
We will now focus on how to measure variation from the mean within a data set. There are several different ways to do this. The first measure we will explore is called the mean absolute deviation, or MAD. **[See Note 7]**

Problem E1. Three line plots are pictured below; each has 9 values, and the mean of each is 5:

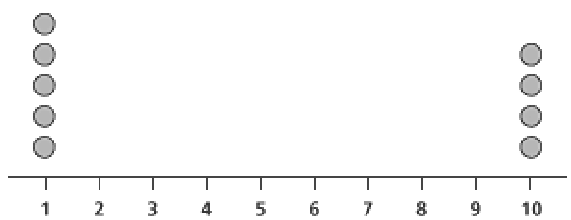
Line Plot A



Line Plot B



Line Plot C



Of the three, which line plot's data has the least variation from the mean? Which has the most variation from the mean? **[See Tip E1, page 155]**

From working on Problem E1, you probably got an intuitive sense of the variation in the three data sets. But is there a way to measure exactly how much the values in a line plot differ from the mean?

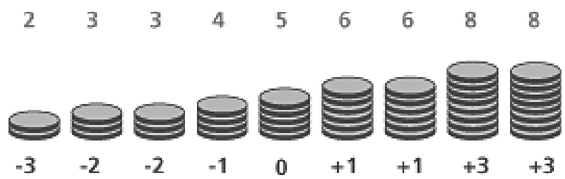
Recall that in Part B, we described the unfairness of allocations of coins by counting the number of moves required to transform an ordered allocation into an equal-shares allocation. The idea of variation from the mean is related to the idea of fairness in the coin allocation.

Note 7. Part E investigates how deviations from the mean can be used to develop a summary measure of the degree of variation in your data. Is there a single number that can describe how much the values in a line plot differ from the mean?

The line plot representation helps us develop an understanding of the mean absolute deviation (MAD). As you become more familiar with the MAD, take some time to think about how this numerical measure relates to your intuitive sense of the degree of variation in the line plots with which you're working.

Part E, cont'd.

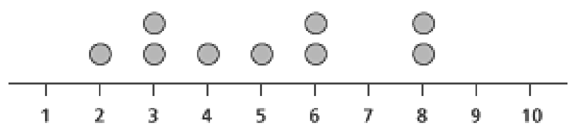
For example, consider this ordered allocation:



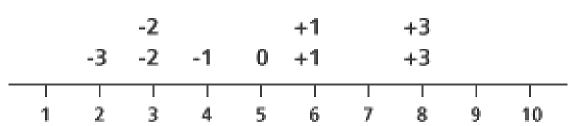
Eight moves are required to make this allocation fair. This is true because there is an excess of 8 coins above the mean from 4 stacks (+1, +1, +3, +3), and a deficit of 8 coins below the mean from 4 other stacks (-3, -2, -2, -1).

The number of moves required to make an allocation fair tells us how much the original allocation differs from the fair allocation and thus gives us a measure of the variation in our data. (The fair allocation has no variation—no moves = no variation.)

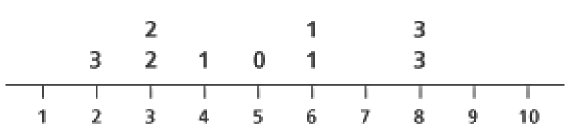
Here is the line plot that corresponds to this allocation:



Here are the deviations from the mean for each value in the set (i.e., how much each value differs from the mean):



Now consider only the magnitude of these deviations—that is, forget for the moment whether they are positive or negative. These are called the absolute deviations. The absolute deviations for this set are plotted below:

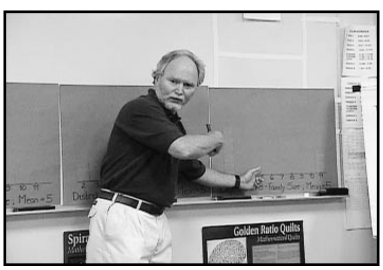


Part E, cont'd.

We are now going to find the mean of these absolute deviations, which is an indicator, on average, of how far (what distance) the values in our data are from the mean. As usual, find the mean by adding all the absolute deviations and then dividing by how many there are. Here is a table for this calculation:

Number of Coins in Stack (x)	Deviation From the Mean (x - 5)	Absolute Deviation From the Mean |x - 5|
2	-3	3
3	-2	2
3	-2	2
4	-1	1
5	0	0
6	+1	1
6	+1	1
8	+3	3
8	+3	3
45	0	16

The mean of these absolute deviations—the MAD (Mean Absolute Deviation)—is 16 / 9 = 1 7/9, or approximately 1.78. This measure tells us how much, on average, the values in a line plot differ from the mean. If the MAD is small, it tells us that the values in the set are clustered closely around the mean. If it is large, we know that at least some values are quite far away from the mean.

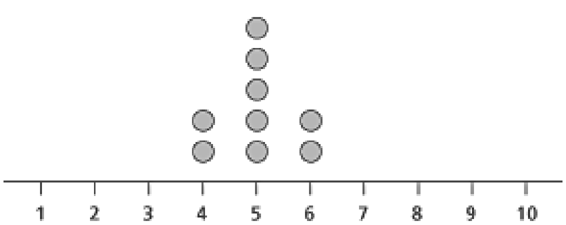


Video Segment (approximate times: 12:08-13:42 and 16:52-17:55): You can find the first part of this segment on the session video approximately 12 minutes and 8 seconds after the Annenberg/CPB logo. Use the video image to locate where to begin viewing. The second part of this segment begins approximately 16 minutes and 52 seconds after the Annenberg/CPB logo.

In this video segment, Professor Kader introduces the MAD as a method for quantifying variation. Watch this segment to review the process of finding the MAD.

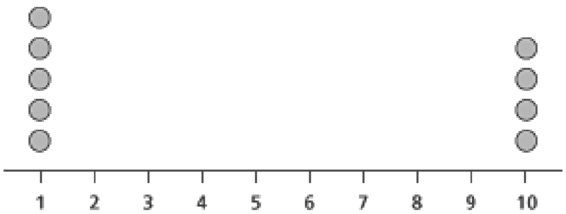
How can the MAD be used to compare different distributions of data?

Problem E2. Below is Line Plot B from Problem E1. Create a table like the one above, find the MAD for this allocation, and compare it to the MAD of Line Plot A from the same problem.



Part E, cont'd.

Problem E3. Below is Line Plot C from Problem E1. Create another table, find the MAD for this allocation, and compare it to the MADs of Line Plots A and B.



Working With the MAD

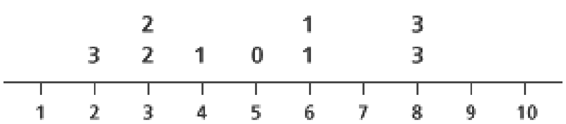
Try It Online! **www.learner.org**

The following problems can be explored online as an Interactive Activity. Go to the *Data Analysis, Statistics, and Probability* Web site at www.learner.org/learningmath and find Session 5, Part D, Problems E4-E6.

For a non-interactive version of this activity, use your paper/poster board to answer the questions in Problems E4-E6.

You will be asked to form several arrangements with a specified total for the absolute deviations. You might want to write the size of each absolute deviation on each adhesive dot or note to help you determine whether you have the desired result. **[See Note 8]**

For example, from the line plot shown on page 148, this version of the line plot was:



Note 8. The idea of variation from the mean is related to the idea of fairness in an allocation of coins, which you discussed earlier in this session. Think back to the method of determining the unfairness of an allocation—counting the number of moves required to transform an ordered allocation to an equal-shares allocation. Here's the connection: The sum of the absolute deviations is equal to twice this required number of moves.

The absolute deviations occur in pairs, since the mean is the "balance point" for the set. Half the absolute deviations are above the mean, and half are below. When a move is made, one coin is moved from a value above the mean to a value below the mean. This removes two of the deviations; the deviation above the mean is reduced, and the deviation below the mean is reduced. Since each move reduces the absolute deviation by two, the sum of the absolute deviations must be twice the required number of moves.

Part E, cont'd.

In the following problems, the mean is 5, and there are 9 values in the data set. See if you can find more than one arrangement for each description.

Problem E4. Create a line plot with a MAD of 24 / 9.

Problem E5. Create a line plot with a MAD of 22 / 9, with no 5s.

Problem E6. Create a line plot with a MAD of 12 / 9, with exactly two 5s, 5 values larger than 5, and 2 values smaller than 5.

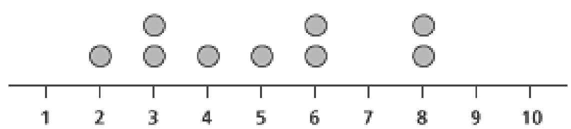
Take It Further

Problem E7. Explain why it is *not* possible to create a line plot with 9 values that has a MAD of 1.

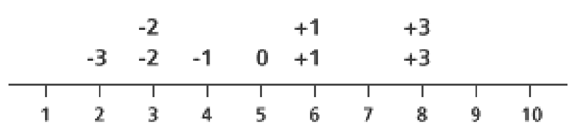
Variance and Standard Deviation

The MAD is a measure of the variation in a data set about the mean. Professional statisticians more commonly use two other measures of variation: the variance and the standard deviation.

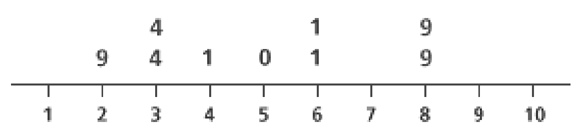
The method for calculating variance is very similar to the method you just used to calculate the MAD. First, let's go back to Line Plot A from Problem E1:



The first step in calculating the variance is the same one you used to find the MAD: Find the deviation for each value in the set (i.e., how much each value differs from the mean). The deviations for this data set are plotted below:



The next step in calculating the variance is to square each deviation. Note the difference between this and the MAD, which requires us to find the absolute value of each deviation. The squares of the deviations for this data set are plotted below:



Part E, cont'd.

The final step is to find the variance by calculating the mean of the squares. As usual, find the mean by adding all the values and then dividing by how many there are. Here is a table for this calculation:

Number of Coins in Stack (x)	Deviation From the Mean (x – 5)	Squared Deviation From the Mean $(x - 5)^2$
2	-3	9
3	-2	4
3	-2	4
4	-1	1
5	0	0
6	+1	1
6	+1	1
8	+3	9
8	+3	9
45	0	38

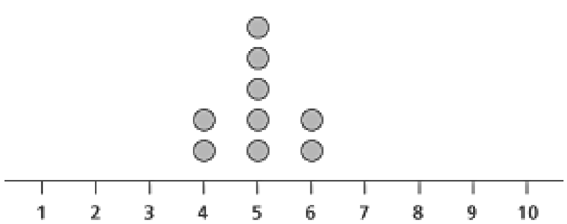
The mean of the squared deviations is 38 / 9 = 4 2/9, or approximately 4.22. This value is the variance for this data set. As with the MAD, the variance is a measure of variation about the mean. Data sets with more variation will have a higher variance.

The variance is the mean of the squared deviations, so you could also say that it represents the average of the squared deviations. The problem with using the variance as a measure of variation is that it is in squared units. To gauge a typical (or standard) deviation, we would need to calculate the square root of the variance. This measure—the square root of the variance—is called the standard deviation for a data set.

For the data set given above, the standard deviation is the square root of 4.22, which is approximately 2.05. Note that this value is fairly close to the MAD of 1.78 that we calculated earlier.

The standard deviation, first introduced in the late 19th century, has become the most frequently used measure of variation in statistics today. For example, the SAT is scaled so that its mean is 500 points and its standard deviation is 100 points. IQ tests are created with an expected mean of 100 and a standard deviation of 15.

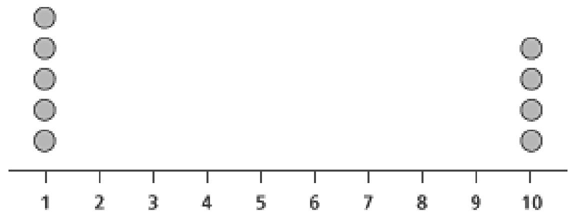
Problem E8. Below is Line Plot B from Problem E1. Create a table like the one above, and find the variance and standard deviation for this allocation. Compare the standard deviation to the MAD of Line Plot B you found in Problem E2 and to the standard deviation of Line Plot A.



[See Tip E8, page 155]

Part E, cont'd.

Problem E9. Below is Line Plot C from Problem E1. Create another table, and find the variance and standard deviation for this allocation. Compare the standard deviation to the MAD of Line Plot C you found in Problem E3 and to the standard deviations of Line Plots A and B.



[See Tip E9, page 155]

Take It Further

Problem E10

a. What would happen to the mean of a data set if you added 3 to every number in it?

b. What would happen to the MAD of a data set if you added 3 to every number in it?

c. What would happen to the variance of a data set if you added 3 to every number in it?

d. What would happen to the standard deviation of a data set if you added 3 to every number in it?

e. What would happen to the mean of a data set if you doubled every number in it?

f. What would happen to the MAD of a data set if you doubled every number in it?

g. What would happen to the variance of a data set if you doubled every number in it?

h. What would happen to the standard deviation of a data set if you doubled every number in it?

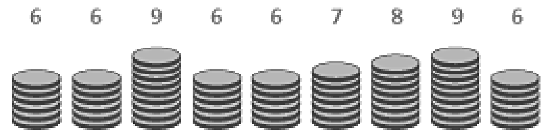
Video Segment (approximate times: 21:22-24:18): You can find this segment on the session video approximately 21 minutes and 22 seconds after the Annenberg/CPB logo. Use the video image to locate where to begin viewing.

In this video segment, Andrea Rex, director of the Massachusetts Water Resources Authority, discusses the use of statistics in assessing water quality in Boston Harbor. Watch this segment for a real-world application of the mean and standard deviation.

How does Andrea Rex use the mean and standard deviation to assess water quality?

Homework

Problem H1. For the allocation below, find and interpret the MAD, variance, and standard deviation. **[See Tip H1, page 155]**



Problem H2. For the allocation below, find and interpret the MAD, variance, and standard deviation. (Note that in this problem the mean is not a whole number, so there will be a fractional deviation for each value. To simplify your work, round off each calculation to one decimal place.) What do these calculations tell you about the data in this problem as compared to the data in Problem H1?

Suggested Readings

These readings are available as downloadable PDF files on the *Data Analysis, Statistics, and Probability* Web site. Go to:

www.learner.org/learningmath

Kader, Gary (March, 1999). Means and MADS. *Mathematics Teaching in the Middle School, 4* (6), 398-403.

Uccellini, John C. (November-December, 1996). Teaching the Mean Meaningfully. *Mathematics Teaching in the Middle School, 2* (3), 112-115.

Zawojewski, Judith and Shaughnessy, Michael (March, 2000). Mean and Median: Are They Really So Easy? *Mathematics Teaching in the Middle School, 5* (7), 436-440.

Tips

Part A: Fair Allocations

Tip A1. To find the mean, divide the number of coins (45) by the number of stacks.

Tip A5. The median is the value in the center of an *ordered* data set. Since there are 9 stacks, the median is in the position (9 + 1) / 2; that is, the median is in position (5) after ordering.

Tip A7. As you manipulate the coins, remember that the median can only be found in an ordered data set, so keep your 9 stacks in ascending order. Try to manipulate the stacks so that the fifth stack does not contain 5 coins.

Part B: Unfair Allocations

Tip B1. Although there are many different ways to think about "unfairness," one way is to consider the number of exchanges that would have to take place in order to achieve a fair allocation.

Part C: Using Line Plots

Tip C4. Don't forget that the mean must always be equal to 5. If you move a dot to the right, it will increase the mean. Each time you move a dot to the right, you must balance this by moving another dot an equal distance to the left. Also, keep in mind that each dot represents a stack of coins, and that by moving the position of the dot, you change the number of coins in the stack. The total number of coins must remain 45.

Part E: Measuring Variation

Tip E1. The plot with the most variation will have data values that are, in general, farthest away from the mean. The plot with the least variation will generally have the values closest to the mean.

Tip E8. Remember that in these problems, the mean is 5. Calculate the variance first, then take its square root to find the standard deviation.

Tip E9. Again, remember to calculate the variance first, then take its square root.

Homework

Tip H1. You will need to find the mean first in order to calculate the deviation for each value in the set.

Solutions

Part A: Fair Allocations

Problem A1. The mean is 5. To calculate the mean, add the numbers in each stack. The sum is 45. Then divide by the total number of stacks, 9, to find the mean.

Problem A2.

a. The mean is still 5.

b. Since the total number of coins remains 45, and the total number of stacks remains 9, the mean must still be 45 / 9 = 5.

c. You could increase or decrease the number of coins, or you could increase or decrease the number of stacks.

Problem A3.

a. As before, the mean is 5.

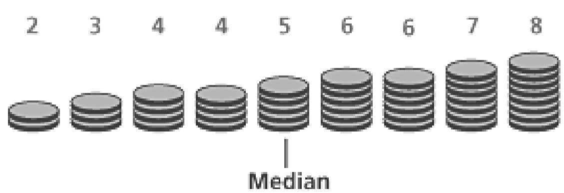
b. Here, the mean is equal to the number of coins in each stack. Since all the stacks have the same number of coins, the average is equal to the number of coins in any given stack.

Problem A4. Write and reflect.

Problem A5.

a. The median is not the fifth stack because the allocation is not in order. It must first be placed in order.

b. Here's the ordered allocation:



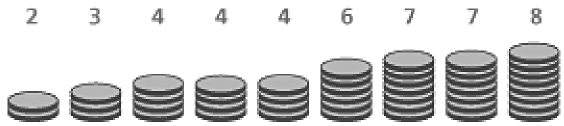
The stack in position (5), the middle, has 5 coins, so the median size is 5. For this allocation, the median is the same as the mean.

Problem A6. Answers will vary. Here is one possible allocation:

Solutions, cont'd.

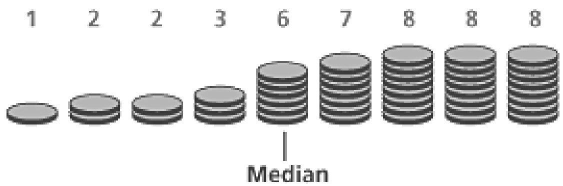
Problem A7.

a. Answers will vary. Here is one such allocation, with a median size of 4 coins:



b. As before, the mean must remain 5, since the number of coins and the number of stacks have not changed.

Problem A8. Answers will vary. Here is one such allocation, with a median size of 6 coins:



Problem A9. The smallest possible value for the median is 1. It cannot be smaller, since each stack must contain at least 1 coin. One possible allocation with this median is {1, 1, 1, 1, 1, 10, 10, 10, 10}; many other allocations are possible if a stack may contain more than 10 coins.

The largest possible value for the median is 8. One way to think about this problem is to recognize that there must be 4 stacks below the median, and each must have at least 1 coin. This leaves 41 or fewer coins for the remaining five stacks, and the median must be the smallest of these stacks. Using an equal-shares allocation of the remaining 41 coins results in 4 stacks of 8 coins and 1 of 9, so the median cannot be greater than 8. There are only two possible allocations with this median: {1, 1, 1, 1, 8, 8, 8, 8, 9} and {1, 1, 1, 2, 8, 8, 8, 8, 8}.

Part B: Unfair Allocations

Problem B1. Selecting the most fair allocation should be fairly easy. Most people immediately pick Allocation A because it has 5 stacks of size 5, and the other 4 stacks are of size 4 or 6. It is the most similar to the fair allocation.

Selecting the least fair allocation is also relatively easy. Most people immediately pick Allocation B as the least fair. Its stacks are only of size 1 or 10, which is quite different from the fair allocation of 5 for each stack.

Otherwise, answers will vary depending on how people measure "fairness." One method is to find the allocation where, on average, the stack sizes are closest to 5. This is accomplished by finding, in total, how far away the stacks are from 5. Using this criterion, Allocation A is the most fair, followed by Allocations D, C, E, and B.

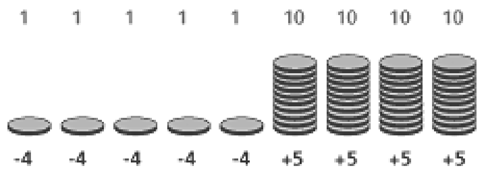
Problem B2.

- Allocation A requires 2 moves.
- Allocation B requires 20 moves.
- Allocation C requires 7 moves.

Problem B3. Allocation A is the most fair, and Allocation B is the least fair. The number of moves required to make an allocation fair is one way to measure how close to being fair the allocation is. That is, the more moves required to make an allocation fair, the more unfair the allocation is.

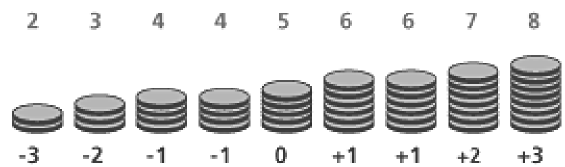
Solutions, cont'd.

Problem B4.



- Four stacks are above the mean. Each of these has 10 coins, an excess of 5 coins (+5) above the average. The total excess of coins above the average is 20.
- Five stacks are below the mean. Each of these has 1 coin, a deficit of 4 coins (-4) below the average. The total deficit of coins below the average is 20.
- No stacks are exactly average.

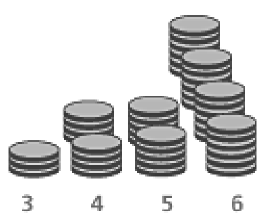
Problem B5.



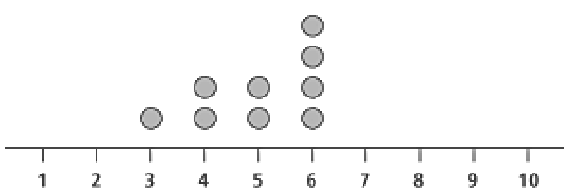
- Four stacks are above the mean. The total excess of coins above the average is 7 ([+1] + [+1] + [+2] + [+3]).
- Four stacks are below the mean. The total deficit of coins below the average is 7 ([-3] + [-2] + [-1] + [-1]).
- One stack is exactly average and has no excess or deficit (0).

Part C: Using Line Plots

Problem C1. The arrangement of the coins should look like this:



Here is the corresponding line plot:

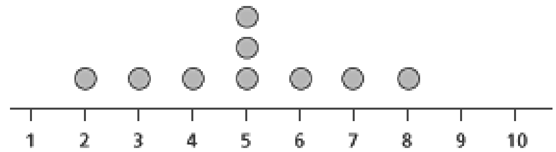


Note that there are nine dots, one for each stack, which are placed in the line plot according to the number of coins in each stack.

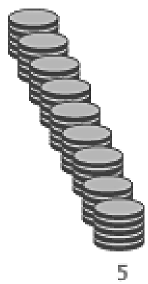
Solutions, cont'd.

Problem C2. The arrangement of the coins should look like this:

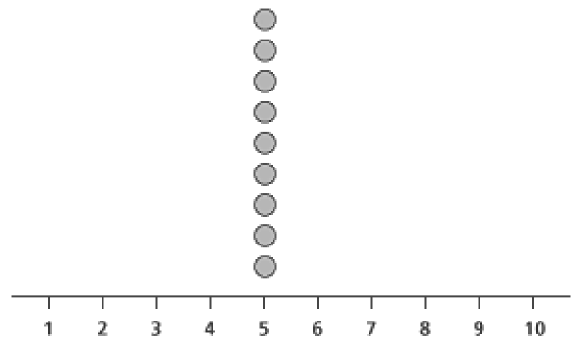
Here is the corresponding line plot:



Problem C3. The arrangement of the coins should look like this:

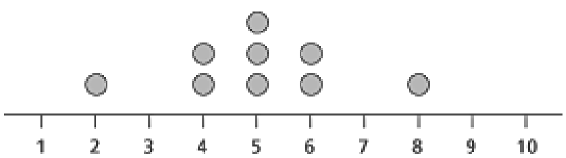


Here is the corresponding line plot:

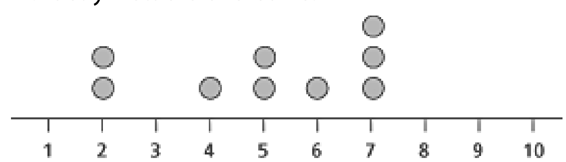


Problem C4. There are multiple solutions to each of these problems, so answers will vary. Here are some possible solutions:

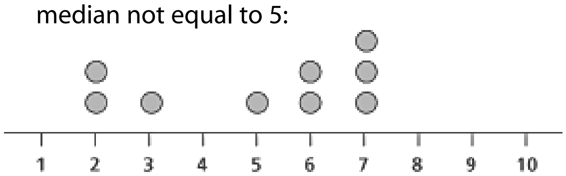
a. Here is a different line plot with a mean equal to 5:



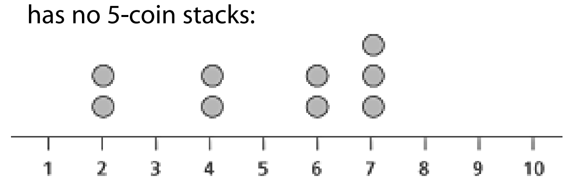
b. Here is a line plot with a mean equal to 5 that has exactly 2 stacks of 5 coins:



c. Here is a line plot with a mean equal to 5 but a median not equal to 5:



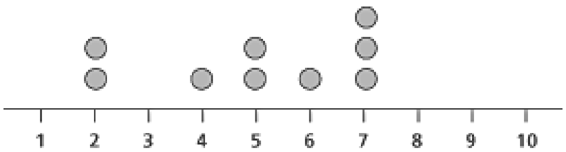
d. Here is a line plot with a mean equal to 5 that has no 5-coin stacks:



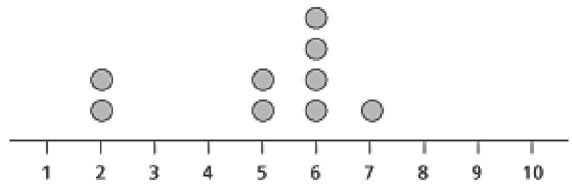
Solutions, cont'd.

Problem C4, cont'd.

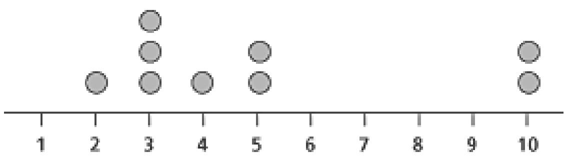
e. Here is a line plot with a mean equal to 5 that has two 5-coin stacks, 4 stacks with more than 5 coins, and 3 stacks with fewer than 5 coins:



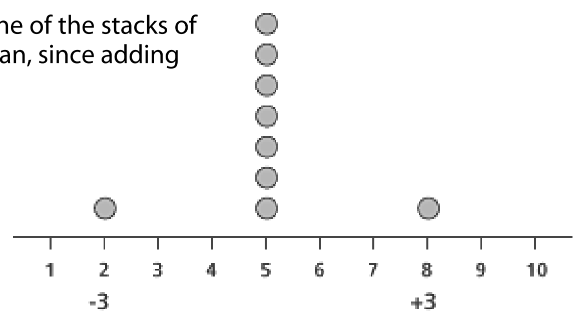
f. Here is a line plot with a mean equal to 5 that has two 5-coin stacks, 5 stacks with more than 5 coins, and 2 stacks with fewer than 5 coins:



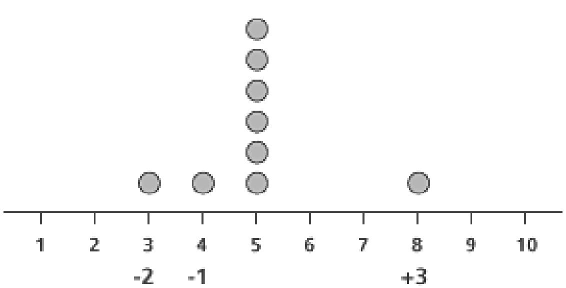
g. Here is a line plot with a mean equal to 5 that has two 5-coin stacks, two 10-coin stacks, and 5 stacks with fewer than 5 coins:



Problem C5. You would need to remove 3 coins from one of the stacks of 5, resulting in a stack of 2 coins. This would reset the mean, since adding 3 coins to a stack changed the mean.



Problem C6. Yes, you could remove 1 coin from one stack and 2 coins from another.



Alternately, you could remove 1 coin from each of three stacks. You could even add 1 coin to a stack and remove 4 from another. The only requirement is that you subtract a total of 3 coins from the total number of coins (to counterbalance the 3 coins you added previously).

Solutions, cont'd.

Problem C7. Answers will vary. Here are some possible solutions:

- You could add 4 coins to the stack of 3 to create a stack of 7 coins.
- You could add 2 coins to each of 2 stacks of 5 to create 2 stacks of 7 coins each.
- You could add 1 coin apiece to 4 of the stacks of 5 to create 4 stacks of 6 coins each.

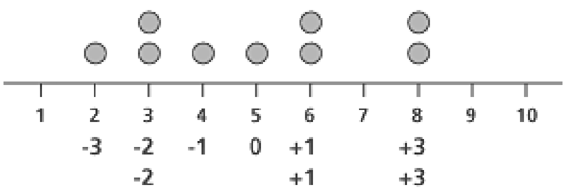
The only requirement is that you must add 4 coins to the total number of coins to return the sum to 45, which will return the mean to 45 / 9 = 5.

Problem C8. As with Problem C4, answers will vary. See Problem C4 for some sample solutions.

Part D: Deviations From the Mean

Problem D1.

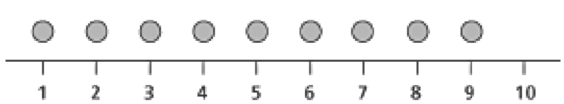
Here is the corresponding line plot, with the deviation for each dot included under the line:



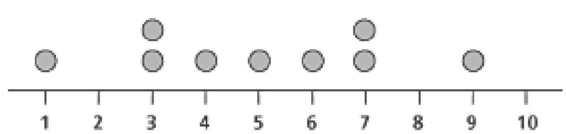
Here is the completed table:

	Number of Coins in Stack	Deviation From the Mean
	2	-3
	3	-2
	3	-2
	4	-1
	5	0
	6	+1
	6	+1
	8	+3
+	8	+3
	45	0

Problem D2. Here is the line plot:

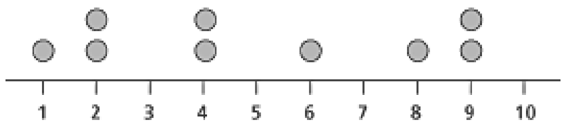


Problem D3. Here is the line plot:



Solutions, cont'd.

Problem D4. The total of the deficits is (-4) + (-3) + (-3) + (-1) + (-1) = -12, so the remaining 4 stacks will need a total excess of +12 to maintain the balance required by the mean. One possible answer is for the remaining 4 values to be (+1), (+3), (+4), (+4), although many other answers are also possible, which corresponds to the following line plot:



Problem D5. The line plots would all move one unit to the right. This would not change the degree of fairness, since the degree of fairness relies on the deviations from the mean. Since the mean will move right along with the values in the line plot, the deviations will not change, and likewise the degree of fairness will not change.

Part E: Measuring Variation

Problem E1. By inspection, it is clear that Line Plot B has the least variation and Line Plot C has the most variation. Remember that the variation is *from the mean*, not between the values themselves, which is why Line Plot C has the most variation.

Problem E2.

Number of Coins in Stack (x)	Deviation From the Mean (x - 5)	Absolute Deviation From the Mean |x - 5|
4	-1	1
4	-1	1
5	0	0
5	0	0
5	0	0
5	0	0
5	0	0
6	+1	1
6	+1	1
45	0	4

The MAD is 4 / 9, or approximately 0.44. This is much smaller than the MAD for Line Plot A, which indicates that the values in this allocation are much more closely grouped around the mean.

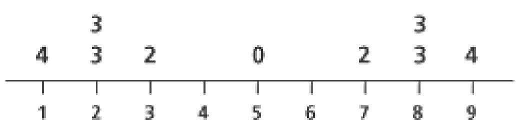
Solutions, cont'd.

Problem E3.

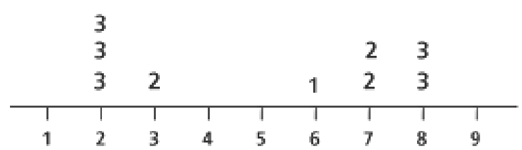
Number of Coins in Stack (x)	Deviation From the Mean (x - 5)	Absolute Deviation From the Mean |x - 5|
1	-4	4
1	-4	4
1	-4	4
1	-4	4
1	-4	4
10	+5	5
10	+5	5
10	+5	5
10	+5	5
45	0	40

The MAD is 40 / 9, or approximately 4.44. This is more than twice the MAD for Line Plot A and 10 times as large as the MAD for Line Plot B. The much larger MAD indicates that the values of Line Plot C are very far from the mean, as compared to the other two.

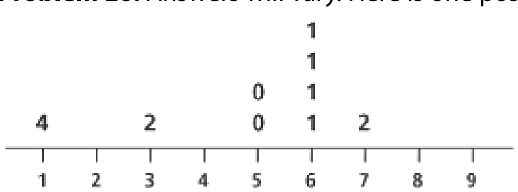
Problem E4. Answers will vary. Here is one possible line plot:



Problem E5. Answers will vary. Here is one possible line plot:



Problem E6. Answers will vary. Here is one possible line plot:



Problem E7. The reason it is impossible is that the MAD is the total of all absolute deviations. You may have noticed that in these problems the MAD is the sum of the deviations divided by 9. For the MAD to equal 1, the sum of the deviations would have to be exactly 9 (9 / 9 = 1). But the only way that could happen is if the total excess and the total deficit each were equal to 4.5. This would require splitting the coins, which cannot be done.

Solutions, cont'd.

Problem E8.

Number of Coins in Stack (x)	Deviation From the Mean (x - 5)	Squared Deviation From the Mean $(x - 5)^2$
4	-1	1
4	-1	1
5	0	0
5	0	0
5	0	0
5	0	0
5	0	0
6	+1	1
6	+1	1
45	0	4

The variance is 4 / 9, or approximately 0.44.

The standard deviation is the square root of the variance, which is 2/3, or approximately 0.67. The standard deviation is slightly higher than the MAD (which is 0.44), and is significantly smaller than the standard deviation of Line Plot A (which is 2.05). The great difference in the standard deviations indicates that the values in Line Plot B are much more closely distributed around the mean.

Problem E9.

Number of Coins in Stack (x)	Deviation From the Mean (x - 5)	Squared Deviation From the Mean $(x - 5)^2$
1	-4	16
1	-4	16
1	-4	16
1	-4	16
1	-4	16
10	+5	25
10	+5	25
10	+5	25
10	+5	25
45	0	180

The variance for Line Plot B is 180 / 9 = 20.

The standard deviation — the square root of 20 — is approximately 4.47. This is very close to the MAD calculated in Problem E3 (4.44), and is much higher than the standard deviations for Line Plot A (2.05) and Line Plot B (0.67).

Solutions, cont'd.

Problem E10.

a. The mean would increase by 3.

b. The MAD would not change. Since the values in the list are each 3 larger, and the mean is also 3 larger, the deviations from the mean would remain the same.

c. The variance would not change, since it depends only on the deviation from the mean, not the values themselves. Since the mean increases by 3 along with the rest of the data set, none of the deviations will change.

d. Since the standard deviation is the square root of the (unchanged) variance, it will not change.

e. The mean would be doubled.

f. The MAD would be doubled, since all the deviations are now doubled, and the MAD is the average of these deviations.

g. The variance would be multiplied by 4. Since calculating the variance involves squaring the deviations, the newly doubled deviations would all be squared, resulting in values that are four times as large. For example, if a deviation was (+3), it now becomes (+6). The value used in the variance calculation changes from $3^2 = 9$ to $6^2 = 36$, which is four times as large.

h. The standard deviation would be doubled, since it is the square root of the variance.

Homework

Problem H1. First you'll need to find the mean by adding all the values and dividing by 9:

6 + 6 + 9 + 6 + 6 + 7 + 8 + 9 + 6 = 63 / 9 = **7**

The mean is 7. (Note that in this calculation, the stacks are not in order. You do not need to order this list before calculating the MAD, variance, or standard deviation.) Here is the table for calculating the MAD:

Number of Coins in Stack (x)	Deviation From the Mean (x - 7)	Absolute Deviation From the Mean |x - 7|
6	-1	1
6	-1	1
9	+2	2
6	-1	1
6	-1	1
7	0	0
8	+1	1
9	+2	2
6	-1	1
63	0	10

The MAD is 10 / 9, or approximately 1.11.

Solutions, cont'd.

Problem H1, cont'd.

Here is the table for calculating the variance:

Number of Coins in Stack (x)	Deviation From the Mean (x - 7)	Squared Deviation From the Mean $(x - 7)^2$
6	-1	1
6	-1	1
9	+2	4
6	-1	1
6	-1	1
7	0	0
8	+1	1
9	+2	4
6	-1	1
63	0	14

The variance is 14 / 9, or approximately 1.56.

The standard deviation is the square root of 1.56, which is approximately 1.25. Note that the MAD and the standard deviation are roughly the same.

Problem H2. First you'll need to find the mean by adding all the values and dividing by 9:

1 + 4 + 10 + 4 + 4 + 6 + 7 + 10 + 4 = 50 / 9 = 5.6 (to one decimal place)

We will use 5.6 for the mean. Here is the table for calculating the MAD:

Number of Coins in Stack (x)	Deviation From the Mean (x - 5.6)	Absolute Deviation From the Mean $|x - 5.6|$
1	-4.6	4.6
4	-1.6	1.6
10	+4.4	4.4
4	-1.6	1.6
4	-1.6	1.6
6	+0.4	0.4
7	+1.4	1.4
10	+4.4	4.4
4	-1.6	1.6
50	0	21.6

The MAD is approximately 21.6 / 9 = 2.4 (to one decimal place). (If you are using fractions, the exact answer is 194/81.)

Solutions, cont'd.

Problem H2, cont'd.

Here is the table for calculating the variance:

Number of Coins in Stack (x)	Deviation From the Mean (x - 5.6)	Squared Deviation From the Mean $(x - 5.6)^2$
1	-4.6	21.16
4	-1.6	2.56
10	+4.4	19.36
4	-1.6	2.56
4	-1.6	2.56
6	+0.4	0.16
7	+1.4	1.96
10	+4.4	19.36
4	-1.6	2.56
50	0	72.24

The variance is 72.24 / 9 = 8.0 (to one decimal place). (If you were using fractions, the exact answer is 650/81.)

The standard deviation is the square root of 8, which is 2.8 (to one decimal place). (The exact fractional answer, the square root of 650/81, cannot be expressed by a rational number, since 650 is not a square number.)

The results of Problems H1 and H2 suggest that the data in Problem H2 have greater variation about the mean than the data in Problem H1.

Notes