Choosing Samples

About Illustrations: Illustrations of the Standards for Mathematical Practice (SMP) consist of several pieces, including a mathematics task, student dialogue, mathematical overview, teacher reflection questions, and student materials. While the primary use of Illustrations is for teacher learning about the SMP, some components may be used in the classroom with students. These include the mathematics task, student dialogue, and student materials. For additional Illustrations or to learn about a professional development curriculum centered around the use of Illustrations, please visit mathpractices.edc.org.

About the *Choosing Samples* Illustration: This Illustration's student dialogue shows the conversation among three students who are investigating what samples of 5 rectangles will give them the best estimate of the average area of a set of 100 rectangles. They generate and test two ideas for how to take samples—having their peers choose 5 random numbers and using those numbers to select samples and having their peers choose 5 rectangles by looking at the rectangles—and then they begin to discuss the difference in the two estimates their two methods generate.

Highlighted Standard(s) for Mathematical Practice (MP)
MP 1: Make sense of problems and persevere in solving them.
MP 3: Construct viable arguments and critique the reasoning of others.
MP 6: Attend to precision.

Target Grade Level: Grades 6–7

Target Content Domain: Statistics and Probability

Highlighted Standard(s) for Mathematical Content

7.SP.A.1 Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.

7.SP.A.2 Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions. *For example, estimate the mean word length in a book by randomly sampling words from the book; predict the winner of a school election based on randomly sampled survey data. Gauge how far off the estimate or prediction might be.*

6.SP.B.5c Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

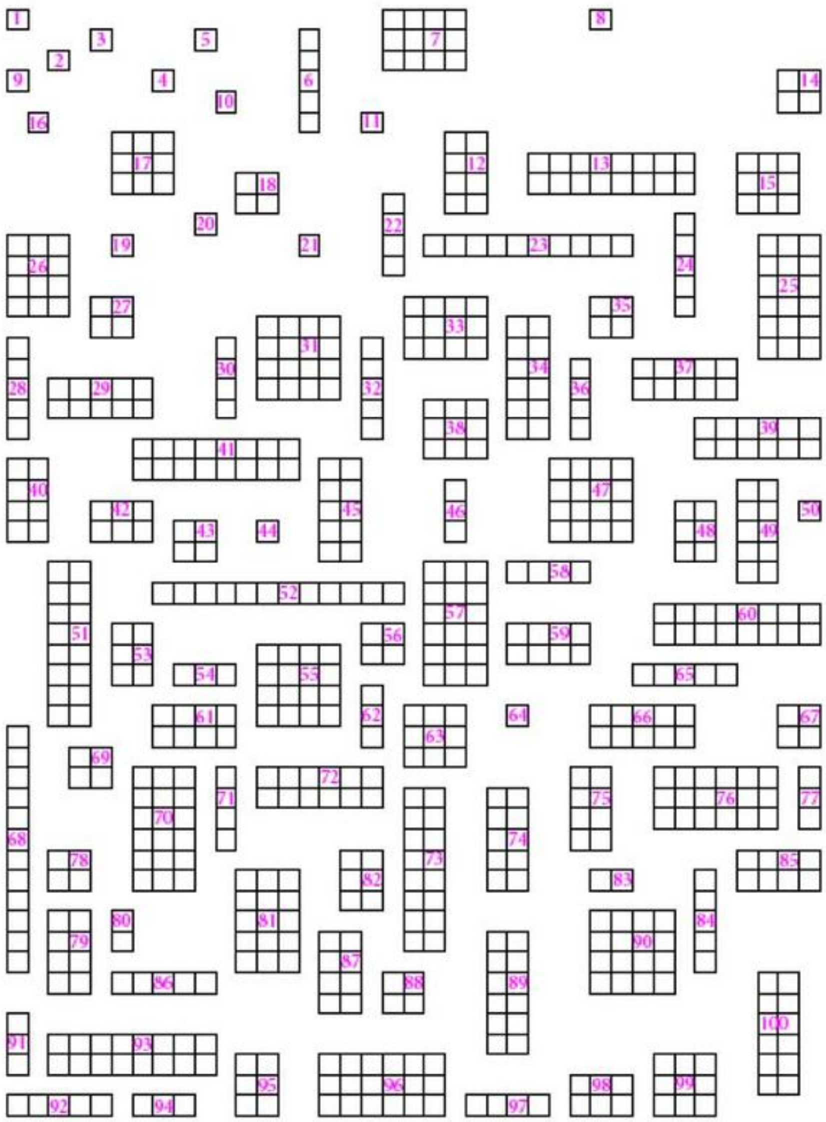
Math Topic Keywords: sample, random sampling, population, inference, area

This material is based on work supported by the National Science Foundation under Grant No. DRL-1119163. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Mathematics Task

Suggested Use
This mathematics task is intended to encourage the use of mathematical practices. Keep track of ideas, strategies, and questions that you pursue as you work on the task. Also reflect on the mathematical practices you used when working on this task.

Estimate the average area of the 100 rectangles below using samples of only five rectangles.



Task Source: Adapted from Barbella, P., Kepner, J., & Scheaffer, R. (1994). *Exploring measurements*. Palo Alto, CA: Dale Seymour Publications.

Student Dialogue

Suggested Use
The dialogue shows one way that students might engage in the mathematical practices as they work on the mathematics task from this Illustration. Read the student dialogue and identify the ideas, strategies, and questions that the students pursue as they work on the task.

Students are at the beginning of a unit on sampling. They are trying to discover what the best way is to choose a sample and what some possible challenges are in doing so. The task has enough rectangles to make a sampling approach possible while still giving students the opportunity to find the average area for the population as a check to the sampling methods they develop. In this way, students are able to try different sampling methods and experimentally verify which produces a better estimate.

(1) Dana: We're supposed to estimate the average area of all the rectangles using only five rectangles, so let's pick 5, average their areas, and we're done.

(2) Sam: Average?! Does this mean the *mean*? Or the median? And how do we decide which 5 to pick? Are some choices better than others? They're very different and we only have 5 to choose.

(3) Dana: I guess it means the mean. And does it even matter which 5 we pick? The question only asks for an estimate.

(4) Sam: True, but an estimate can be better or worse. We want the best we can get, don't we?

(5) Anita: But the question doesn't say we have to use only one sample of five rectangles. It says using *samples* of five rectangles. What if we pick a whole bunch of 5-rectangle samples?

(6) Dana: Good idea! I'm sure that would get us closer to the real average.

(7) Anita: So let's ask everyone in the class to pick five rectangles for us, and we can use those samples to estimate the average area of all the rectangles.

(8) Sam: Wait! Isn't that sort of silly? There are 20 of us in this class. If we each pick a *different* set of five rectangles, we'd have picked them all! Let each person find their own average and then just average those 20 results and we're done! Who cares about the sampling?

(9) Anita: Ooh! That raises a good question, Sam. Would the average of those 20 averages give the same result as getting the average of one set of 100? Lemme think a bit.

(10) Sam: Anyway, that still doesn't answer if there are better or worse ways to choose samples of five rectangles.

(11) Dana: We could always ask people how they chose the rectangles they did.

(12) Anita: Or why don't we try two totally different approaches. We can make each kid choose a sample of rectangles by picking 5 different random numbers between 1 and 100 from a hat, and we can compare that to what happens if we show the rectangles and ask them to pick 5 different rectangles visually.

(13) Sam: Well, that only tests two different ways, and it's not clear why we'd even expect those two ways to generate different results, but I like that plan. Let's go survey the whole class.

[Sam, Dana, and Anita survey all 20 students in the class, asking them to first choose five numbers, without replacement, between 1–100 from a hat and then choose five different rectangles. Next they calculate the average area from each sample, and Dana puts it all in a table.]

(14) Dana: So here's the table. It has the averages of all the samples. *[Dana shows the table they made.]*

(15) Sam: Sorry. Say one more time, just to be clear, what exactly do we have here?

(16) Dana: OK, each kid picked five different random numbers from a hat. Then they averaged the areas of the rectangles that were labeled with those numbers. Those averages are listed in one column. Then they looked at the whole set of rectangles and tried to choose a set of 5 that they think will be a good representative sample. The averages of *those* sets of five are in the other column. And now we get to see which is better!

Person	Average area based on samples chosen by random numbers	Average area based on samples chosen by hand-picking rectangles
1	5	7.6
2	6.8	15.2
3	5	4.2
4	5	9.8
5	7	10.8
6	6.4	5.4
7	7.8	13.6
8	10.8	10.6
9	4.2	14.4
10	4.8	9.8
11	4.6	10.8
12	4.4	8.2
13	3	10.4
14	8	7.6
15	7.8	8.2

16	3.2	11.2
17	8.2	11
18	6.6	15.4
19	5.6	6.4
20	7.8	7.8

(17) Anita: Wow, these vary a lot! The random column varies from 3 to… it looks like 10.8 is the largest. And the hand-picked column varies from… Is 4.2 the lowest? All the way to, I think, 15.2. Oh, no, 15.4. That's an even larger range! Let's see what the mean values are.

(18) Dana: Yes, but we need more than just the mean values because, you're right, these really do vary a lot. We also need to look at that variation beyond just looking at range. The mean absolute deviation would help us describe the variation of each sampling method.

(19) Sam: Yup. Good idea, Dana. But we still need to figure out the mean for each column first, so let's do that. *[Two of them set to work to calculate the averages.]* The samples chosen by random numbers have an average area of 6.1 and the samples chosen by looking and trying to pick a representative set have an average area of 9.92.

(20) Anita: OK, so now we need to find the difference between each sample's average area and those column averages.

(21) Dana: Right, and we want the absolute value of the differences….

[Students take several minutes to calculate the mean absolute deviation and summarize their findings in a table.]

	Samples chosen by random numbers	Samples chosen by hand-picking rectangles
Mean	6.1	9.92
Mean absolute deviation	1.62	2.42

(22) Sam: OK, so we have this table. What is it really telling us?

(23) Dana: Well, hmmm… Looking just at the mean absolute deviation tells us that we're getting lots more agreement, lots less variation, with the random number method than when we picked by hand.

(24) Anita: So when we chose samples using random numbers we got an average area of 6.1 for the rectangles; however, samples differ by an average of 1.62 so really the samples' averages were anywhere between… *[calculates for a minute]* 4.48 and 7.72. And when we chose samples by hand, the average area was 9.92 but sample averages could vary from 5.5 to 12.34.

(25) Sam: Well, no. That's not what it means. We had sample averages outside of those intervals in both sampling methods! Look at the numbers in our first table.

(26) Dana: Right! But mean absolute deviation does give an idea of the *average* variation of samples. It's telling us the average distance a sample will be from the mean. All we can say is that sample averages were closer together when we used random numbers than when we handpicked rectangles. Do Anita's limits tell us where *most* of the sample averages will fall?

(27) Sam: I don't know, Dana, but it does sound like using the random numbers sampling method gives us more consistent results. What does that mean in terms of which sampling method is better? Couldn't the handpicking method end up giving us the better estimate even if there is more variation?

(28) Dana: Should we just average the two means for the sampling methods and get something around 8?

(29) Anita: Yeah, we could, but why *would* we? One of these might already be best and we'd be making it worse. We need some theory before we just take averages.

(30) Sam: So this does mean one method must be better than the other….

(31) Anita: No, not necessarily.

(32) Sam: Right, they could be equally wrong. But I have an idea. Let's check to see which method is better.

(33) Dana: How?

(34) Sam: Let's just find the actual average area of all 100 rectangles and see which of the two sampling methods gave us the closer estimate.

(35) Dana: I guess we could. This seems like a lot of work just for an estimate, though.

(36) Anita: Well, it won't *be* an estimate any more. It'll be the actual average, but that might also help us figure out whether the average of averages is the average.

(37) Sam: I don't know what you just said, Anita, but I'm glad you're happy. I agree with you, Dana, that this'll be a lot of work, but I'm curious now to find out which method is best.

Teacher Reflection Questions

Suggested Use
These teacher reflection questions are intended to prompt thinking about 1) the mathematical practices, 2) the mathematical content that relates to and extends the mathematics task in this Illustration, 3) student thinking, and 4) teaching practices. Reflect on each of the questions, referring to the student dialogue as needed. Please note that some of the mathematics extension tasks presented in these teacher reflection questions are meant for teacher exploration, to prompt teacher engagement in the mathematical practices, and may not be appropriate for student use.

1. What evidence do you see of students in the dialogue engaging the Standards for Mathematical Practice?

2. In the Student Dialogue, students are asked to "estimate the average area of the 100 rectangles below using samples of only five rectangles." How is this similar to or different from being asked to "estimate the average size (square-footage) of single-family homes in the states of Florida and Texas"? What is the significance of Sam's question in line 2: "How do we decide which 5 to pick? Are some choices better than others?"

3. How might choosing one sample of 10 rectangles instead of one sample of five rectangles using the random sampling method affect the sample's ability to estimate the average area of the 100 rectangles?

4. In line 16, students create a table describing the sampling distribution for the average area of the rectangles using two different sampling methods. The students then describe the mean and the mean absolute deviation for their sampling distributions (line 21). How might increasing the sample size (i.e., the number of rectangles used in each sample) affect (1) the mean and (2) the variability of the distribution using each sampling method?

5. In line 24, Anita seems to think that all the sample means will be less than 1 mean absolute deviation away from the sampling distribution mean. In the Student Dialogue example, this is not true. Can it ever be true?

6. In line 26, Dana seems to think that Anita's idea of creating a range of ± 1 MAD of the sampling distribution mean is likely to contain *most* of the sample averages. Is Dana correct? Explain.

7. In line 9, Anita asks if "the average of those 20 averages gives the same result as getting the average of one set of 100"? What is the answer to Anita's question?

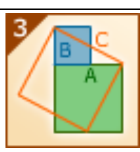
8. Revisit Anita's question about the average of averages from line 9. What would happen if the groups were of unequal size? Would the average of those groups' averages be equal to the average of the population?

9. How might you help students make sense of how the average of averages of unequal groups relates to the overall average across groups *without* working through numerical examples?

Mathematical Overview

Suggested Use
The mathematical overview provides a perspective on 1) how students in the dialogue engaged in the mathematical practices and 2) the mathematical content and its extensions. Read the mathematical overview and reflect on any questions or thoughts it provokes.

Commentary on the Student Thinking

Mathematical Practice	**Evidence**
 Make sense of problems and persevere in solving them.	The task asks students to "estimate the average area" and sets a restriction "using samples of only five," but provides no other guidance: no specification of how many samples, how to select the samples, how (if at all) to assess the accuracy of the estimate, or even what "average" to use. Students in this Student Dialogue make sense of the problem partly by making sense of the situation, concluding that the problem makes sense only if they are trying to find a sampling method that produces the best results. In lines 1–5, students are "explaining to themselves the meaning of a problem" by "analyz[ing] givens." Sam also raises the important question, "How do we decide which five to pick? Are some choices better than others?" (line 2), and this guides the students' exploration of the problem. In lines 7–13, students "plan a solution pathway rather than simply jumping into a solution attempt." They develop two different ways of choosing samples and are open-minded, noting that these may or may not lead to different solutions. They also check whether their solution path makes sense in line 8, when Sam notes that choosing 20 disjoint samples of five rectangles would lead to a total of 100 rectangles, the same number in the entire population of triangles. Even after two different average areas are calculated using their two methods, students come up with a way to "check their answers to problems using a different method." Students realize in lines 26–28 that they could calculate the population mean as a way to judge which sampling method yields the more accurate estimate.
 Construct viable arguments and critique the reasoning of others.	As students interpret the results of the estimates produced by their two sampling methods, they are proposing different ways to handle the two estimates. During this process students also critique one another's ideas and "respond to the arguments of others." For example, in lines 28 and 29, Dana proposes averaging the two estimates for average area, and Anita responds by saying that if one method is better than the other, averaging could decrease accuracy. Later, Sam and Anita argue about whether they can be sure that one estimate must be best or whether it is possible that both estimates "could be equally wrong" (line 32).

Attend to precision.	In the Student Dialogue, students “use clear definitions in discussion with others and in their own reasoning.” They begin in lines 2–3 when they question what is meant by average, and they establish a common understanding. Similarly, they use their definition of mean absolute deviation (lines 18–21) to question what that calculation tells them about the two sampling methods (lines 22–26). Not only are the students thinking about precision of vocabulary and the meaning of results, but they are also considering the precision of their sampling method. The entire Student Dialogue centers on the question, What sampling method gives the most precise estimate of a population parameter (in this case, the average area)? The students keep that question in mind throughout the Student Dialogue (lines 16, 27, 32).

Commentary on the Mathematics

The deep question behind an exercise like this is, “What makes a ‘good’ sample, one that reasonably faithfully represents the population from which it is drawn?” (A secondary but important question is also raised in the Student Dialogue about averaging averages.)

When populations are small, one doesn’t need fancy statistical techniques. If you want to know the favorite kind of vacation spot, candidate, or sweater color, just ask everyone. How you handle the diversity of answers may still be a problem, but it’s not a statistical one. And finding a mean or median, if such things are appropriate, also does not require sampling. You have *all* the information and can use it.

When the numbers are large, getting *all* the information is not practical, and so opinion research, market research, effectiveness research, and so on depend on estimates. That requires attention to the sampling technique and statistical methods to ensure validity.

With a mere 100 data points, the students can divide and conquer as Sam points out in lines 8 and 34, and get an exact answer; they don’t need statistical thinking and don’t need to estimate. This problem is an *exercise* designed deliberately to be small enough to allow students to get an exact answer, so that they can use that to explore the way their sampling technique affects the estimate they get. Furthermore, the problem constrains the *size* of the sample, so the remaining variable is only the selection method. Students aren’t, for example, comparing the merits of larger or smaller samples.

Though Sam does raise the question (line 8) of whether or not sampling is even needed, the students don’t discuss anything about the purpose of sampling. They just take the problem “as given” and work from there.

In line 9, Anita raises a question about the mathematics *behind* the statistical technique they call averaging (finding the mean). The students don’t tackle this question within the brief snapshot of their activities represented in the Student Dialogue, but it comes up in Teacher Reflection Questions 7 and 8. In lines 17–27, the students notice that the two sampling methods that they used gave sample averages that “vary a lot” (line 17), and they proceed to calculate and interpret the mean absolute deviation of samples collected using each sampling method. Finally, they decide on a concrete check but, in both a scientific and mathematical sense, the work is really not done even when they identify which of their methods came closer by computing the *actual* mean. Without explicitly hypothesizing *why* one technique might bias the result differently from the other, they will learn only *that* one technique came closer. As these students progress in their studies, they will come to learn both why random sampling provides the better estimate (as a result of the Central Limit Theorem) and why random sampling is useful, namely that it gives you the ability to generalize to the population based on sample information. This is especially helpful in scenarios where calculating a population mean is not feasible and using random sampling allows us to more accurately estimate a mean while controlling for some unwanted sources of variation.

Evidence of the Content Standards

The students in this Student Dialogue are using samples (both random and handpicked) to make inferences regarding the average area of the population of rectangles (7.SP.A.1 and 7.SP.A.2). In the process, they use measures of center (both sample means and sampling distribution means) and measures of variability (mean absolute deviation), and they interpret these measures (6.SP.B.5c).

Choosing Samples

Student Materials

Suggested Use
Student discussion questions and related mathematics tasks are supplementary materials intended for optional classroom use with students. If you choose to use the mathematics task and student dialogue with your students, the discussion questions can stimulate student conversation and further exploration of the mathematics. Related mathematics tasks provide students an opportunity to engage in the mathematical practices as they connect to content that is similar to, or an extension of, that found in the mathematics task featured in the student dialogue.

Student Discussion Questions

1. Two sampling methods are compared in the Student Dialogue: choosing samples at random (with random numbers) and choosing them by looking deliberately for what seems to be the most representative set. If you were required to choose *one* sample of exactly 10 rectangles to use for estimating the average area, decide which method—random choice or handpicking—*you* think is a "better" way to get a good estimate of the actual average of all 100 rectangles. Explain your choice.

2. In the Student Dialogue, students are asked to "estimate the average area of the 100 rectangles below using samples of only five rectangles." How is this similar to or different from being asked to "estimate the average size (square-footage) of single-family homes in the states of Florida and Texas" or being asked to conduct a poll to find out which presidential candidate most Americans favor for an election?

3. How do you think the estimated average area of the 100 rectangles would be different if students used samples of 10 instead of samples of 5?

4. In line 9, Anita imagines the 20 students in class dividing up the 100 rectangles, each taking a different set of five, and computing their averages. She then poses the question, "Would the average of those 20 averages give the same result as getting the average of one set of 100"? Just using intuition, and without working this out, what do you *think* is the answer to Anita's question? You'll get a chance to investigate more deeply later.

Related Mathematics Task

1. What is the average area of all 100 rectangles? Which sampling method provided the better estimate: using random numbers or picking rectangles by hand? Why do you think that method works better than the other?

2. Imagine you are asked to survey the students in your class for their favorite sport. You are asked to use a sample of five students only. Is it best to choose five classmates by (A) picking five names out of a bag, (B) picking five of your friends, or (C) picking the first five classmates that walk in the room? Explain your reasoning.

3. In line 24, Anita generates an interval around the mean, using the mean absolute deviation to define its limits. In line 26, Dana asks if "Anita's limits tell us where *most* of the sample averages will fall." Do they? Explain why or why not.

4. Split the 100 rectangles into disjoint groups of five (rectangles 1–5, rectangles 6–10, etc.) similar to Sam's suggestion in line 8. Find the average area of each of the smaller groups. Take the average of those averages. How does the average of the averages compare with the average area of the 100 rectangles found in Question 1?

5. Split the 100 rectangles in 10 disjoint groups of five (rectangles 1–5, rectangles 6–10, etc.) and five disjoint groups of 10 (rectangles 51–60, rectangles 61–70, etc.). Find the average area of each of the smaller groups. Take the average of those averages. How does the average of the averages compare with the average area of the 100 rectangles found in Question 1?

6. In Questions 4 and 5, you experimented with taking the average of averages for both groups of equal and unequal sizes. What did you notice about the average of averages and why do you think this happens? To help in your response, consider the following:
   A) Imagine you have 100 rectangles that are separated into 20 equal groups. The average areas of the 20 groups are $a_1, a_2, a_3, \ldots, a_{20}$. Show algebraically that the average area of the 20 groups is the same as the average area of the 100 rectangles.
   B) Imagine you have 100 rectangles that are separated into two groups: one group with 1 rectangle whose area is 10 and one group with 99 rectangles each with an area of 1. How does the average area of the 100 rectangles compare with the average of the average area of the two groups?

7. Two employees of the Census Bureau are trying to calculate the average age for the population across the 50 states. Employee A says they need to take the average of everyone's age across the 50 states. Employee B says they can use a shortcut by averaging the average age for each state. What do you think of the two employee's proposals?

8. Budget cuts at the US Census Bureau prevent the agency from collecting *every* person's age (along with other information). Propose a way to estimate the average age of people living in the United States.

Answer Key

Suggested Use
The possible responses provided to the teacher reflection questions are meant to be used as an additional perspective after teachers have thought about those questions themselves. The possible responses to the student discussion questions and related mathematics tasks are intended to help teachers prepare for using the student materials in the classroom.

Possible Responses to Teacher Reflection Questions

1. What evidence do you see of students in the dialogue engaging the Standards for Mathematical Practice?

    Refer to the Mathematical Overview for notes related to this question.

2. In the Student Dialogue, students are asked to "estimate the average area of the 100 rectangles below using samples of only five rectangles." How is this similar to or different from being asked to "estimate the average size (square-footage) of single-family homes in the states of Florida and Texas"? What is the significance of Sam's question in line 2: "How do we decide which 5 to pick? Are some choices better than others?"

    In the Student Dialogue, students have access to complete information about all 100 rectangles, so they can compute the exact average and don't need *any* sampling technique nor do they need to estimate. The single-family homes situation is very different. While it might theoretically be possible to find every single-family home in Florida and Texas, measure its square-footage (the area of its floor plan), and calculate the average, that is vastly impractical. In fact, during the time it would take to do this, some dwellings might be built or demolished, so it is, in a sense, not even *theoretically* possible to get the exact figures. If a practical study could be done with, say, a sample of 1,000 dwellings, then one really does care about finding a way to choose that sample to get the best estimate.

    In the Student Dialogue, students use samples because the assignment requires it, not because the situation does. This, in fact, is the significance of Sam's statement in line 8. In real life, we often cannot study all the members of a population due to cost, time, access, or other constraints, which is why we draw a sample. By studying a sample we are then able to make generalizations about a population. The significance of Sam's question about how to choose a sample is an important one, given that a sample is meant to allow us to generalize to the entire population, within a level of certainty. Using some form of random sampling is often viewed as the best way to choose a sample since it decreases possible bias.

3. How might choosing one sample of 10 rectangles instead of one sample of five rectangles using the random sampling method affect the sample's ability to estimate the average area of the 100 rectangles?

   You might turn this question into an exploration by actually choosing samples of 10 versus samples of 5 using both sampling methods. If the samples are unbiased (as in the case of random sampling), the larger sample will give the better estimate. You can see why by considering the analogy of trying to find the average height of players in the NBA. The more people you measure, the closer you are to measuring all the players and going from an estimate to the actual average height. However, if the sample is biased (which may occur in the hand-picking method used in the Student Dialogue), then taking a larger sample may not increase the sample's ability to estimate the average area of the rectangles. The larger sample could skew the estimate in the direction of the bias.

4. In line 16, students create a table describing the sampling distribution for the average area of the rectangles using two different sampling methods. The students then describe the mean and the mean absolute deviation for their sampling distributions (line 21). How might increasing the sample size (i.e., the number of rectangles used in each sample) affect (1) the mean and (2) the variability of the distribution using each sampling method?

   In the random-sampling method, the sampling distribution mean would get closer to the true population mean as sample size increases. The variability would decrease as a result of the Central Limit Theorem.

   In the handpicked rectangles sampling method, the sampling distribution mean would not necessarily get closer to the true population mean since each sample may have a bias and the sampling distribution may center around a biased sampling distribution mean. If the handpicked rectangles sampling method had no bias, then increasing the sample size would most likely improve the estimate of the population mean, because it will most likely decrease the variation within a sampling distribution of sample means regardless of sampling bias.

5. In line 24, Anita seems to think that all the sample means will be less than 1 mean absolute deviation away from the sampling distribution mean. In the Student Dialogue example, this is not true. Can it ever be true?

   Mean absolute deviation (MAD) is the average distance between sample means and the sampling distribution mean. The notion of "average" implies that some samples will be closer and some will be farther. Therefore, it can never be the case that all sample means will be less than 1 MAD from the sampling distribution mean.

6. In line 26, Dana seems to think that Anita's idea of creating a range of ± 1 MAD of the sampling distribution mean is likely to contain *most* of the sample averages. Is Dana correct? Explain.

   Dana is not correct in thinking that most sample averages would be within ± 1 MAD of the sampling distribution mean. Looking at the table from line 16, we do see that a majority (14 out of 20) of the sample averages from the "hand-picking rectangles" column are within ± 1 MAD of the sampling distribution mean of that column. However looking at the "random numbers" column, only half (10 out of 20) of the sample averages are within ± 1 MAD of the sampling distribution mean of that column.

   We can also create scenarios where *most* of the sample averages are outside ± 1 MAD of the sampling distribution mean. For a simple case, the sampling distribution mean in the set 10, 10, 10, 19, 21, 30, 30, 30 would be 20 and the MAD would be 7.75. Dana's hypothesis is that most of the sample averages should fall in the interval 12.25 and 27.75, which is clearly not the case.

7. In line 9, Anita asks if "the average of those 20 averages gives the same result as getting the average of one set of 100"? What is the answer to Anita's question?

   Because the sets are all the same size, and because they are disjointed, the answer to Anita's question is yes. You can think of it this way. Name the 20 averages that the students found $a_1, a_2, a_3, a_4, ..., a_{20}$. Each one of those 20 averages was generated by dividing a sum of areas by 5. So, for example, $5a_1$ is the sum of the areas of the five rectangles in the first sample. Thus, $5(a_1 + a_2 + a_3 + a_4 + ... + a_{20})$ is the sum of the areas of all 100 rectangles. Divide that by 100 to get the average area of all 100 rectangles.

   $$\frac{5(a_1 + a_2 + a_3 + a_4 + ... + a_{20})}{100} = \frac{a_1 + a_2 + a_3 + a_4 + ... + a_{20}}{20}$$

   This is the same result that is achieved by adding the 20 averages ($a_1 + a_2 + a_3 + a_4 + ... + a_{20}$) and dividing that by 20. To see a numerical example of averaging the averages of sets of equal size, look at Related Mathematics Task 4.

8. Revisit Anita's question about the average of averages from line 9. What would happen if the groups were of unequal size? Would the average of those groups' averages be equal to the average of the population?

   If the samples were *not* all the same size, the answer to Anita's question would be no in most cases: we could not rely on an "average of the averages of subsets" to be the same as the average of the whole set. An easy counter example: Imagine partitioning the hundred rectangles into two groups, one of which contained just a single rectangle, and the other of which contained the 99 others. We should expect the average of the 99 to be quite close to the average of all 100. Let's suppose that the average of the 99 is about 7.48 and the one missing rectangle has an area of 1. The average of those two averages is 4.24, *very* different from the estimate given by the average of the 99. In fact, if the *average* of the 99 is, say, 7.48, then the *sum* of those 99 areas is 99 × 7.48 = 740.52, so the sum of all 100 is 741.52, and therefore the *average* of all 100 would be 7.42. To see a numerical example of averaging the averages of sets of unequal size, look at Related Mathematics Task 5.

There are, however, special cases when the average of averages of different sized subsets is the same as the average of the whole set. For example, if all the rectangles in a population are the same size then it does not matter how you subdivide them and what average of averages you calculate—the result will always be the same.

9. How might you help students make sense of how the average of averages of unequal groups relates to the overall average across groups *without* working through numerical examples?

   One way to make sense of the average of averages of unequal groups is by looking at extreme cases, just as we did for Question 7. For example you may ask students to consider a scenario where you have 1,000,000 rectangles each with an area of 10 and 1 rectangle with an area of 1. If you take the average area of these two unequal groups of rectangles, you would get 10 and 1. Averaging those two values would give you 5.5, which is definitely not the average area across all 1,000,001 rectangles. Looking at extreme cases often helps evaluate claims, partly by exaggerating the effects of whatever it is that one is testing. It can also focus attention on the *logic* of the situation, allowing students to draw a conclusion without getting lost in calculations.

Possible Responses to Student Discussion Questions

1. Two sampling methods are compared in the Student Dialogue: choosing samples at random (with random numbers) and choosing them by looking deliberately for what seems to be the most representative set. If you were required to choose *one* sample of exactly 10 rectangles to use for estimating the average area, decide which method—random choice or handpicking—*you* think is a "better" way to get a good estimate of the actual average of all 100 rectangles. Explain your choice.

   Students may, of course, have differing opinions as to which method is best. They could argue that if they have only one sample to work from, a random choice might, by accident, get a badly biased sample but, by choosing deliberately, they could pick some small ones, some large ones, and be "less biased." Or they could argue that random sampling decreases the likelihood of bias from the person choosing samples. Both arguments make sense, but, in general, deliberate choice is not nearly as representative and unbiased as we like to think it is, and random sampling is often best.

   Note that in the Related Mathematics Tasks, one question asks students to find the average area of the 100 rectangles and determine the accuracy of the estimates provided by the two sampling methods in the Student Dialogue. During the student discussion you may also wish to discuss other sampling methods used in practice by researchers, such as stratified sampling, where a population is first broken into categories and then random samples from each category are chosen based on the size of the categories. This method ensures that individuals from all categories of a population are represented.

Choosing Samples

2. In the Student Dialogue, students are asked to "estimate the average area of the 100 rectangles below using samples of only five rectangles." How is this similar to or different from being asked to "estimate the average size (square-footage) of single-family homes in the states of Florida and Texas" or being asked to conduct a poll to find out which presidential candidate most Americans favor for an election?

   In many problems, we often do not have access to all the information in a given population, which is one reason we work with a sample. We choose a sample that we believe is representative of the population and gain information about the sample, which we then generalize to the larger population. For example, in the single-family-homes situation, it might theoretically be possible to find every single-family home, measure the area of the floor plan, and calculate the average; however, that is vastly impractical. In fact, during the time it would take to do this, some homes might be built or demolished, so it is, in a sense, not even *theoretically* possible to get the exact figures. Instead, a sample can be chosen for each state and then a generalization can be made at the state-level. Another example is that of a newspaper polling Americans about who they favor in a presidential election. Since a newspaper cannot go around to every person in the country and ask them who they would like to win the election, the newspaper instead chooses a sample and polls those individuals. Based on the sample results, they then make an estimate about the overall country's feeling towards the candidates.

3. How do you think the estimated average area of the 100 rectangles would be different if students used samples of 10 instead of samples of 5?

   You might turn this question into an exploration for students to try out. Students should eventually recognize that the larger the sample chosen the more accurate the prediction of that sample because there are more data points (assuming the same number of 10-rectangle samples are chosen as 5-rectangle samples). You can give the analogy of trying to find the average height of a basketball team. The more people you measure, the closer you are to measuring all the players and going from an estimate of the average height to the actual average height.

4. In line 9, Anita imagines the 20 students in class dividing up the 100 rectangles, each taking a different set of five, and computing their averages. She then poses the question, "Would the average of those 20 averages give the same result as getting the average of one set of 100"? Just using intuition, and without working this out, what do you *think* is the answer to Anita's question? You'll get a chance to investigate more deeply later.

   This question is meant to get students to debate their initial ideas. In the Related Mathematics Tasks, there are problems that consider taking the average of averages for both equal and unequal categories. This discussion question can then be revisited.

Choosing Samples

Possible Responses to Related Mathematics Task

1. What is the average area of all 100 rectangles? Which sampling method provided the better estimate: using random numbers or picking rectangles by hand? Why do you think that method works better than the other?

   The average area of all 100 rectangles is 7.42. The random sampling method produced the more accurate estimate of 6.1, which is only 1.32 away from the true mean. The picking-rectangles method produced an estimate of 9.92, which is 2.5 away from the true mean. Using random numbers worked best because it helped get rid of bias when picking rectangles to include in the sample. In at least this example, picking rectangles visually seemed to favor the larger rectangles, and this is what causes that estimate to be larger and less accurate than the estimate provided by the random sampling.

2. Imagine you are asked to survey the students in your class for their favorite sport. You are asked to use a sample of five students only. Is it best to choose five classmates by (A) picking five names out of a bag, (B) picking five of your friends, or (C) picking the first five classmates that walk in the room? Explain your reasoning.

   Method A is the best since it provides the least bias. Method B might have bias since your friends may enjoy the same sport you do and, therefore, might not be representative of the whole class. Method C might also have bias if certain students are consistently early (or late) to class. Let's say athletes always show up to class early since they know they need to leave class a little early for practice; method C would be favoring the sports that group enjoys.

3. In line 24, Anita generates an interval around the mean, using the mean absolute deviation to define its limits. In line 26, Dana asks if "Anita's limits tell us where *most* of the sample averages will fall." Do they? Explain why or why not.

   Anita's limits do not tell us where most of the sample averages will fall. Looking at the table from line 16 we do see a majority (14 out of 20) of the sample averages from the "hand-picking rectangles" column are within ± 1 MAD of the sampling distribution mean of that column. However looking at the "random numbers" column, only half (10 out of the 20) sample averages are within ± 1 MAD of the sampling distribution mean of that column.

   We can also create scenarios where *most* of the sample averages are outside ± 1 MAD of the sampling distribution mean. For a simple case, the sampling distribution mean in the set 10, 10, 10, 19, 21, 30, 30, 30 would be 20 and the MAD would be 7.75. Dana's hypothesis is that most of the sample averages should fall in the interval 12.25 and 27.75 which is clearly not the case.

4. Split the 100 rectangles into disjoint groups of five (rectangles 1–5, rectangles 6–10, etc.) similar to Sam's suggestion in line 8. Find the average area of each of the smaller groups. Take the average of those averages. How does the average of the averages compare with the average area of the 100 rectangles found in Question 1?

   When you split the rectangles in disjoint groups of five as described in the problem, you get the following average area for each group:

Groups	Average Area
Rectangles 1–5	1
Rectangles 6–10	4
Rectangles 11–15	7.6
Rectangles 16–20	3.2
Rectangles 21–25	7.6
Rectangles 26–30	7
Rectangles 31–35	9.8
Rectangles 36–40	8.6
Rectangles 41–45	7.4
Rectangles 46–50	7.2
Rectangles 51–55	10.6
Rectangles 56–60	10
Rectangles 61–65	5.2
Rectangles 66–70	9.6
Rectangles 71–75	10
Rectangles 76–80	7
Rectangles 81–85	7.2
Rectangles 86–90	9
Rectangles 91–95	6.6
Rectangles 96–100	9.8

   Taking the average of the averages, we get 7.42, which is the same as the average area for all 100 rectangles as found in Question 1.

5. Split the 100 rectangles in 10 disjoint groups of five (rectangles 1–5, rectangles 6–10, etc.) and five disjoint groups of 10 (rectangles 51–60, rectangles 61–70, etc.). Find the average area of each of the smaller groups. Take the average of those averages. How does the average of the averages compare with the average area of the 100 rectangles found in Question 1?

   When you split the rectangles in disjoint groups of 5 and 10 as described in the problem, you get the following average area for each group:

Groups	Average Area
Rectangles 1–5	1
Rectangles 6–10	4
Rectangles 11–15	7.6

Rectangles 16–20	3.2
Rectangles 21–25	7.6
Rectangles 26–30	7
Rectangles 31–35	9.8
Rectangles 36–40	8.6
Rectangles 41–45	7.4
Rectangles 46–50	7.2
Rectangles 51–60	10.3
Rectangles 61–70	7.4
Rectangles 71–80	8.5
Rectangles 81–90	8.1
Rectangles 91–100	8.2

Taking the average of the averages, we get 7.06, which is different from the average area for all 100 rectangles as found in Question 1.

6. In Questions 4 and 5, you experimented with taking the average of averages for both groups of equal and unequal sizes. What did you notice about the average of averages and why do you think this happens? To help in your response, consider the following:
   A) Imagine you have 100 rectangles that are separated into 20 equal groups. The average areas of the 20 groups are $a_1, a_2, a_3, ..., a_{20}$. Show algebraically that the average area of the 20 groups is the same as the average area of the 100 rectangles.
   B) Imagine you have 100 rectangles that are separated into two groups: one group with 1 rectangle whose area is 10 and one group with 99 rectangles each with an area of 1. How does the average area of the 100 rectangles compare with the average of the average area of the two groups?

   In Questions 3–5, students see that the average of group averages is equal to the average of the whole population only when the groups are of equal size. In this question, they are asked to consider two scenarios to explore why. In the first scenario with equal groups, students can express the sum of the areas in each group as $5a_1, 5a_2, ..., 5a_{20}$. Adding these and dividing by 100 gives us the average area of all 100 rectangles, which can be shown to equal the average of the 20 groups' averages as seen below:

$$\frac{5a_1 + 5a_2 + ... + 5a_{20}}{100} = \frac{5(a_1 + a_2 + ... + a_{20})}{100} = \frac{a_1 + a_2 + ... + a_{20}}{20}$$

   A similar algebraic process can be used for other scenarios with equal group sizes.

   In the second scenario, students are asked to consider what happens to the average of group averages when group sizes are unequal. In the group with only 1 rectangle, the average area is 10, while in the group with 99 rectangles, the average area is 1. Taking the average of 10 and 1 will give us 5.5, which is clearly larger than the average we would expect for the 100 rectangles, almost all of which have an area of 1. Looking at an exaggerated case where group sizes are unequal helps us understand why the average of group averages is not (usually) equal to the population average in those cases.

Choosing Samples

7. Two employees of the Census Bureau are trying to calculate the average age for the population across the 50 states. Employee A says they need to take the average of everyone's age across the 50 states. Employee B says they can use a shortcut by averaging the average age for each state. What do you think of the two employee's proposals?

    Employee A's method is guaranteed to work since that employee is taking a straightforward average of everyone's age, but it is a huge job and likely to be quite costly. Employee B's method, while appearing to be a shortcut, is incorrect. Averaging the averages would not give the same as the average age for all the people across the 50 states, because different states have different populations. For example, if we take the average of the average age for California (a state with a high population) and Montana (a state with a low population), we would get a value closer to that of Montana's than if we were to take the average age of everyone in California and Montana combined. Using the average age at the state level increases the weight of a state with a low population and decreases the weight of a high population state when taking the average of averages.

8. Budget cuts at the US Census Bureau prevent the agency from collecting *every* person's age (along with other information). Propose a way to estimate the average age of people living in the United States.

    Instead of collecting information from every person, one could choose a random sample from across the 50 states and then take the average age in the sample. This sample would have to be randomly selected from across *all* the states and couldn't consist of an equal random sample from each state since that would favor states with lower populations (see Related Mathematics Task 7). Of course, the estimate can be improved by having a bigger sample size as long as the sample is unbiased. The more people included in the sample, the closer the sample is to including the entire population and, therefore, the closer the sample estimate is to the actual population's average age.