1

Variability: Mean Absolute Deviation and Standard Deviation

Introduction to Statistics

Elizabeth G. Arnold, *Colorado State University*
Elizabeth W. Fulton, *Montana State University*
Katharine M. Banner, *Montana State University*
Rachel Tremaine, *Colorado State University*

1.1 Overview and Outline of Lesson

Variability in data is a central concept in statistics, addressed throughout middle school, high school, and college statistics curricula. This lesson focuses on how two measures of variability—mean absolute deviation (MAD) and standard deviation (SD)—can be used to quantify the variation in a dataset. Both MAD and SD measure variability in terms of "average distance from the mean," although they use different methods to calculate "average distance." This lesson highlights how SD builds upon the understanding of variability students are introduced to in middle school with MAD. Understanding how to measure variation is an essential component of making statistical inferences, and statistical inference plays a large role in any undergraduate introduction to statistics course. Undergraduates have likely had different experiences learning about variability in their K–12 schooling, and this lesson respects what undergraduates may have previously encountered and presents a unified framework for developing a deeper understanding of "average distance from the mean" as a measure of variability.

1. Launch—Pre-Activity
   Prior to the lesson, undergraduates complete a Pre-Activity which introduces the importance of describing variability in data. In the Pre-Activity, undergraduates examine three dotplots to compare the center and variability of three different datasets. Instructors can launch the lesson by reviewing the solutions to the Pre-Activity.

2. Explore—Class Activity

   - *Problem 1*:
     Undergraduates quickly play a memory game online and record their time in order to generate a class dataset. Undergraduates then create a graphical summary of the data and describe what they notice about the data. The context of the memory game is used throughout the Class Activity.
   - *Problems 2 & 3*:
     Undergraduates analyze hypothetical student work to make sense of how to compute and interpret the mean absolute deviation of a dataset. The instructor provides a brief discussion to formally define and interpret the mean absolute deviation.

- *Problem 4*:
  Undergraduates analyze hypothetical student work to make sense of how to compute and interpret the standard deviation of a dataset. The instructor provides a brief discussion to formally define and interpret the standard deviation.
- *Problem 5*:
  Undergraduates return to their class dataset from Problem 1 and compute and interpret the mean absolute deviation and standard deviation of their data.

3. Closure—Wrap-Up
   The instructor concludes the lesson by revisiting the calculations and interpretations of mean absolute deviation and standard deviation and discussing how the concept of standard deviation compares to and builds on the concept of mean absolute deviation.

1.2 Alignment with College Curriculum

Given that variability plays a central role in teaching and learning statistics, a deep focus on this topic at the collegiate level will serve all undergraduates, including prospective teachers, exceptionally well. This lesson offers undergraduates an opportunity to develop an understanding of the concept, "average distance from the mean" by examining two different measures of variability (mean absolute deviation and standard deviation). The lesson fits well after instructors have taught different ways to visualize univariate quantitative data and have discussed different measures of center (such as mean). Because this lesson only focuses on two measures of variability, instructors may wish to implement another lesson to address other measures of variability, such as range and interquartile range.

1.3 Links to School Mathematics

Statistics content standards are integrated throughout middle school and high school mathematics curricula, and it is important for prospective teachers to understand how and why variability is fundamental to teaching and learning statistics. By studying connections between the mean absolute deviation and the standard deviation, prospective teachers will develop a deeper understanding of variability and why the study of mean absolute deviation in middle school serves as a precursor to the study of standard deviation in high school.

This lesson highlights:

- Computing and interpreting the mean absolute deviation and the standard deviation of a dataset;
- Connections between the mean absolute deviation and the standard deviation.

This lesson addresses several statistical knowledge and mathematical practice expectations in common high school standards documents, such as the Common Core State Standards for Mathematics (CCSSM, 2010). Middle school students are expected to understand mean absolute deviation as one way to quantify the amount of variability present in data. Middle school students learn how to compute and interpret the mean absolute deviation of a dataset in the context of a problem (c.f. CCSS.MATH.CONTENT.6.SP.B.5.C). High school students are expected to build on their understanding of mean absolute deviation to develop an understanding of how to compute and interpret the standard deviation of a dataset, which is a more common measure of variability used in practice (c.f. CCSS.MATH.CONTENT.HSS.ID.A.2 and CCSS.MATH.CONTENT.HSS.ID.A.3). This lesson also provides opportunities for prospective teachers to think about the reasoning of others and construct sound statistical arguments.

1.4 Lesson Preparation

Prerequisite Knowledge

Undergraduates should know:

- Measures of central tendency (mean, median, mode);
- How to visualize univariate quantitative data with dotplots, boxplots, and histograms.

Learning Objectives

In this lesson, undergraduates will encounter ideas about teaching mathematics, as described in Chapter 1 (see the five types of connections to teaching listed in Table 1.2). In particular, by the end of the lesson undergraduates will be able to:

- Compute the mean absolute deviation and the standard deviation of a dataset;
- Interpret the mean absolute deviation and the standard deviation of a dataset in the context of a problem;
- Describe how the concept of standard deviation builds on the concept of mean absolute deviation;
- Examine hypothetical student work to make sense of "average distance from the mean";
- Evaluate and pose questions to help guide students' understanding about mean absolute deviation and standard deviation.

Anticipated Length

Two 50-minute class sessions.

Materials

The following materials are required for this lesson.

- Pre-Activity (assign as homework prior to Class Activity)
- Class Activity (print Problems 1, 2–3, 4, and 5 to pass out separately)
  - Computer/Tablet/Phone (for undergraduates) to play the memory game at the beginning of the Class Activity
  - Computer (for instructor) to compile the class dataset and create a graphical summary of the data during the Class Activity
- Homework Problems (assign at the end of the lesson)
- Assessment Problems (include on quiz or exam after the lesson)

All handouts for this lesson appear at the end of this lesson, and LaTeX files can be downloaded from INSERT URL HERE.

1.5 Instructor Notes and Lesson Annotations

Before the Lesson

Assign the Pre-Activity as homework for undergraduates to complete in preparation for the lesson, and ask undergraduates to bring their solutions to class on the day you start the Class Activity. At your discretion, allow undergraduates to use technology to compute the mean number of pets for each class in part (a).

> *The goal of the Pre-Activity is to introduce the importance of accounting for variability in data, so we purposely constructed each dataset to have the same mean number of pets (i.e., 1.5 pets) but the variability (and the distributions) are different.*

Pre-Activity Review (10 minutes)

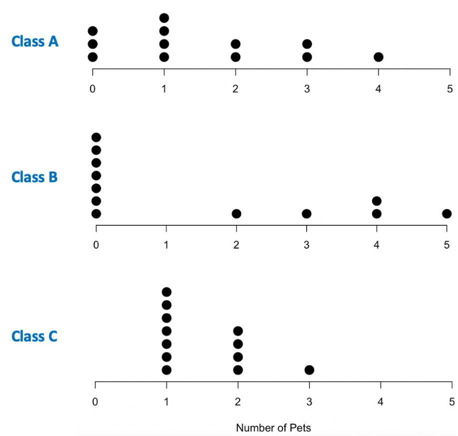
As a class, review the solutions to the Pre-Activity. Focus most of the discussion on part (c) define variability (e.g., "variability, sometimes referred to as spread, is commonly described by specifying how far the data are from a measure of center"), and discuss the following connection to teaching. Undergraduates' answers in the Pre-Activity may vary but the main point is for them to recognize that the amount of variability present in each distribution is different.

Discuss This Connection to Teaching

"Learning from data and making informed choices on the basis of data depend on understanding and describing variability" (Peck et al., 2013, p. 25). It is important for prospective teachers to understand that finding the mean of a dataset is often not enough to summarize quantitative data or characterize a distribution. In K–12 mathematics, students will be asked to compare distributions and they will need to consider the shape, center, and spread of the distributions.

Pre-Activity

1. Students from three different classes reported the number of pets in their household. The results are summarized graphically as dotplots and in a frequency table below.



Class A	Class B	Class C
1	0	1
0	0	1
4	0	1
3	4	1
2	4	2
1	3	1
1	0	2
3	5	2
0	0	3
2	0	2
1	0	1
0	2	1

(a) Compute the mean number of pets for each class.

Solution:
For Classes A, B, and C: $\bar{x}_A = \bar{x}_B = \bar{x}_C = 1.5$ pets.

(b) What is similar about the three dotplots?

Sample Responses:

- The mean number of pets in each class is the same.
- There are 12 dots in each dotplot.
- All have the same mean of 1.5 pets.

(c) What is different about the three dotplots?

Sample Responses:

- The shape of each dotplot is different.
- The medians are different. Class A and Class C have a median of 1 pet but Class B has a median of 0 pets.
- The modes are different. Most of the students in Class B have 0 pets and most of the students in Class A and Class C have 1 pet.
- The spread of each dotplot is different.
- The standard deviation of each dotplot is different. $s_A = 1.31$ pets, $s_B = 1.98$ pets, and $s_C = 0.67$ pets.

Conclude the Pre-Activity Review by discussing the following connection to teaching and letting undergraduates know that the purpose of this lesson is to explore two different measures of variability: mean absolute deviation (MAD) and standard deviation (SD).

Discuss This Connection to Teaching

National and state standards now emphasize statistics content standards in middle school and high school with the central theme of variability being developed throughout. The Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report states that "Statistical thinking, in large part, must deal with the omnipresence of variability in data (e.g., variability within a group, variability between groups, sample-to-sample variability in a statistic). Statistical problem solving and decision making depend on understanding, explaining, and quantifying variability in the data within the given context" (Bargagliotti et al., 2020, p. 7). Because the concept of variability is foundational to teaching and learning statistics, undergraduates are expected to develop a deep understanding of this concept, and this is especially true for prospective teachers who will teach their students statistics.

Class Activity: Problem 1 (10 minutes)

Pass out **Problem 1** of the Class Activity. Ask each undergraduate to access the *Census at School* webpage and play the *Memory Game* (`http://ww2.amstat.org/education/cas/1.cfm`). Explain to the undergraduates why you are using a game from this particular website and its connection to teaching.

Discuss This Connection to Teaching

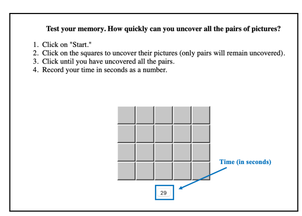
The *Census at School* site is an international resource that engages students in statistical problem solving. It is a useful resource for prospective teachers to be aware of and use in their future classrooms for class activities and projects. The site also serves as a good source for obtaining authentic data for statistical investigations.

Give undergraduates about a minute to play the game. Remind them to record their time (in seconds) from their **first attempt only**. As undergraduates play the game, set up a way to compile, share, and present the class dataset of memory game times (e.g., GoogleSheets, StatKey, etc.) so that you (or your undergraduates) can create a graphical summary of the data.

Class Activity Problem 1 : Part a

1. **Test Your Memory!**

   Play the *Census at School Memory Game* where you will need to uncover and match 10 pairs of pictures. The time it takes you to complete the game will be tracked. Go to the following link to access the game:
   `https://ww2.amstat.org/education/cas/1.cfm`



   (a) Play the game once and record your time (in seconds) as a number.

   Solution:
   Answers will vary. Most undergraduates complete the game in less than a minute.

   Commentary:
   If undergraduates play the game more than once, consider having a class discussion about why, from a statistical perspective, we should only use the time from their first attempt of the game. This may include the following ideas:

- You may get quicker at completing the game by playing it more than once, so using times from everyone's first attempt is more standardized among everyone.
- If we gave everyone 5 minutes to play the game and used all of the attempts, then those who were faster at completing the game would have more of their times recorded in the dataset. This would result in repeated measures (i.e., non independent observations from the same person), which might make the measure of variability smaller than what it is expected.

Complete Problems 1(b) and 1(c) as a class. Start by asking undergraduates to identify and explain what type of graph would be appropriate to summarize the class data. Next, create the type of graph recommended by undergraduates, display it to the class, and ask them to sketch the graph on their Class Activity handout for Problem 1(c).

Class Activity Problem 1 : Parts b & c

(b) As a class, compile everyone's time in a dataset. What graphical summary would be appropriate to visualize the class's distribution of times on the memory game? Explain your reasoning.

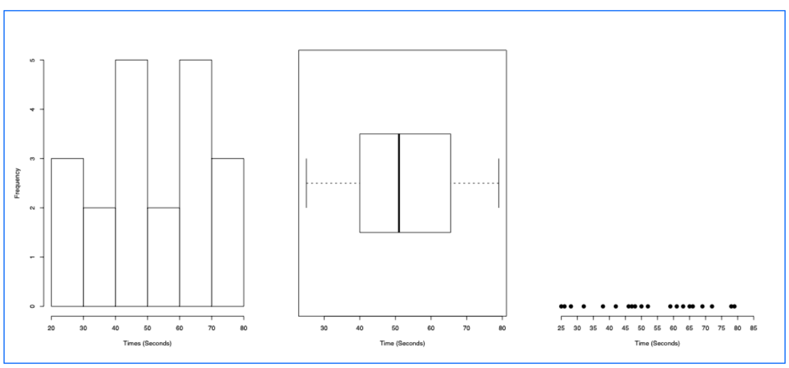
Sample Responses:

- A histogram because the times are quantitative.
- We can make a dotplot of the memory game times since the times are quantitative.
- I would make a boxplot because we have a single dataset of quantitative values.
- A stem-and-leaf plot would appropriately display the class's memory game times.

Commentary:
If all undergraduates suggest the same type of plot, ask them to think of other appropriate ways to visualize the data. If undergraduates suggest inappropriate types of graphs (such as a scatterplot), discuss why they are not appropriate to visualize these data.

(c) As a class, create a graphical summary to visualize the class's distribution of times on the memory game and sketch it below.

Sample Graphical Summaries:



Commentary:
Dotplots and histograms are the primary types of graphs used throughout the lesson so it may be worthwhile to quickly create both of these types of plots. If you create a dotplot, ask undergraduates, *What does one dot on this dotplot represent?* and lead them to an appropriate understanding (e.g., "one dot represents how long a single student in the class took to successfully complete the memory game.").

Finish Problem 1 by asking undergraduates to document what they notice about the class's distribution of memory game times, based on the graphical summary, in Problem 1(d).

Class Activity Problem 1 : Part d

(d) Describe what you notice about the class's distribution of times on the memory game.

Sample Responses:
Undergraduates may describe features about

- The center, shape, and spread of the distribution.
- The mode(s) of the distribution.
- Any noticeable outliers.

Class Activity: Problems 2 & 3 (30 minutes)

Pass out **Problems 2 and 3** of the Class Activity, which focus on computing and interpreting the MAD of a dataset. Undergraduates will work on these two problems in groups. (See Chapter 1 for guidance on facilitating group work and selecting and sequencing student work for use in whole-class discussion.) Before they begin working, discuss the following connection to teaching.

Discuss This Connection to Teaching

These next several problems in the Class Activity focus on analyzing hypothetical students' thinking in order to develop undergraduates' skills in understanding school student thinking and developing questions to guide school students' understanding. It is important for undergraduates, especially prospective teachers, to understand how others use, reason with, and communicate statistics. These problems also give prospective teachers (and tutors and future graduate students) an opportunity to think about how they would respond to student work in ways that nurture students' assets and understanding and in ways that help develop students' statistical understanding.

Emphasize to undergraduates that the hypothetical students' work they will be examining is procedurally correct so they do not need to be looking for any arithmetic errors. Rather, encourage undergraduates to focus on the reasoning behind the calculations that Jasmine is using to compute the MAD. We intend for Problem 2 to be an informal introduction to MAD, and you will formally define it after undergraduates complete this problem. From previous implementations of this lesson, we have noticed that some undergraduates are familiar with MAD, but most do not recall learning MAD prior to this lesson.

In Problem 2, Jasmine drew a histogram and is computing the MAD. She first computes the mean. Then she computes the absolute value of the differences between each data point to the mean (i.e., the distances). Finally, she computes the average of those distances to compute the MAD.

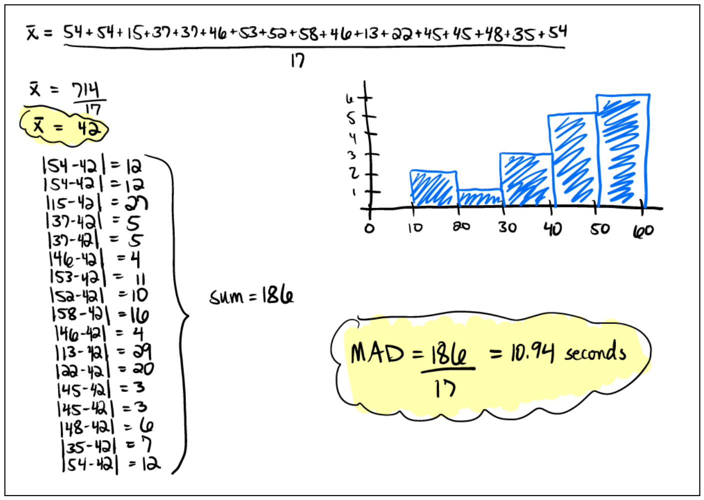
Class Activity Problem 2

2. **Quantifying Variability with Mean Absolute Deviation**

Amaury is teaching a high school intermediate algebra class and his students are learning about different measures of variability. The students in his class played the *Census at School Memory Game* and recorded their times, in seconds:

54, 54, 15, 37, 37, 46, 53, 52, 58, 46, 13, 22, 45, 45, 48, 35, 54

Amaury asked his students to create a graphical summary of the data, compute the mean, and quantify the amount of variability present. One of his students, Jasmine, did the following:



Jasmine, recalling what they learned in middle school, quantified the amount of variability by computing the **mean absolute deviation (MAD)**. All of their calculations are correct. Describe mathematically what Jasmine did to compute the MAD.

Sample Responses:

- Jasmine subtracted the value of each data point from the mean and took the positive value of that result by taking the absolute value. Then they averaged all of these values.
- Jasmine found the mean (42) and calculated the distance each data value was from the mean. Jasmine added all those distances and then divided by 17 (# of data values).
- Jasmine first got the mean of the data by adding all the numbers together, then dividing by the total number. In order to get the mean absolute deviation, Jasmine used absolute value and subtraction in order to find the distance from each data point and the mean. They then computed the average of those distances.

Commentary:
If undergraduates may have difficulty recognizing how deviations from the mean are used in Jasmine's calculation, focus their attention to these deviations and how absolute value is used.

Discussion: Mean Absolute Deviation (MAD)
Formally define MAD, discuss how it is calculated, and emphasize the following connection to teaching. At your discretion, show undergraduates how to compute the MAD with technology. Note that R has an `MAD` command, but this computes the median absolute deviation.

To compute the mean absolute deviation (in R) of a dataset stored as the object "x", you can use the following code:

```
meanAD <- function(x){
  avg <- mean(x)
  mad <- mean(abs(x) - avg)
  return(mad)
}
```

Example:

```
x <- c(0.84, -0.18, -1.28, -1.07, -0.05)
meanAD(x)
# [1] 1.032
```

Discuss This Connection to Teaching

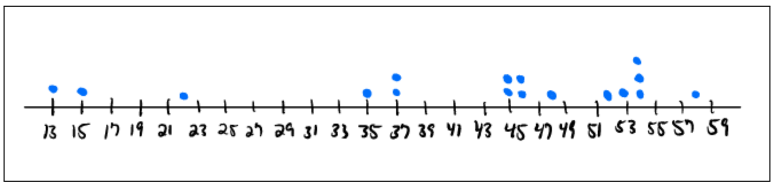
MAD is a measure of variability typically taught to middle school students and it is a precursor to the study of standard deviation in high school.

Definition:

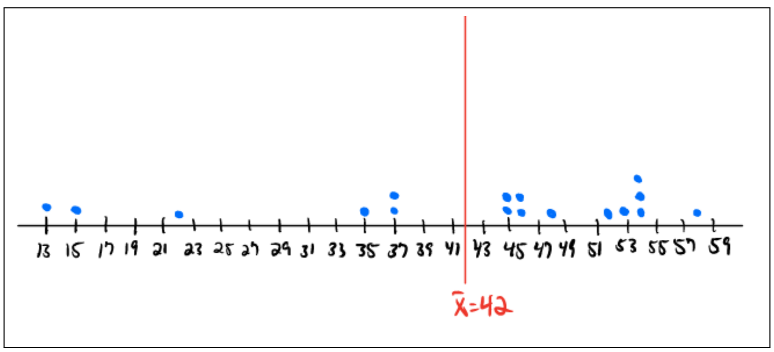
- MAD stands for the **mean absolute deviation** (not "mean average deviation" or "median absolute deviation").
- Informally, the MAD measures how spread out the data are and provides a numerical value to quantify the amount of variability present in data.
- Specifically, the MAD measures the average distance the data values are from the mean.

Conceptual Understanding:
Construct the following sequences of dotplots to help undergraduates conceptually understand MAD; this includes an understanding of how deviations from the mean are treated in the formula for MAD. Note that one of the Assessment Problems includes a dotpot with deviations drawn in.

1. Draw a dotplot of the high school class's data.

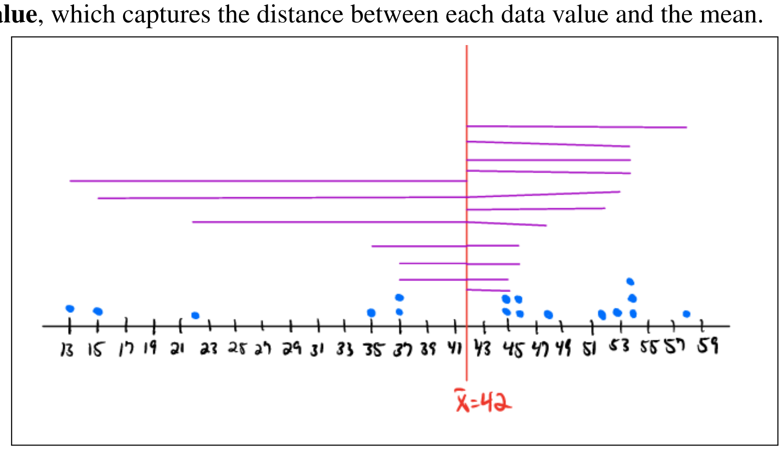


2. Indicate the mean (42 seconds) by drawing a vertical line through 42.



3. Ask undergraduates how far each data point is from the mean and then draw in each deviation from the mean using horizontal lines. Label them as indicated by the undergraduates.

   Some undergraduates may give negative values for points to the left of the mean and positive values for points to the right of the mean. If this occurs, ask undergraduates what happens when we sum those positive and negative values. Here, they will notice the sum will be zero which may help them understand why the formula for MAD uses **absolute value**, which captures the distance between each data value and the mean.



Emphasize that to compute the MAD, we find the distance between each data point and the mean, and then take the average of those distances.

Thus, MAD represents the "average distance from the mean." A smaller MAD generally indicates that the data are closer to the mean. A larger MAD generally indicates that the data are farther from the mean.

Formula:

- The formula for MAD is given by:

$$\text{MAD} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Problem 3 focuses on interpreting the MAD in the context of a problem, which can be challenging for undergraduates. This problem is intended to guide undergraduates in developing their knowledge about what a correct (and complete) interpretation of MAD includes. All three of the interpretations in Problem 3 are real examples from students. Tarryn's interpretation is correct. Jasmine's and Benny's interpretations are a good start, but they are incomplete and need to be expanded upon.

Class Activity Problem 3 : Parts a & b

3. **Interpreting Mean Absolute Deviation**

Two other students, Tarryn and Benny, also correctly computed the MAD. When Amaury asked his students to write a sentence interpreting their measure of variability in the context of the problem, Jasmine, Tarryn, and Benny wrote the following sentences:

Jasmine	The MAD is 10.94 seconds.
Tarryn	On average, the memory game times were 10.94 seconds away from the mean of 42 seconds.
Benny	A memory game time is 10.94 from the mean.

(a) One student correctly (and completely) interpreted the MAD in the context of the problem. Identify who it was, and describe what components of their interpretation make it correct and complete.

Sample Response:
Tarryn's interpretation is correct. She focuses on "average distance from the mean" in her interpretation and includes units (seconds) in her response.

Commentary:
From our experience, most undergraduates correctly identify Tarryn's interpretation as correct. Take time to highlight why her interpretation is correct, focusing on accounting for "average distance from the mean" in the interpretation.

(b) The other two students gave incomplete interpretations of the MAD. Based on their interpretations, describe what each

i. may understand about interpreting the MAD, and
ii. may not yet understand about interpreting the MAD.

Sample Responses:

- Jasmine
  - Jasmine understands that the MAD has a unit (seconds).
  - They may not fully understand what MAD measures because they did not write out what MAD measures in the context of the problem.

- Benny
  - Benny understands that MAD is a measure of distance from the mean.
  - He doesn't yet understand that MAD is an average distance from the mean.

Commentary:
In previous implementations of this lesson, we have noticed that undergraduates tend to focus only on what the hypothetical students do not yet understand. Make sure that undergraduates also attend to what the students do understand.

After discussing the solutions to Problems 3(a) and 3(b), emphasize the following connection to teaching:

Discuss This Connection to Teaching

The practice of evaluating student work requires knowledge to assess what a student does and does not yet understand, and prospective teachers need practice with the skill of nurturing students' assets and understanding in the classroom. Addressing different perspectives in a manner that conveys respect for student thinking and reasoning, both when a student's work is correct and when it is incorrect, is a critical practice for teachers.

Undergraduate responses to Problem 3(c) will vary. A key idea to include in an answer is that MAD is a measure of variability that describes the average distance from the mean.

Class Activity Problem 3 : Part c

(c) In a general context, describe what MAD measures.

Sample Response:
MAD tells us, on average, how far the data are from their sample mean. The higher the number, the more spread out the data are.

After discussing the solutions to Problem 3, emphasize the importance of statistical interpretations in this connection to teaching.

Discuss This Connection to Teaching

Part of the statistical problem-solving process involves interpreting results, and interpretations should integrate the context of the problem. Statistical thinking is different from mathematical thinking because of the omnipresence of variability and because of the central role context plays in understanding and interpreting data. All undergraduates benefit from having to interpret statistical concepts in the context of the problem. This skill is especially important for prospective teachers who will teach their students how to compute and interpret various measures of variability.

Advice on Delivering the Lesson Over Two Class Sessions

If you are teaching this lesson over two class sessions, stopping around Problem 3 is a good place. See Chapter 1 for guidance on using exit tickets to facilitate instruction in a two-day lesson.

Class Activity: Problem 4 (20-25 minutes)

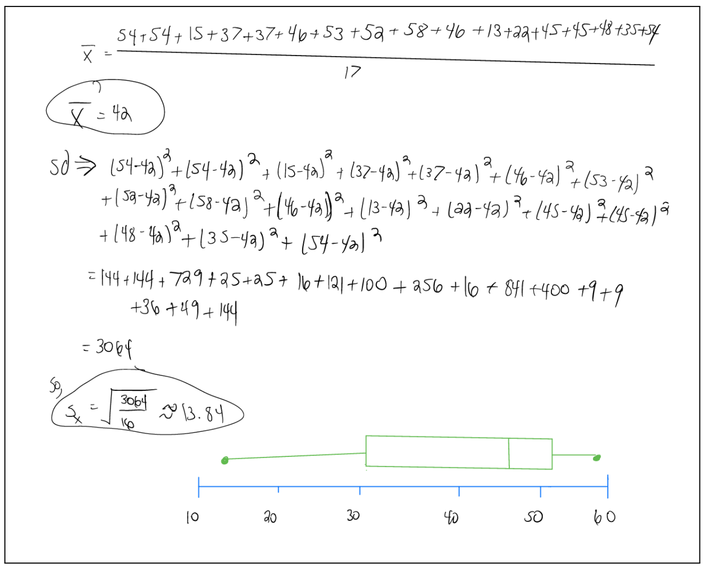
Pass out **Problem 4** of the Class Activity, which focuses on calculating and interpreting the standard deviation (SD) of a dataset. Instruct undergraduates to work in small groups on Problem 4 and let them know they will continue to examine hypothetical students' work. For Problem 4(a), emphasize that the student's work is procedurally correct so they do not need to be looking for any arithmetic errors. Rather, encourage undergraduates to focus on the reasoning behind the calculations that Josief is using to compute the SD. We intend for Problem 4(a) to be an informal introduction to SD, and you will formally define it after undergraduates complete this problem.

In Problem 4, Josief drew a boxplot and is computing the standard deviation (SD). He is finding deviations from the mean and squaring them. He adds these values, divides the sum by 1 less than the total number of data values, and concludes by taking the square root to compute the SD.

Class Activity Problem 4 : Part a

4. **Quantifying Variability with Standard Deviation and Interpreting Standard Deviation**

   Josief, another student in Amaury's class, did something different, as shown below:



   (a) Josief quantified the amount of variability by computing the **standard deviation (SD)**. All of his calculations are correct. Describe mathematically what he did to compute the SD.

   Sample Responses:

   - Josief first found the mean. Then he computed the distance from each data point and the mean, squared those distances, and then added all of those numbers. With the sum of those numbers, he then divided by the amount of numbers minus one and then took the square root.
   - To calculate standard deviation, calculate the difference between each value and the mean, square those differences and add them together, divide by one less than the total number of data points, and take the square root of that value.
   - Josief computed the mean. Then he calculated $(x_i - \bar{x})^2$ for each data value, added them up, and divided by 16. Finally, he took the square root of that number to get the SD.
   - He first found the mean. Then subtracted the mean from each data point and squared those values. He divided the sum of those values by 16 and took the square root.

   Commentary:
   If undergraduates have difficulty recognizing how deviations from the mean are used in Josief's calculation, focus their attention to these deviations and how they are squared.

Discussion: Standard Deviation (SD)
Formally define SD, discuss how it is calculated, and discuss the following connection to teaching. At your discretion, show undergraduates how to compute the SD of a dataset with technology.

Discuss This Connection to Teaching

SD is a measure of variability taught to high school students, and it builds on the concept of MAD. Both MAD and SD measure the same concept but they are calculated differently. SD is one of the more common measures of variability used in practice.

Definition:

- SD represents the **standard deviation** of a dataset.
- Informally, the SD measures how spread out the data are and provides a numerical value to quantify the amount of variability present in data.
- Specifically, the SD measures the average distance the data values are from the mean.

Formula:

- The formula for SD is given by:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

- To compute the SD, you first compute the mean of the dataset. Then, you compute the deviations from the mean by subtracting the mean from each data value. You will square all of those values, add them up, and divide the sum by one less than the number of total data values. Finally, you take the square root of that number to compute the SD.

> *Some undergraduates may ask about taking the square root and dividing by $n-1$ instead of n when computing the SD. Discuss this idea as you normally would in your class (e.g., The reason for using $n-1$ instead of n is beyond the scope of this course, but is related to statistical theory. In short, the sampling distribution for s^2 can be approximated by a known distribution if we divide by $n-1$ instead of n (as long as n is large), which makes statistical inference more straightforward for many applications.).*

We have found that undergraduates find it challenging to interpret the SD in the context of a problem. Problem 4(b) focuses on guiding undergraduates in developing their knowledge about what a correct (and complete) interpretation of SD includes. Problem 4(c) focuses on giving undergraduates an opportunity to think about how they would respond to student work in ways that help the student develop an understanding of interpreting SD in the context of a problem.

Class Activity Problem 4 : Parts b & c

(b) When asked to write a sentence to interpret the SD in the context of the problem, Josief wrote the following:

The memory game times varied by 13.84 seconds.

Describe why Josief's interpretation is not completely correct.

Sample Response:
Josief is not attending to the fact that standard deviation measures an average distance from the mean.

Commentary:
The main point is that Josief's interpretation is missing the "average from the mean" aspect. More specifically, not all memory game times will vary exactly 13.84 seconds from the mean but that "on average" the times were about 13.84 seconds faster or longer than the average time of 42 seconds.

(c) Consider the following questions that one might ask Josief to help him with his interpretation of the standard deviation (in the context of the memory game times).

i. Explain how the following question might help Josief to advance in his understanding of interpreting standard deviation in the context of a problem:

Can you say more about how the memory game times varied?

Sample Response:
Asking Josief to describe how the memory game times varied will push him to explain what he means by "varied" in his interpretation. This may help him understand that standard deviation is one particular measure of variability that measures average distance to the mean, but its interpretation differs from other measures of variability. For instance, the range also describes how the data vary but in a way that is different from how the standard deviation describes variability.

ii. Explain how the following question might help you assess what Josief understands about interpreting the standard deviation in the context of a problem:

What does standard deviation measure?

Sample Response:
Asking this will help me know if he wrote the interpretation based on his understanding of what standard deviation measures or if he fully understands what standard deviation measures and is having difficulty translating that interpretation to the context of the problem.

iii. Explain why the following question might not help Josief:

Why is your interpretation incorrect?

Sample Response:
Josief probably thinks his answer is correct, so asking this question isn't helpful.

Undergraduate responses to Problem 4(d) will vary. A key idea to include in an answer is that SD is a measure of variability that describes the average distance from the mean.

Class Activity Problem 4 : Part d

(d) In a general context, describe what SD measures.

Sample Response:
SD tells us how far the data points deviate from their sample mean, on average.

Class Activity: Problem 5 (10 minutes)

Pass out **Problem 5** of the Class Activity, which asks undergraduates to return to the class's dataset from Problem 1 so they can compute the MAD and SD of that dataset. At your discretion, allow undergraduate to use technology for these computations. Focus on undergraduates' interpretations, making sure they are complete and correct. If you have run out of time, it may be helpful to assign this problem as homework.

Class Activity Problem 5

5. Return to the class dataset from Problem 1.

(a) Compute the mean absolute deviation of your class's data of memory game times and write a sentence interpreting the mean absolute deviation in the context of the problem.

Sample Response:
The average distance from the mean time of [*value of mean*] seconds is [*value of MAD*] seconds.

(b) Compute the standard deviation of your class's data of memory game times and write a sentence interpreting the standard deviation in the context of the problem.

Sample Response:
The average distance from the mean time of [*value of mean*] seconds is [*value of SD*] seconds.

Wrap-Up (15 minutes)

Conclude the lesson by revisiting the computations and interpretations of MAD and SD, discussing how the concept of SD compares to and builds on the concept of MAD. Further, it may be appropriate to discuss how MAD and SD are sensitive to outliers and why SD is used more often in practice. See below for specific considerations.

Importance of Computing and Interpreting MAD and SD:

- Variability is central to teaching and learning statistics because it is a foundational concept for understanding sampling distributions and sampling variability/error (i.e., the SD of the sampling distribution which is discussed later in the semester).
- In this lesson we focused on two different measures of variability, MAD and SD, both of which rely on deviations from the mean, and quantify how far, on average, data values are from the mean.
  - MAD uses the absolute value of the deviations and SD uses the squares of the deviations, both resulting in positive values.
- We also focused on interpreting statistics, a crucial part of the statistical problem-solving process.

MAD and SD in Middle School, High School, and College:

- MAD is a measure of variability taught in middle school while standard deviation is a measure of variability taught in high school.
- To conduct statistical inference, we need to quantify the background variability in the estimator of a population parameter of interest. In this lesson, undergraduates developed a deeper understanding of how two measures of variability can be used to quantify the variation in a dataset. The role of standard deviation in doing statistical inference will be present throughout the rest of an undergraduate introduction to statistics course, particularly when undergraduates quantify the standard deviation of a sampling distribution to find a margin of error and build a confidence interval.

How SD Builds on MAD:

- There's a reason MAD is a precursor to SD in school mathematics. MAD is easier to compute, and the concept of SD builds on the understanding students have of MAD.
- After students learn to compute and interpret the MAD of a dataset, they will learn that SD has a similar interpretation and a slightly more complex computation. Both MAD and SD have the same units and measure "average distance from the mean."
- The MAD is a good introduction to the concept of measuring "average distance from the mean" and more intuitive to understand compared to the SD because the computation involves taking a "true" average and the concept of average is taught in middle school.
  - MAD is a considered a more "true" average since it's calculated with the formula $\frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$ (i.e., a sum divided by the total number of values), whereas the SD is calculated using $\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ (i.e., a sum divided by one less than the total number of values).

Sensitivity to Outliers:

- MAD and SD are both sensitive to outliers and are not robust measures of spread.

  - MAD and SD are sensitive to outliers because of the way they are calculated. MAD takes the absolute value of deviations from the mean and the SD squares those deviations. The SD then sums all of the individual squared deviations, divides by the number of observations minus one and then takes the square root of that quantity. This process makes large deviations from the mean have a greater influence in the calculation of SD compared to what happens in the calculation of MAD, so generally SD is a larger number than the MAD.
- MAD/SD measure deviations from the mean so it makes sense to report a SD/MAD when a mean is reported as a measure of center, and these measures are more appropriate when a distribution is unimodal and symmetric. When data contain an outlier, other measures of variability, such as IQR, are more appropriate to quantify the amount of variability present.

Discuss This Connection to Teaching

High school students are asked to relate the choice of measures of center and spread to the shape of the data distribution. Depending on the shape of the distribution, appropriate descriptors of center and spread may change. Prospective teachers should understand how different measures of variability correspond to different measures of center and the shape of a distribution.

Why SD is Used More in Practice:

- In practice, we often report the SD (more than the MAD) because much of statistical theory is based on using the SD. Undergraduates will learn more about the statistical theory if they continue with their statistics education and take a probability theory and mathematical statistics course sequence.
  - Working with squared deviations (that is, $(x_i - \bar{x})^2$) is computationally easier than working with absolute deviations (that is, $|x_i - \bar{x}|$). In other words, it's much easier to take the derivative of squares than to take the derivative of an absolute value. For example, taking a derivative is necessary for finding the maximum likelihood estimator (MLE) for the true population variance.
  - SD has a tie to a variance estimator that's important for applied statistics. The variance is the SD squared, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. When we have a random sample from a normal population (i.e., if $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$), this estimator of the population variance, σ^2, is unbiased (i.e., $E(s^2) = \sigma^2$), and as $n \to \infty$, s^2 of the approximate sampling distribution for the sample variance is known. This makes conducting statistical inference for the population variance mathematically convenient (see Wackerly et al. (2008) for more information).

You can ask undergraduates to complete an exit ticket, if you choose. See Chapter 1 for guidance on using exit tickets.

Homework Problems

At the end of the lesson, assign the following homework problems; we've focused on including problems that highlight connections to teaching. Assign any additional homework problems at your discretion.

Problem 1 prompts undergraduates to compute and interpret the MAD and the SD for a given dataset. At your discretion, allow undergraduates to use technology for these computations.

Homework Problem 1

1. Ten movies were randomly selected and the length of each movie (in minutes) is given below.

   152, 156, 98, 173, 68, 122, 92, 105, 138, 126

   (a) Compute the mean absolute deviation (MAD) and write a sentence that interprets the MAD in the context of the problem.

Solution:
$\bar{x} = (152 + 156 + 98 + 173 + 68 + 122 + 92 + 105 + 138 + 126)/10 = 123.$

Then,

$$\begin{aligned}\text{MAD} = (&|152-123| + |156-123| + |98-123| + |173-123| + |68-123| + \\ &|122-123| + |92-123| + |105-123| + |138-123| + |126-123|)/10 \\ = {}& 26 \text{ minutes}\end{aligned}$$

One sample interpretation is: "On average, the length of a movie is 26 minutes away from the mean of 123 minutes."

(b) Compute the standard deviation (SD) and write a sentence that interprets the SD in the context of the problem.

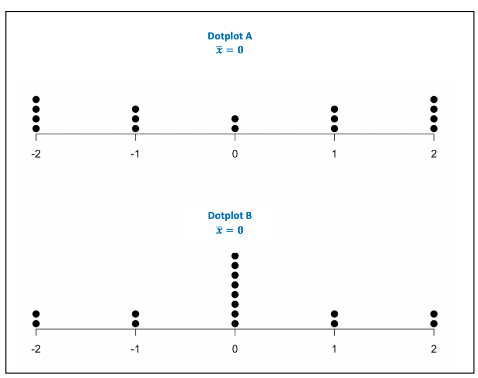
Solution:
The standard deviation (found using technology) is approximately 32.7 minutes. One sample interpretation is: "The difference between a movie length and the mean length of the movies is, on average, about 32.7 minutes."

High school students are asked to compare distributions in terms of their center and spread. It is important for future teachers to be able to visually assess and compare the center and spread of distributions so they can help their students do the same. The inspiration for Problem 2 came from delMas (2001).

Homework Problem 2

2. In this problem, you will *visually* compare the mean absolute deviation (MAD) between pairs of dotplots.

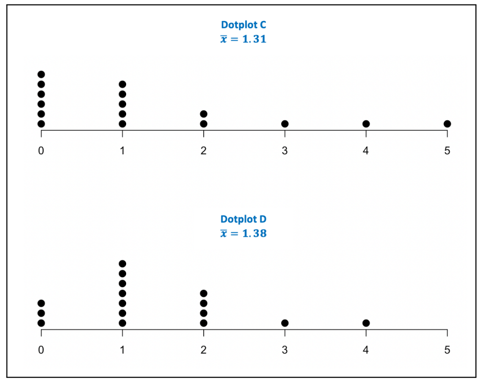
(a) **Dotplots A and B**.
Without doing any calculations, identify which dotplot (A or B) has the *larger* MAD. Explain your reasoning.



Sample Responses:
Dotplot A has the larger MAD. Sample responses are provided below.

- Dotplot A has more dots at the extreme ends (e.g., -2 and 2).
- The mean of Dotplot B is closer to its mode compared to the mean of Dotplot A and its mode. Thus, Dotplot B has less variability.

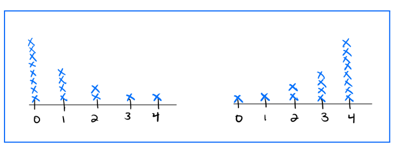
(b) **Dotplots C and D**.
Without doing any calculations, identify which dotplot (C or D) has the *larger* MAD. Explain your reasoning.



Sample Responses:
Dotplot C has the larger MAD. Sample responses are provided below.

- Dotplots C and D have similar means, but C has more skew than D. This means that Dotplot C has more variability.
- The dots in Dotplot D are more tightly clustered around its mean (compared to those in Dotplot C). Thus, the MAD of Dotplot D is smaller than the MAD of Dotplot C.

(c) Draw two different dotplots that have the same MAD. Describe your thought process when creating these two different dotplots.
Sample Response:



These are two different plots with the same MAD. I wanted to create mirror images that have the same range. Thus, the right dotplot is a reflection of the left dotplot. This makes it so that the distance between a data point and mean in the left dotplot is equivalent to distance between the reflected data point and reflected mean in the right dotplot. Thus, I know that the MAD of each will be the same.

Problem 3 prompts undergraduates to describe why the MAD is a statistical concept taught before the SD in K–12 schooling.

Homework Problem 3

3. Consider the mean absolute deviation (MAD) and the standard deviation (SD). Typically, MAD is first taught in middle school and SD is taught in high school. Describe why it is helpful for school students to learn MAD before SD.

Solution:
Answers will vary; key ideas to include in a correct solution are described below:

- Both the MAD and the SD have similar interpretations and measure, on average, how much quantitative data deviates from the mean. Because the MAD is simpler to compute, students can focus on understanding how to interpret the MAD and transfer that knowledge to interpreting the

SD.

- MAD is a more intuitive measure of spread because it describes the exact average distance each data point deviates from the mean.
- The MAD is easier to calculate than the SD. We don't need to square deviations and then take the square root when computing the MAD.

Sample Response:

- MAD is more obvious and understandable than SD. We have calculated the mean many times and when we do the MAD, the equation $\frac{\sum_{i=1}^{n}|x_i-\bar{x}|}{n}$ tells us exactly what we are doing. We're finding the average of the absolute deviations. The equation makes more sense to what we are finding. SD has additional components we need to comprehend when doing it. The equation $\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$ has a square, an $n-1$, and a square root which we have to learn more about to understand. They are similar measures because they measure "average distance from the mean," but it is easier to understand MAD.

Problem 4 has been adapted from a problem created by the *LOCUS Project* (Levels of Conceptual Understanding in Statistics). You can view commentary and correct answers to the original problem at `https://locus.statisticseducation.org/professional-development/questions/analyze-data?type=prodev_multiple_choice_question&field_prodev_level_tid=8&page=5`. In Problem 4, undergraduates have the opportunity to examine hypothetical student work and to identify what the students do and do not yet understand. Undergraduates will also practice posing questions that may guide the students' statistical understanding.

Overall, the LOCUS Project is a useful resource for future teachers to be aware of as they can potentially use these kinds of questions in their future classrooms. The problems from the LOCUS Project have been developed to measure students' understanding across levels of development (elementary, middle school, high school) as identified in the Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II) Report (Bargagliotti et al., 2020), and they align with the Common Core State Standards for Mathematics.

Homework Problem 4

4. Four students (Daveed, Monica, Juliana, and Bryant) are working on the following problem together, but they all pick a different answer.

> The director of the City Transportation System is interested in the amount of time required for a bus to make the trip from Downtown Station to City Mall. After collecting data for several months by recording the time it takes to make the trip, she finds that the distribution of times has a standard deviation of 3 minutes.
>
> Which of the following is the best interpretation of the standard deviation?
>
> A. A bus that leaves from Downtown Station typically arrives at City Mall 3 minutes later than the scheduled time.
>
> B. A bus typically takes about 3 minutes to get from Downtown Station to City Mall.
>
> C. The time a bus takes to get from Downtown Station to City Mall never varies more than 3 minutes from the mean trip time.
>
> D. The difference between the actual time a bus takes to get from Downtown Station to City Mall and the mean trip time is, on average, about 3 minutes.

Daveed selects option A, Monica selects option D, Juliana selects option C, and Bryant selects option B.

(a) Who selected the correct answer?

Solution:

Monica (Choice D) selected the correct answer.

(b) For *each* student who selected an incorrect answer, examine the choice they selected and describe a statistical concept they do understand.

Sample Responses:

- Daveed (Choice A):
  - Daveed may understand that the typical difference from the expected time is 3 minutes.
- Bryant (Choice B):
  - Bryant may understand that the typical time can be used to characterize part of a distribution.
- Juliana (Choice C):
  - Juliana may understand that standard deviation quantifies the deviation from the mean.

(c) For *each* student who selected an incorrect answer, examine the choice they selected and describe a statistical concept they might not yet fully understand.

Solution:

Answers can vary. We provide sample responses below and quotes from the *Correct Answer and Commentary* link at the LOCUS Project website.

- Daveed (Choice A) and Bryant (Choice B):
  - Daveed's and Bryant's choice demonstrates that they may not yet understand the concept of standard deviation as a measure of spread.
- Juliana (Choice C):
  - Based on her choice, Juliana may be struggling with the concept that "the standard deviation is a statement about average variability from the mean and not maximum variability from the mean. This option is incorrect because it prescribes absolute bounds on the variability, e.g. 'never varies by more than 3 minutes.'" (LOCUS Project).
- "Options (A) and (B) both make statements regarding the typical time of the bus trip, which would be measures of center . . . " (LOCUS Project)

(d) For *each* student who selected an incorrect answer, write a question you could ask them to help guide their understanding of interpreting a standard deviation in the context of the problem. Briefly explain how your question may help guide their statistical understanding.

Sample Responses:

- Daveed (Choice A):
  - Can you tell me what standard deviation means in your own words?
  - This question will help me assess to what degree Daveed understands what standard deviation represents. Then, we can work backwards from his understanding to interpret the standard deviation in the context of the problem.
- Bryant (Choice B):
  - What property of the distribution is the standard deviation measuring?
  - Bryant seems to be interpreting 3 minutes as the mean and not the standard deviation. Based on his response to my question, I would follow up and ask him to explain what about his choice describes variability.
- Juliana (Choice C):
  - Let's assume the typical time for a bus to travel from Downtown Station to City Mall is 5 minutes. Is it possible for the bus to take 10 minutes to make that trip?

– Juliana is close to correctly interpreting the standard deviation in the context of the problem, but her use of "never" is problematic. By asking Juliana this question, it may help her see that it could take longer than 8 minutes (i.e., typical time + 3 minutes) due to unforeseen circumstances (like traffic, construction delays, accidents, etc.). Hopefully this will help her understand why the use of "never" is incorrect in the interpretation that she selected.

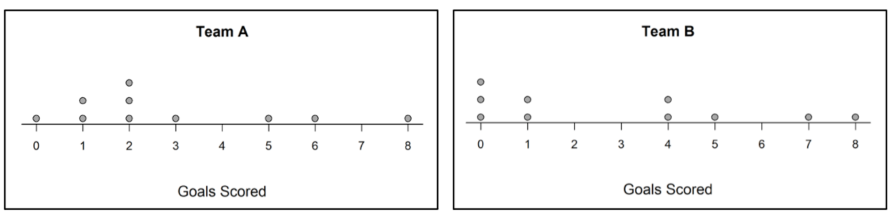
Assessment Problems

The following two problems address ideas explored in the lesson, with a focus on connections to teaching and mathematical content. You can include these problems as part of your usual course quizzes or exams.

Parts of Problem 1 have been adapted from a problem created by the *LOCUS Project* (see `https://locus.statisticseducation.org/professional-development/questions/by-grade/grade-6?field_prodev_level_tid=All&page=6`). The primary purpose of this problem is to assess undergraduates' statistical understanding of computing and interpreting standard deviation.

Assessment Problem 1

1. Two soccer teams will be meeting in the city championship game. Each team played 10 games and averaged 3 goals scored per game for the season. The two dotplots below show the number of goals scored by each team per game for the season.



(a) Describe what one dot in the dotplot for Team A represents.

Solution:
One dot represents how many goals Team A scored in one of the games during the season.

(b) Compute the standard deviation (SD) for Team A.

Solution:

$$\begin{aligned}\bar{x} &= \frac{0+(2)1+(3)2+3+5+6+8}{10}\\ &= \frac{30}{10}\\ &= 3 \text{ goals}\end{aligned}$$

$$\begin{aligned}\text{SD} &= \sqrt{\frac{(0-3)^2+(2)(1-3)^2+(3)(2-3)^2+(3-3)^2+(5-3)^2+(6-3)^2+(8-3)^2}{10-1}}\\ &= \sqrt{\frac{58}{9}}\\ &\approx 2.54 \text{ goals}\end{aligned}$$

(c) Write a sentence that interprets the SD for Team A in the context of the problem.

Sample Responses:

- The SD tells us that Team A's goals scored per game differed from the mean by an average of 2.54 goals.
- The number of goals scored per game is typically 2.54 goals away from the mean, which is 3 goals.
- Since the SD for Team A is 2.54, then that tells us that on average, the goals scored were 2.54 away from the mean.

(d) Based on the dotplots (and without computing the SD for Team B), does Team A or Team B have more variability in the number of goals scored per game over the course of the season? Explain your reasoning.

Sample Responses:

- Team B has more games with goal counts further from the mean (3 goals), so it seems like Team B's season had more variability than Team A's.
- Team B has more variability. Visually, there are more data points at the extreme left, which indicates more variability from the mean.
- Team B has more variability because there are more numbers that are further away from the mean, making the SD larger.

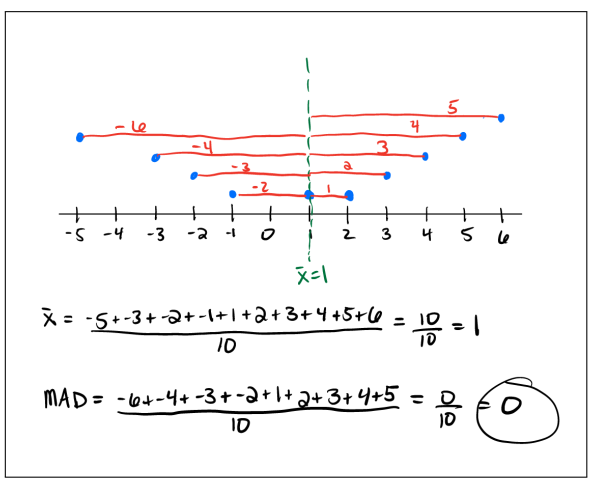
The purpose of Problem 2 is to assess undergraduates' knowledge of (1) examining hypothetical student work and identifying what the student does and does not yet understand about mean absolute deviation, and (2) posing questions to help guide the student's statistical understanding.

Assessment Problem 2

2. Delia was given the following dataset and asked to compute the MAD.

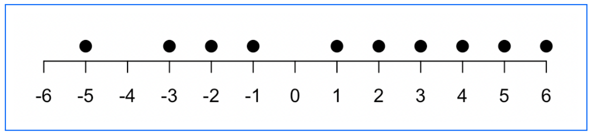
$$1, 5, 2, 6, -2, 3, 4, -1, -3, -5$$

She incorrectly computed the MAD, as shown in her work below.



(a) Draw a dotplot to graphically display the data. Explain how your dotplot shows that the MAD cannot equal 0.

Sample Response:



If the MAD is 0, then all dots on the dotplot should have the same value (like all 1's). But the dotplot clearly shows that variability is present in the data because the dots are spread out. Thus, the MAD can't be 0.

(b) Examine Delia's work and describe what she understands about MAD.

Sample Responses:

- Delia understands how to identify and compute deviations from individual data points to the mean of the dataset.
- Delia understands that computing the MAD involves using the mean of a dataset and deviations.
- Delia understands that MAD is an exact average of deviations.

(c) Identify the error(s) Delia made in her work.

Sample Response:

Delia incorrectly used the *signed* deviations (e.g., -3) rather than the absolute value of the deviations in her computation.

(d) Write one question you could ask Delia to help her correct her work. Briefly explain how your question might help guide Delia's statistical understanding of MAD.

Sample Responses:

- "What do you expect the sum of the deviations from the mean to be?" This will help me determine whether the student understands the mean as a balance point. Then I can assess what they conceptually understand about the mean (i.e., balance point) and the MAD (i.e., average distance from the mean).
- "If the MAD is 0, what does that mean in the context of the problem?" This question will hopefully guide Delia to recognize that if the MAD is 0, then no variability should be present in her dotplot.
- "What does MAD stand for and how are you using the deviations in your calculation?" Asking a student what MAD stands for will help me understand if they know they need to compute "absolute deviations." Then I can help guide them to recognize that they are using signed deviations rather than taking the absolute value of the deviations in their calculation of MAD.
- "Can you draw a dotplot where the MAD is 0? How does that dotplot differ from your work here?" These questions will help me assess whether a student visually understands what a MAD of 0 looks like. If they correctly draw a dotplot with a MAD of 0, then they can visually see how the MAD of this dataset should not be 0.

1.6 References

[1] Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., Spangler, D. A. (2020). *Pre-K–12 guidelines for assessment and instruction in statistics education II (GAISE II): A framework for statistics and data science education.* American Statistical Association.

[2] Census at School (Accessed April 7, 2023). `http://ww2.amstat.org/censusatschool/`

[3] delMas, R. C. (2001). What makes the standard deviation larger or smaller? *Statistics Teaching and Resource Library*. Available at `https://www.causeweb.org/cause/archive/repository/StarLibrary/activities/delmas2001/`

[4] *LOCUS Project* (Accessed April 7, 2023). `https://locus.statisticseducation.org/professional-development`

[5] National Governors Association Center for Best Practices & Council of Chief State School Officers (2010). *Common Core State Standards for Mathematics.* Authors. Retrieved from `http://www.corestandards.org/`

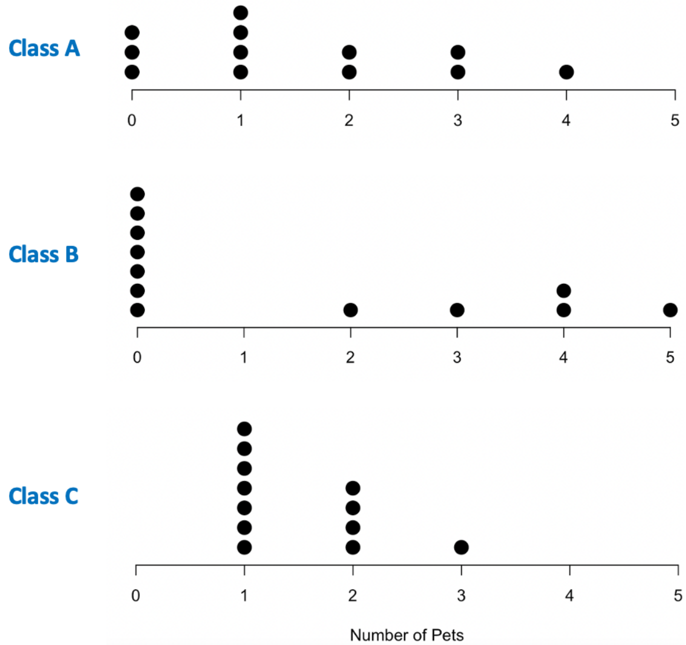
[6] Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). *Mathematical statistics with applications*. Thomson Brooks/Cole.

1.7 Lesson Handouts

Handouts for use during instruction are included on the pages that follow. LaTeX files for these handouts can be downloaded from INSERT URL HERE.

NAME: **PRE-ACTIVITY:** VARIABILITY: MAD AND SD (page 1 of 1)

1. Students from three different classes reported the number of pets in their household. The results are summarized graphically as dotplots and in a frequency table below.



Class A	Class B	Class C
1	0	1
0	0	1
4	0	1
3	4	1
2	4	2
1	3	1
1	0	2
3	5	2
0	0	3
2	0	2
1	0	1
0	2	1

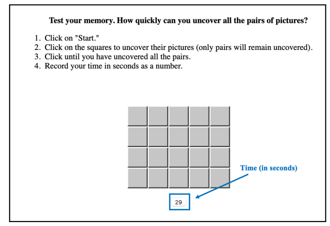
(a) Compute the mean number of pets for each class.

(b) What is similar about the three dotplots?

(c) What is different about the three dotplots?

NAME: **CLASS ACTIVITY:** VARIABILITY: MAD AND SD (page 1 of 6)

1. **Test Your Memory!**



Play the *Census at School Memory Game* where you will need to uncover and match 10 pairs of pictures. The time it takes you to complete the game will be tracked. Go to the following link to access the game: `https://ww2.amstat.org/education/cas/1.cfm`

(a) Play the game once and record your time (in seconds) as a number.

(b) As a class, compile everyone's time in a dataset. What graphical summary would be appropriate to visualize the class's distribution of times on the memory game? Explain your reasoning.

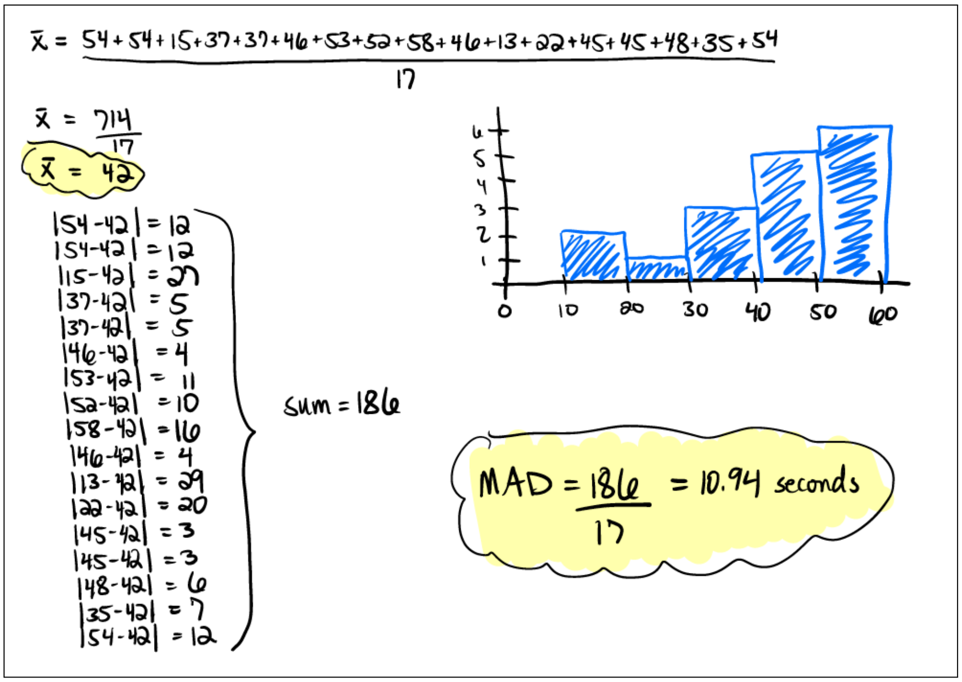
(c) As a class, create a graphical summary to visualize the class's distribution of times on the memory game and sketch it below.

(d) Describe what you notice about the class's distribution of times on the memory game.

2. **Quantifying Variability with Mean Absolute Deviation**
Amaury is teaching a high school intermediate algebra class and his students are learning about different measures of variability. The students in his class played the *Census at School Memory Game* and recorded their times, in seconds:

54, 54, 15, 37, 37, 46, 53, 52, 58, 46, 13, 22, 45, 45, 48, 35, 54

Amaury asked his students to create a graphical summary of the data, compute the mean, and quantify the amount of variability present. One of his students, Jasmine, did the following:



Jasmine, recalling what they learned in middle school, quantified the amount of variability by computing the **mean absolute deviation (MAD)**. All of their calculations are correct. Describe mathematically what Jasmine did to compute the MAD.

3. **Interpreting Mean Absolute Deviation**
Two other students, Tarryn and Benny, also correctly computed the MAD. When Amaury asked his students to write a sentence interpreting their measure of variability in the context of the problem, Jasmine, Tarryn, and Benny wrote the following sentences:

Jasmine	The MAD is 10.94 seconds.
Tarryn	On average, the memory game times were 10.94 seconds away from the mean of 42 seconds.
Benny	A memory game time is 10.94 from the mean.

(a) One student correctly (and completely) interpreted the MAD in the context of the problem. Identify who it was, and describe what components of their interpretation make it correct and complete.

(b) The other two students gave incomplete interpretations of the MAD. Based on their interpretations, describe what each

i. may understand about interpreting the MAD, and

ii. may not yet understand about interpreting the MAD.

(c) In a general context, describe what MAD measures.

4. **Quantifying Variability with Standard Deviation and Interpreting Standard Deviation**
Josief, another student in Amaury's class, did something different, as shown below:

$$\bar{X} = \frac{54+54+15+37+37+46+53+52+58+46+13+22+45+45+48+35+54}{17}$$

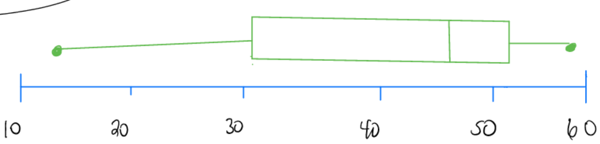
$$\bar{X} = 42$$

$$sd \Rightarrow (54-42)^2+(54-42)^2+(15-42)^2+(37-42)^2+(37-42)^2+(46-42)^2+(53-42)^2$$
$$+(52-42)^2+(58-42)^2+(46-42)^2+(13-42)^2+(22-42)^2+(45-42)^2+(45-42)^2$$
$$+(48-42)^2+(35-42)^2+(54-42)^2$$

$$= 144+144+729+25+25+16+121+100+256+16+841+400+9+9$$
$$+36+49+144$$

$$= 3064$$

$$SD, \quad S_x = \sqrt{\frac{3064}{16}} \approx 13.84$$



(a) Josief quantified the amount of variability by computing the **standard deviation (SD)**. All of his calculations are correct. Describe mathematically what he did to compute the SD.

(b) When asked to write a sentence to interpret the SD in the context of the problem, Josief wrote the following:

> The memory game times varied by 13.84 seconds.

Describe why Josief's interpretation is not completely correct.

(c) Consider the following questions that one might ask Josief to help him with his interpretation of the standard deviation (in the context of the memory game times).

i. Explain how the following question might help Josief to advance in his understanding of interpreting standard deviation in the context of a problem:

Can you say more about how the memory game times varied?

ii. Explain how the following question might help you assess what Josief understands about interpreting the standard deviation in the context of a problem:

What does standard deviation measure?

iii. Explain why the following question might not help Josief:

Why is your interpretation incorrect?

(d) In a general context, describe what SD measures.

5. Return to the class dataset from Problem 1.

   (a) Compute the mean absolute deviation of your class's data of memory game times and write a sentence interpreting the mean absolute deviation in the context of the problem.

   (b) Compute the standard deviation of your class's data of memory game times and write a sentence interpreting the standard deviation in the context of the problem.

NAME: **HOMEWORK PROBLEMS:** VARIABILITY: MAD AND SD (page 1 of 2)

1. Ten movies were randomly selected and the length of each movie (in minutes) is given below.

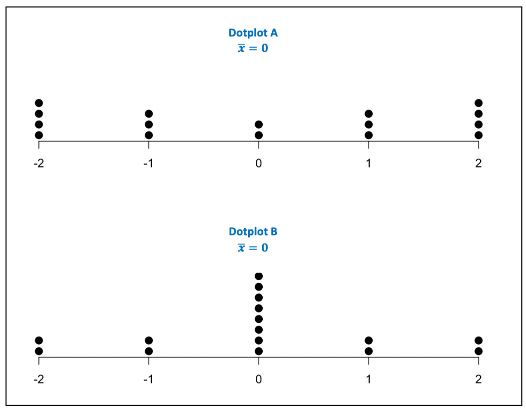
   152, 156, 98, 173, 68, 122, 92, 105, 138, 126

   (a) Compute the mean absolute deviation (MAD) and write a sentence that interprets the MAD in the context of the problem.

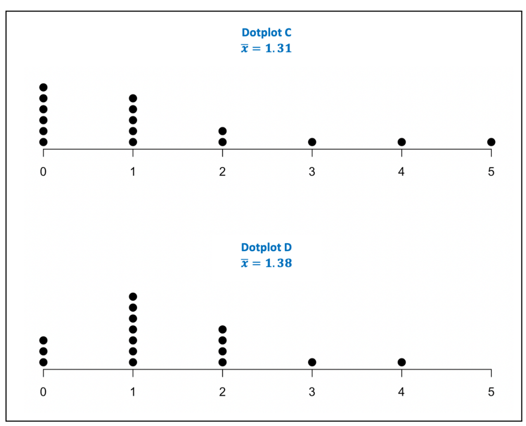
   (b) Compute the standard deviation (SD) and write a sentence that interprets the SD in the context of the problem.

2. In this problem, you will *visually* compare the mean absolute deviation (MAD) between pairs of dotplots.

   (a) **Dotplots A and B**.
   Without doing any calculations, identify which dotplot (A or B) has the *larger* MAD. Explain your reasoning.



   (b) **Dotplots C and D**.
   Without doing any calculations, identify which dotplot (C or D) has the *larger* MAD. Explain your reasoning.



   (c) Draw two different dotplots that have the same MAD. Describe your thought process when creating these two different dotplots.

3. Consider the mean absolute deviation (MAD) and the standard deviation (SD). Typically, MAD is first taught in middle school and SD is taught in high school. Describe why it is helpful for school students to learn MAD before SD.

4. Four students (Daveed, Monica, Juliana, and Bryant) are working on the following problem together, but they all pick a different answer.

> The director of the City Transportation System is interested in the amount of time required for a bus to make the trip from Downtown Station to City Mall. After collecting data for several months by recording the time it takes to make the trip, she finds that the distribution of times has a standard deviation of 3 minutes.
>
> Which of the following is the best interpretation of the standard deviation?
>
> A. A bus that leaves from Downtown Station typically arrives at City Mall 3 minutes later than the scheduled time.
>
> B. A bus typically takes about 3 minutes to get from Downtown Station to City Mall.
>
> C. The time a bus takes to get from Downtown Station to City Mall never varies more than 3 minutes from the mean trip time.
>
> D. The difference between the actual time a bus takes to get from Downtown Station to City Mall and the mean trip time is, on average, about 3 minutes.

Daveed selects option A, Monica selects option D, Juliana selects option C, and Bryant selects option B.

(a) Who selected the correct answer?

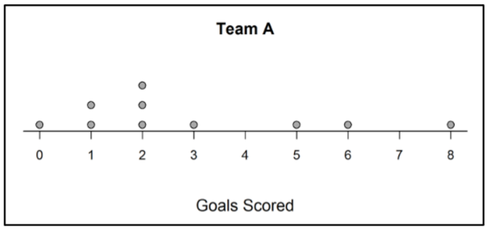
(b) For *each* student who selected an incorrect answer, examine the choice they selected and describe a statistical concept they do understand.

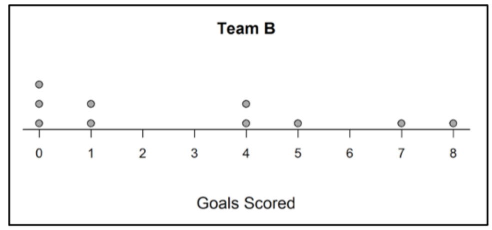
(c) For *each* student who selected an incorrect answer, examine the choice they selected and describe a statistical concept they might not yet fully understand.

(d) For *each* student who selected an incorrect answer, write a question you could ask them to help guide their understanding of interpreting a standard deviation in the context of the problem. Briefly explain how your question may help guide their statistical understanding.

NAME: **ASSESSMENT PROBLEMS:** VARIABILITY: MAD AND SD (page 1 of 3)

1. Two soccer teams will be meeting in the city championship game. Each team played 10 games and averaged 3 goals scored per game for the season. The two dotplots below show the number of goals scored by each team per game for the season.





(a) Describe what one dot in the dotplot for Team A represents.

(b) Compute the standard deviation (SD) for Team A.

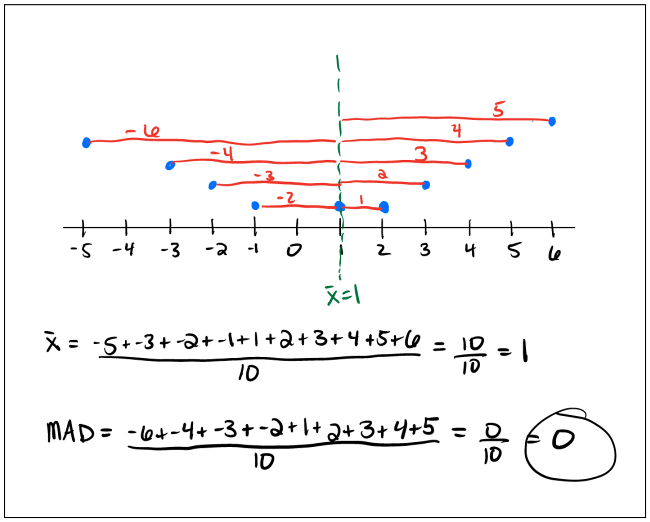
(c) Write a sentence that interprets the SD for Team A in the context of the problem.

(d) Based on the dotplots (and without computing the SD for Team B), does Team A or Team B have more variability in the number of goals scored per game over the course of the season? Explain your reasoning.

2. Delia was given the following dataset and asked to compute the MAD.

$$1, 5, 2, 6\ , -2, 3, 4, -1, -3, -5$$

She incorrectly computed the MAD, as shown in her work below.



(a) Draw a dotplot to graphically display the data. Explain how the dotplot shows that the MAD cannot equal 0.

(b) Examine Delia's work and describe what she understands about MAD.

(c) Identify the error(s) Delia made in her work.

(d) Write one question you could ask Delia to help her correct her work. Briefly explain how your question might help guide Delia's statistical understanding of MAD.