CCSSM Grade 7 Sampling: Content for Teachers and Classroom Activities

for Fresno Unified, Spring 2014
Tim Erickson • Epistemological Engineering

Introduction

Here are the standards in question:

7.SP.A Use random sampling to draw inferences about a population.

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.

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.*

This introduction will help you (a teacher) understand what the standards-folks probably mean by this. We'll also discuss some pedagogical issues and important connections to other content, especially direct proportion.

Activities Overview

This packet also contains three substantial activities that center around sampling.

In **The Bean Counters**, students sample from a large "population" of beans, trying to figure out what percentage of the beans are which type (e.g., pinto beans, white beans, or pink beans). On a class graph, they see how different students' samples give different percentages. Through that, they start to gain insight into how far off their sample values are likely to be. Here the "quantity of interest" is a percentage.

Words on a Page has students sampling words—just as in the example given in the Standards. Here the "quantity of interest" is the average (mean) number of letters in the words. The point of this activity is that it matters how you choose a sample. It's better to have a random sample, but even then, you have to take care that your random sample is not *biased*.

In **Estimating the Crowd**, students invent a procedure to estimate the size of a population, in this case, the number of people in a photograph of a crowd.

Basic Ideas Behind Sampling

Suppose you want to know the average length of a word (we do this in the central activity, *Words on a Page*). You pick ten words off a page at random and find the average length.

Do you know the average length of all words? No! Because you didn't count the letters in all of the words on the page. That would be impractical. So instead you counted ten.

The ten you used were a *sample*. All the words on the page constitute the *population*.

To make valid inferences, the sample has to be *representative* of the population. It turns out that taking a *random* sample is a reliable way of getting a representative sample.

Note that the population is not "all words in English." It's better to say that our population is "the words on this page"— not even "the words in this book." It can still be hard to get a random sample—but we try, which is pretty much as far as they get in real life.

> Note important vocabulary: *sample* and *population*. Sample is a noun as well as a verb. And population is a collection of things, not just a number.

In a different example (*Bean Counters*, the first activity in this unit), we want to know what percent of the beans in a bowl are pinto beans. There are thousands of beans, so again, counting is impractical. Instead, we take a sample. If we sample 20 beans, and 7 are pintos, our best guess is that 35% of all the beans are pinto beans.

Two Types of Settings

Elementary statistics and sampling deal with *categorical* and *numerical* variables.

In the "word length" scenario, we computed an average. Our variable—word length—is *numerical*.

In the "beans" scenario, we computed a percentage (or fraction, or decimal). Our variable—the type of bean—is *categorical*. It divides the population of beans into *categories*: pinto, black-eyed pea, kidney, whatever.

It may be asking too much for seventh-graders to classify problems by whether they're numerical or categorical. Mostly, students do the right thing intuitively—or we ask the "right" question explicitly. The standards do not specify which we should focus on at this stage. But you, the teacher, will benefit from recognizing this distinction. And it might be good to bring up the issue in discussion: why do we find percentages here and averages there?

Yet the principle of sampling—use a sample to represent the population because it's too hard to get the answer for the whole thing—is the same in both cases. But you use different mathematical tools: averaging for numerical settings and proportion for categorical.

One more thing about categorical problems: sometimes you're studying the number, and sometimes the proportion. Students need to convert between the two (good practice!) depending on the context.

The Broader Picture: Stats and Probability

The two standards at the top of this document are the first part of a terrifying and gigantic set: the entire Statistics and Probability standard at Grade 7. They go on to address inferences about the difference between two populations, and then leap into probability with both feet. If you were an AP Stats teacher at high school, or a professor of statistics at a college, and looked at those standards, you might say, "yep, that's pretty much what we do."

So what can they be thinking, putting this is seventh grade?

We'll adopt a kinder, charitable answer: here in grade 7, we're doing a much more informal version of the topics than you might think if you were teaching AP. And when we're not as informal, we'll be doing only the simple versions of these often very complex topics.

For example, look at standard 2 above. If you google "inferences about a population," you'll find out that there are really two kinds of inference procedures: hypothesis testing and interval estimates. That is, are you doing *t*-tests (or chisquare), or are you computing confidence intervals? Yikes. And that's if you're doing a Normal-based, frequentist course. You might instead use randomization techniques, or even go so far as to use a Bayesian paradigm. Got that? Neither do I.

What do we mean at grade 7? Think of the example above about the beans. We got 7 pintos in our sample of 20. What percent are pintos? 35%. It's just a proportion problem.

But there's a twist.

Variability

When we truly study stats at grade 7, we need to be aware of *variability*. So we add, "what's the *range* of the percentage you think is correct?" That is, would we be surprised if the correct value for the whole bowl were 40% pintos instead of 35%? (Probably not.) How about 90% pintos? (You bet!)

In high school and college, students will learn to *calculate* that range. But for now, we want students only to (a) calculate that "expected value," that is, 35%; (b) realize that they will not get the same answer every time, and learn to cope with that; and (c) start developing intuition for how much we expect things to vary. The short answer: things vary more than you think.

You should also be aware that there are at least three types of variability. Your students do not need to distinguish between them formally, but you should. Here they are:

1. **Inherent, natural variability**. You measure a number of things, and they're not identical. So if you look at a distribution of values, you see a spread. For example, if you plot all the students' heights, you'll see spread. There is variability in height.

2. **Measurement variability** (or measurement error). Even with the best intention, and even if you don't make any mistakes, when you measure the *same thing* more than once, you get different values. This is more science-y; we won't deal with that here.

3. **Sampling variability**. Suppose you estimate some quantity (e.g., the mean height of middle-school students or the percentage of people wearing sneakers) by taking a sample —say, a random sample of 100. You will not get the true exact value. What's more, if you sample again, you'll get a different value, because you got different people in your sample.

In this unit, students experience some natural variability, but also sampling variability. And we'll sample repeatedly in order to figure out how much sampling variability there is.

> Example: In *Words on a Page*, students study the lengths of words on a page. The words on that page are the population. To estimate the average word length in the population, they sample 10 words from the page. Those 10 words are not all the same length—that's natural variability. But each student will get a different mean, because they sampled different words. That's sampling variability.

The Big Picture about Sampling Variability: Making Skeptics

In this unit, students experience sampling variability by seeing how different students or groups get different results when they sample. This presents some logistical challenges in the classroom—students will have to contribute to a "class graph" somehow—but it's immediate and almost tactile. And it makes sense to students: *I don't expect Maria to get the same answer as I did? We have different samples*.

Where are we going with this? In a broad sense, we're developing *intuition* and *skepticism*.

Look at it this way: if we use sampling to figure out the mean length of the words on the page, and everybody gets a different answer, then it's possible that *everybody is wrong*. Of course, our wrong answers are better than just guessing. But they're still wrong. We have to ask, then, *how wrong are we*? How far off is our result likely to be?

We answer that by looking at the class distribution of sample means: the class graph that everybody contributes to. (Official statspeak: this is called a *sampling distribution*.) You can see from the spread of that distribution, how far off, in general, the samples are from the true mean. We used the computer to pull 100 samples of 10 words each. The dot plot shows the mean lengths for each sample.

The middle of the distribution is at about 4.25. Chances are pretty good you'll be at least 0.5 letters off, but it's unlikely you'd be more than about 1.0 letters off. Think about that for a moment.

Now think about this: *in real life, you get only one sample*. You don't get to see the "class graph." So you have no idea which direction you're off, or exactly how much. Suppose you got a mean of 3.5 letters per word. You might say with some confidence that the true mean was between, say, 2.5 and 4.5. Why that range? Because when we simulated taking many samples, it looked like plus-or-minus 1.0 encompassed nearly all of the samples.
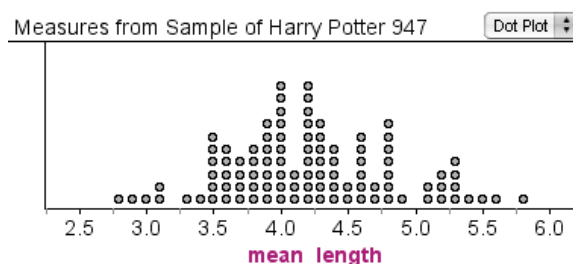
This is exactly the same idea as the "margin of error" statisticians quote when they do opinion polling. When you poll people, you'll be wrong. You don't know how much or in what direction, but you can put a range on your answer, and be pretty confident that the true answer falls in that range.

More things that should make sense:

- If your sample is biased somehow, your answer can be even more wrong, and you have no idea where the true answer lies.

- If the population is spread out more, your samples will vary more.

- If your sample gets larger—you sample more words—your mean gets closer to the true mean: the samples will vary less.

All that was for you, the teacher. What do we want from students at Grade 7?

Let's repeat what we said above: students should (a) calculate that "expected value," that is, 3.5 letters; (b) realize that they will not get the same answer every time, and learn to cope with that; and (c) start developing intuition for how much we expect things to vary, that is, be appropriately skeptical about conclusions they draw from samples.

There are lots of details, of course. But that's enough for now.

Where do students get confused?

Here's a deep, advanced one. Do not expect everyone to get this at this time: Students get confused between the distribution of their data (the words in their sample, which might range from 1 to 12) and the class graph, which is a distribution of means (which ranges from about 2.5 to 6).

And some simpler issues; it's good to probe for understanding:

- What does it mean that the average length of a word is 3.5 letters? Are there *any* words with 3.5 letters?

- Your average word length is 4.6 letters. There are 947 words on the page. How many letters is that altogether?

A Few More Issues

Sampling and Randomness

Sampling is a random process. How is it related to other random processes such as rolling dice and flipping coins?

The amazing deep answer is, they're the same. Think of it this way: rolling a fair die is the same as sampling from an *infinite population* that's evenly divided among the set {1, 2, 3, 4, 5, 6}.

Vocabulary: *Imply* and *Infer*

Students will learn about inference. How can we teach them what "infer" means? One explanation is to contrast it with "imply"; you can also use the idea of cause and effect.

Imply: If you know about a cause, you can reason about the effect. A cause *implies* an effect.

Infer: If you know about an effect, you can reason *backwards* to the cause. You can infer a cause from an effect.

For example, think about this statement: *if it is raining, then the streets are wet*. The *cause* is rain; the *effect* is wet streets. Rain *implies* wet streets. It's inevitable.

On the other hand, if you see that the streets are wet, you might *infer* that it was raining. Here you can't be sure, though. There might be other reasons streets can be wet, for example, a broken water main or a huge water-balloon fight.

This also works in a sampling situation. Suppose we have a box with 1000 pennies. Some are older (before 2000). Most are newer. The box of pennies is our population. We will pull out a sample of 20.

If we know beforehand that there are 250 old pennies in the box (in the population), we expect that our sample will contain 5 old pennies. That's just proportion and common sense. It's implied by the situation. Of course, it won't always be exactly five.

But usually we don't know how many old pennies there are. So we look at the sample of 20. We count the 5 old pennies and *infer* that there are 250 old ones in the population.

What about Polling? Why NOT to Do a Poll in This Unit

In real life, we see the results of sampling every time we learn about a poll in the news. We want our students, as citizens, to develop the same skepticism and intuition about professional polling that they do about their own results in classroom activities.

To that end, it is tempting to have students do polling projects, either in the form of opinion polls or some other "about us" survey. This section is intended to make you question that practice. It's not that an "about us" survey isn't a good idea—it's engaging, it helps students develop good questions, it lets them display results in different ways and create effective visualizations—but it's really hard to draw valid conclusions that you can generalize beyond the survey itself. And that's what sampling is for.

So let's look at a polling example. Suppose Fresno County Measure T would raise the sales tax by 1% to increase teacher salaries. Will measure T win?

We don't want to wait 'til after the election. We need to know how it's doing right now; if it's losing, we'll want to buy a TV ad supporting the measure. So we take a poll. We ask 100 people what they think, and 70 of them say they plan to vote yes. Yay! We're winning!

We've used sampling and a proportion to answer our question.

- Ideally, we would have asked every voter in Fresno County. But that's impractical. So we take a sample—a smaller number that will represent the larger number.

- We assume that the proportion in the "population" (i.e., all the voters in Fresno County) is the same as that in the sample. So we conclude that the vote will go 70% for, 30% against.

- In this case, we don't care how many people will vote for Measure T. We only care that the *proportion* of people is more than 50%, so it wins. In other situations, we'll care about the number.

So far so good. But there are all sorts of things that can go wrong. Let's list a few.

- If we had to ask 200 people in order to get 100 responses—after all, you can never get everyone you ask to respond to a poll—we don't know if the 100 non-responders might be generally against Measure T.

- If the way you asked the question in a non-neutral way ("Do you support Measure T, which is designed to help our kids and give hard-working teachers the respect they deserve?") the answers you get may not reflect how they'll really vote.

- If you asked people who might not be representative of the whole population of voters— suppose you collected your data in a parking lot at Fresno State—you might know only what that sub-population (current CSUF students and teachers) thinks. And that may not align with the population in general.

If these problems look like a mess to you, you're right. Polling that actually predicts what people will do is hard. You can do to deal with the problems, but this is are beyond the scope of what seventh-graders need to understand. The basic idea is that we "scale up" the sample to the population.

Intrusion of Real Life

The problems—non-response, the wording of a survey question, biased, non-representative samples—are the kind of real-world issues that don't show up in traditional math instruction. How should we treat them in a math class?

My advice: acknowledge the problems as much as you can. Praise students when they raise these issues. *But don't let them derail you if they're not the main point of the lesson*.

Personally, I therefore try to *avoid using a poll in a problem or project*. But even in a poll, the problems only get in the way of *inference*. That is, if students survey their friends about favorite music (yawn), they can make great visualizations and compute proportions, they just can't justify saying that all the kids in the school (or all seventh-graders) have the same preferences.

Here's a survey setting that somehow doesn't have the same pitfalls as an election poll. A great task is to have students plan how they will do a survey to answer the question; they'd need randomness to get a good sample. Or you can give them results, like this:

> Suppose you're planning lunch for an all-school field trip. There aren't a lot of choices. Each student gets one "main course." These are: PBJ, a ham and cheese sandwich, or two rice crackers with tahini between them. There are 500 students in the school. You can get 525 "mains," so you can over-buy and have some left over. What should you bring?

> You poll two students at random from every classroom in the school, and get these results: 15 PBJ, 15 Ham & Cheese, 2 Rice/tahini, 2 anything is fine, and 6 either sandwich is OK but please, please, no tahini.

By the way: The idea of getting two kids at random from every class is called *stratified sampling*. That's not in the standards at this grade, but students could come up with it anyway; it makes sense. You could discuss why it isn't good enough simply to use the *class* as the sample for the school (because other classes might be systematically different for any number of reasons).

Questions and Claims

You will probably come across resources that say something like, *students should come up with a research question and answer it using data*. What a fine idea! Ah, but how do students come up with a "research question?" Even if they have interesting ideas, turning ideas into questions often saps the life out of them.

How does this happen? Sometimes, students word their question so that it doesn't foreshadow an interesting conclusion. For example, suppose you think that girls in the school like soccer as opposed to basketball, while the boys are the other way around. A typical research question is, "What sports do students prefer?" A more interesting question would have been, "How do girls' preferences for basketball and soccer differ from boys'?" But that's hard for students to come up with.

An alternative is to have students *come up with a claim they can address using data*. This sometimes works better. A "claim" is a statement that's either true or false—we don't care which. The student has to find or collect data that address the claim. The student's report then describes how the data support or refute the claim.

In this case, the student can put the interesting idea into the claim: "Girls prefer soccer, and boys prefer basketball."

Bean Counters:
an Introduction to Sampling

This activity centers around a bowl with lots of beans. There are at least three types of beans in the bowl, all about the same size. The overall question is, "what percent of the beans are of each type?"

There are too many beans to simply count them all, so students will sample.

We want to get several things out of this:

- Experience with sampling, with inferring the percentages in the population, and with understanding how much those percentages are likely to vary.

- Some idea that if you use a bigger sample, you get a more accurate result.

- Experience with graphs and spread, and comparing two distributions. Specifically, we're hitting 7.SP.B.3:

    Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, [measuring the difference between the centers by expressing it as a multiple of a measure of variability].

    I've put brackets around the last bit because if students aren't comfortable with MAD, this will add a day or two.

The *class graph* is a key element of activities in this unit. We'll discuss this more in the teacher notes.

Overview

1. Introduce the beans (or whatever you're using).

2. Individual students guess the percentages of each kind of bean.

3. They post these on class graphs (dot plots). This step also lets you and your students figure out the logistics of the class graphs; make it efficient. You'll be doing it a lot.

4. Working in groups, students sample with a sample size of 20. They post their results for all three bean types on class graphs.

    Important: students must sample with their eyes closed!

5. You hold a class discussion about the results. See the notes for what to be sure to talk about.

6. Now they sample with a larger sample size (see note later on how large). Again, they post.

7. Finish with another discussion.

Bean Counters • Student Page

The teacher has prepared a population of beans.

Your task is to figure out the percentage of each kind of bean. There are too many beans simply to count them all. So you will *sample*.

That is, you'll take a relatively small number of beans and count them. If you mix the beans and choose randomly, your sample will have approximately the same percentages as the whole population.

Part 1: Predict!

1. Just looking at the beans, make your *individual* best guess about the percentages of each kind of bean.

2. Record your predictions on the class dot plot.

Part 2: Sample Size 20

Working in a group…

3. *With eyes closed*, one member chooses 20 beans from the population.

4. Find the percentage of each kind of bean.

5. Return the sample to the population. Don't want to run out of beans!

6. Plot your values on the class graphs (one for each kind of bean).

7. If you finish early, do another sample!

8. Participate in the discussion.

9. Change roles (somebody else will sample, plot, etc.)

Part 3: Bigger Sample Size

10. *With your eyes closed*, get a bigger sample. How big? Your teacher will explain.

11. Find the percentage of each kind of bean.

12. Return the sample.

13. Plot your values on the class graph.

14. Participate in the discussion

If you're done with a part and are waiting around:

Look at the graphs that people are making. Notice how the values are not all the same? Why is there a spread in the values? What does that spread tell you about the underlying percentage?

Bean Counters • Teacher Notes

This is pretty straightforward and deliberately repetitive, in the sense that students plot multiple values on class graphs several times. This gives them practice with the graphs, and reduces the individual workload: no individual has to take lots and lots of samples to make a distribution.

Preparation

Get a lot of dried beans. We chose these because they're cheap. Find three kinds of beans that are about the same size—ideally so you can't tell them apart by feel.

Now mix the class population of beans. Two of the types should be the same amount. The third should be different. For example, two bags of pinto, and one bag each of medium white and pink will make a 50-25-25 mix.

The point is that students should be able to tell from sampling that the percentage of pinto beans (in our example) is greater. And then, even though there will be a difference in the data between white and pink, you can explain that in reality, they're (nearly) the same percentage. Ultimately we want to develop students' *intuition* about how big a difference represents a "real" distinction.

Pre-thinking the logistics

Spend some time thinking about your students, how things can get bogged down, and how you can prevent it.

For example, the first sample is 20. If every group representative painstakingly counts out 20 before the next one can start, you can burn up a lot of time. So maybe give each group a paper cup, in which they get a (suitably mixed and random) scoop of beans. Then the group's sampler can pick out 20 at the same time that every other group is doing it.

Similarly, think about the class graph (more on this soon).

Population and Sample

The question is about the percentages of beans in the class bowl of beans, the one you prepared. So that's the population. Students will infer about the population from their samples.

You should use that vocabulary ("*this is the population of beans*") so that students will start doing that. Also, ask them about their *samples* (not simply their beans). *How many in your sample*? *What percent of this sample are pinto beans*?

The Class Display

Each student will add his or her group's data to the class display of sample percentages. The display will be a dot plot. (Or course, it doesn't have to be a dot. The key is one mark per data point.)

About the axes. Ultimately, we will want to compare all three types of beans using axes we can compare, that is, they should have the same *range* and *scale*, and will ideally be parallel. This may work best if the axes are *vertical*. Even though most dot plots are horizontal, they don't have to be. For these proportions, simply make axes that go from 0 to 100%.

Three typical techniques are to use whiteboard dots (or X's or whatever) or adhesive notes (i.e., Post-Its™); or to use regular markers on poster paper. A student posts the group proportion for each type of bean whenever they're ready. The advantage of poster paper is that you can preserve it for another day if the lesson goes over; also, the graphs can be separated horizontally for easy access by more than one student at a time, and then brought together.

The graph at right shows seven results displayed on a whiteboard. Each person plotted three *x*'s. The three graphs are for pinto, white, and pink beans, and are all vertical. Notice the overlap, but see also how, given all the data, it looks as if the pinto beans are more numerous.
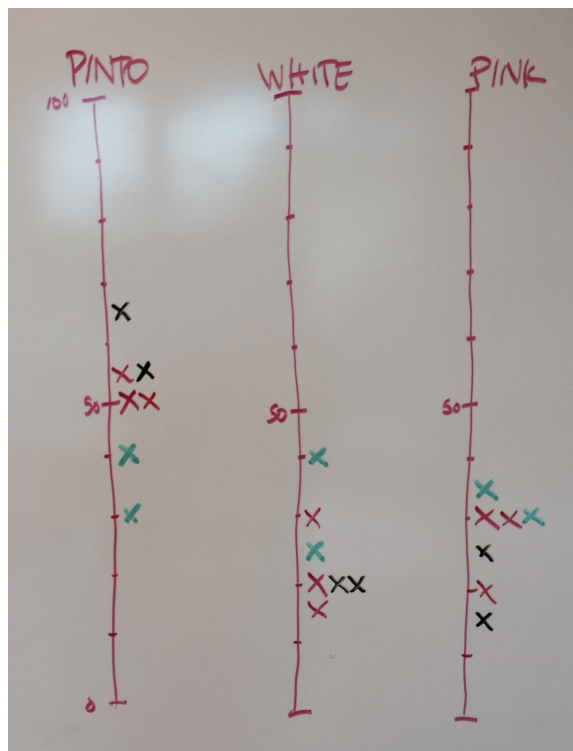
The Larger Sample

Give groups small cups to draw their samples. Now each sample will have more beans, but it probably won't be a friendly number. No matter. They have to compute the percentages anyway, and plot them in the same way, *on a fresh set of axes*.

Try out different-sized sample cups. I'd aim for between 50 and 100 beans—you don't want the whole class to be used up counting. With my beans, I tried a 2-Tbsp coffee measure (13-15-24, for a total of 52) and a ⅛ cup coffee measure (18-21-35, total of 74). Either would work fine.

Materials

- Beans (I used a total of 4 pounds, divided 2:1:1. See the materials notes at the end of this packet.)

- Scoops or equivalent for the larger sample (see above).

- Optionally, additional cups for big samples to improve logistics around the big bean "population" bowl.

Words on a Page

This activity shows how a *random sample* gives you a better result than a sample chosen some other way. It also shows how a technique you think is random might be *biased*.

But that's a surprise. So make that point at the end of the activity.

At the beginning, it's enough to say that we have done some activities about sampling, but we haven't paid a lot of attention to how we sampled. How we sample turns out to be important. We'll try three different techniques and compare them.

The Context: Text from *Harry Potter*
Students get a page of text (as at right). You can ask if they recognize it; it's the opening of the first *Harry Potter* book. The idea is that we want to figure out the average length of the words on this page. Because there are so many words (over 900), we'll do that by sampling: we'll pick 10 words and average those.

Overview
But how shall we sample? We'll use three methods:

Method A: Chosen Sample. The student chooses ten words they think are representative of the words on the page.

Method B: Blind Pencil Point. The student closes his or her eyes and points at the page with a pencil in order to choose a word. If they accidentally point at a number, it doesn't count.

Method C: Random Choice. The small numbers are "word numbers" on the page. Use three rolls of a ten-sided die to get a number between 000 and 999. If that-numbered word is on the page, use it. If not, roll again. You may have to explain the scheme to the students.

For each method, students contribute to a class graph (a dot plot) of all the means. That's one dot per student, representing the mean of his or her sample of 10.

Then you facilitate a discussion about any issues with that method. There are notes later in this packet to help you.

After Method C, the random sampling, you should discover that the center of the class distribution is closer to the true average (4.26 letters, but don't tell them too early!) than either of the other two.

In addition to sampling, students compute means; make dot plots and (optionally) box plots; and reason about procedures.

Words on a Page • Student Pages

How many letters are there in a word?

What an odd question!

It depends on the word. Not all words have the same length. But we can find out about the lengths of words by looking at chunks of text.

One thing we can find is the *average* length of a word. For example, take the first sentence in this paragraph.

3	One
5	thing
2	we
3	can
4	find
2	is
3	the
7	*average*
6	length
2	of
1	a
4	word.

words — Dot Plot
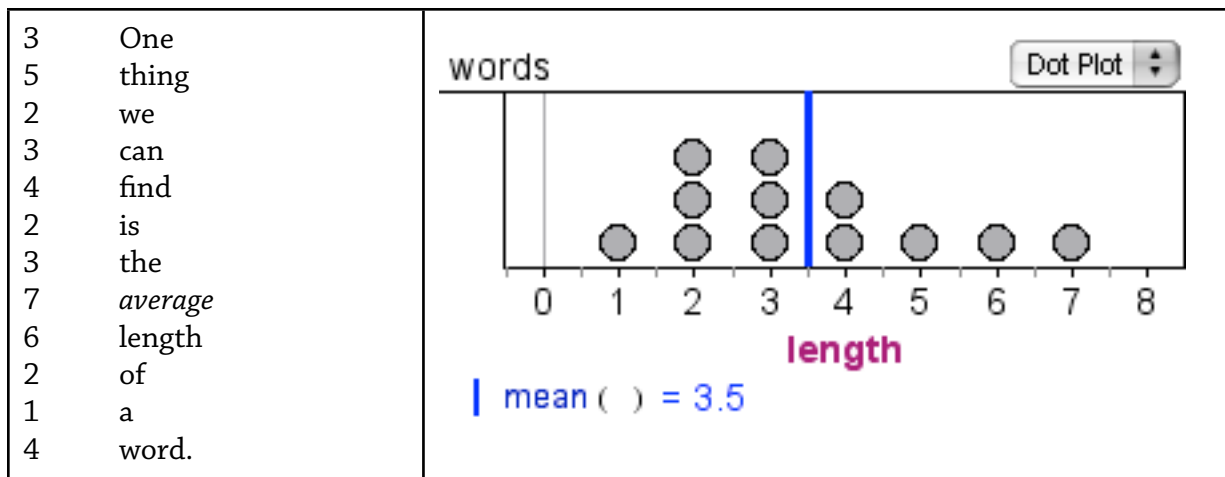
length

mean ( ) = 3.5

It has a total of 42 letters in 12 words. And 42 ÷ 12 = 3.5. That is, the average (mean) length of a word in that sentence is 3.5 letters. You can also see a *dot plot* of the lengths, and the position of the mean (3.5). Each dot corresponds to one word. The graph shows the *distribution* of word lengths.
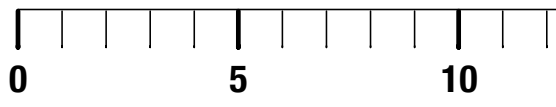
Now suppose you wanted to find the average length of the words on the handout. That would be a pain: you'd have to count the letters in all of the words, add them up, and divide by the number of words. That's too much trouble.
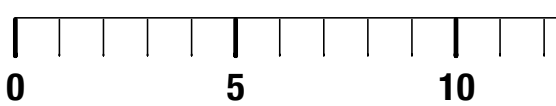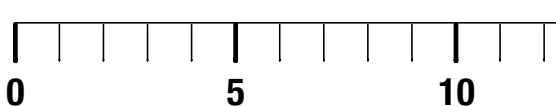
So instead you'll *sample*.

But how? You'll use three methods that you and your teacher will agree on. Each time, you'll sample ten words, write them down, count the letters, add them up, and divide by ten to get the mean.

Why ten? Because it's easy to divide…

You'll also make a dot plot by hand and mark the mean.

word	length	Method A
		Name(s):
		Mean length:
total		0 5 10 word length

word	length	Method B
		Name(s):
		Mean length:
total		0 5 10 word length

num	word	length	Method C
			Name(s):
			Mean length:
total			0 5 10 word length

Text Passage

Words on a Page • Teacher Notes

Opening

Make sure students understand what they're going to be studying: the lengths of words.

For this, you could pass out the first student page, or, at your discretion, simply talk through it or display the relevant parts using a document camera or computer projection.

The key is that students understand the overall question—how long are the words on the handout page? This requires understanding the idea of a *distribution*—that you can only answer the question if you know that there is a *spread* of lengths; students have to use tools and ideas (such as the mean) that deal with this variability.

Watch that students remember what a dot plot is and that they remember how to calculate a mean. These (along with *distribution*) are Common Core grade 6 ideas, but being realistic, some students will stumble. The next few parts of the lesson will reinforce those skills, so try to avoid re-teaching. Get help from as many students as possible.

Three Sampling Strategies

Assuming we buy the idea that it's too hard to count all of the lengths, we agree to sample 10 and use that to *estimate* the average length. (It's an estimate because we don't actually expect to get the exact mean of the population. That's one of the lessons of this exercise.)

Students will use three different sampling strategies, one at a time.

> The secret of the lesson is that the first two strategies have serious problems. That is, they're *biased*: they give answers that are systematically wrong. The third strategy is unbiased, so even though it doesn't give you the right answer every time, on the average, it does.

You will explain the strategies. You can let the students brainstorm how to do the sampling, but arrange to use these three; wait to "decide" on each until you're done with the previous one.

Method A: Chosen Sample. The student chooses ten words they think are representative of the words on the page.

Method B: Blind Pencil Point. The student closes his or her eyes and points at the page with a pencil in order to choose a word. If they accidentally point at a number, it doesn't count.

Method C: Random Choice. The small numbers are "word numbers" on the page. Use three rolls of a ten-sided die to get a number between 000 and 999. If that-numbered word is on the page, use it. If not, roll again. You may have to explain the scheme to the students.

For each strategy, students

1. Select and record the ten words in the sample;

2. Record the lengths (the previous page of this packet has forms you can use);

3. Calculate the mean; and finally

4. Post their mean on a class display. (Optional: post their data slips as well.)

The Class Display

Each student will add his or her mean to the class set of mean word lengths. The display will be a dot plot. (Or course, it doesn't have to be a dot. But it has to be one mark per data point.)

For suggestions about graphs, see the teacher notes for *Bean Counters*.

> **About the axes**. As in *Bean Counters*, the axes should have the same *range* and *scale*, and will ideally be parallel. And again, even though most dot plots are horizontal, making these vertical may help with comparison. For the Harry Potter passage, you should make sure each axis extends at least from 2 to 8. This should encompass all reasonable sample means.

Once the class has posted its data for Method A, everyone looks at the distribution.

Discussing Method A's Distribution

- What's the *population* we're sampling from?

  > It could be the whole book, but really it's just the words on the page.

- What decisions did you have to make about counting letters?

  > Do you count apostrophes? Do you count hyphens?

  > Point out that it's really up to you—but that it's important that everybody do it the same way. If you can, push the class into not counting apostrophes, and have them treat hyphenated words as different words. So "good-for-nothing" is 3 words, not a single 14-letter word. (Then their answers will match with the ones in this packet.)

- How is the distribution of means different from (or the same as) your individual distribution of 10 word lengths?

  > There are probably more points in the class distribution.

  > All of our individual numbers are whole numbers; the class has decimals.

  > More important: the class distribution is *not as spread out* as the individual distributions. Ask students why they think this might be the case. Solid ideas: "I have two and three-letter words in my own sample, but there's no way the average is going to be that low." "Taking the average mushes things together, so everyone will be closer if they take the average.

- Why don't we all get the same average?

  > We have different words in our samples, so the averages will be different.

- Based on this display, what do you think the average of the whole page is?

  > One approach is to average all of these averages. Do not let students calculate it (too much time). Instead, ask them to estimate it based on the graph.

  > Most classes get a result between 5 and 6.

- If we wanted to know the average length of the words, how do you feel about sampling as a way to get an answer, as opposed to actually counting all the letters on the page?

- Why would anyone care about average word length? (One answer: it's a measure of readability. It can help determine what grade-level can read it.)

- One more thing: How many of you picked "Dursley" as one of your words? Why?

- **Extension**: based on our estimate, how many *letters* are on the page?

Onward: Method B

Explain that maybe it would be better to get a random sample. Explain Method B (blind pointing) and have students implement it. New issues in its discussion:

- Is there any reason you might think the blindfold method will give you a better result than just picking words?

   "It's random." No bias.

- What decisions did you have to make when you used the blindfold pencil-point method?

   What happens if you don't hit a word? If you hit punctuation? Usually this results in a do-over, which is fine.

   Some students may have a system of pointing at a different part of the page for each stab; this is good as it ensures that they don't hit the same word by accident.

   Point out that, as with Method A, there are decisions you have to make about your procedure. (Extension: write your procedure clearly so someone else can do it the same way.)

- How do the class results compare with the ones from Method A? (This is why the two plots have to be the same scale: to facilitate comparing.)

   They will be different, but probably not by much; typically between 5 and 6 letters per word. Ideally, the distributions *overlap*. This is an important idea; be sure students see it.

- Now, based on the whole graph, what do you think the page average is?

Method C: the Payoff

Now you explain Method C, which may be the weirdest of the three. The blindfold plan was to choose words randomly. But here is another way to choose a random word…

Explain how the numbers on the page work; ask students to find word number 413. (Make sure everybody says "he.")

Give every student (or group) *one* ten-sided die (see the materials section for sources or alternatives). Explain that they are to roll the one die three times to find the three digits of a word number. Find the word and write it down. Do this ten times to get your sample.

Again, everyone plots their mean on the class graph.

- Now how does this distribution compare to the other two?

   If everything works right, the random averages will be substantially *smaller* than the other two —closer to 4 instead of between 5 and 6.

   You should still see overlap between the distributions, but it should be clear that even though some students' values were comparable to the other methods, in general, the values are lower.

- Why do you suppose it's smaller?

    It's not clear how well students can understand the underlying reasons, but here they are:

    When we pick words on our own, we tend to pick words that we think are characteristic of the passage, such as "Dursley." These are longer than the more typical small words—*a*, *the*, *he*, *and*, *was*, *of*—that make up so much of the text and bring the average down. Randomness avoids this unconscious preference.

    When we close our eyes and point, we're six times more likely to hit a six-letter word than a one-letter word—because it's bigger on the page! So although the procedure is random, it's *biased* towards longer words. And the average in our sample is bigger than the average in our population.

    Accept and discuss other ideas as well.

- Which one is right? Do you think this technique is better or worse than the others? (What do we mean by *better*?) Is this average closer to the true average length of the words on the page?

Harry Potter 947 — Dot Plot

mean ( ) = 4.25871
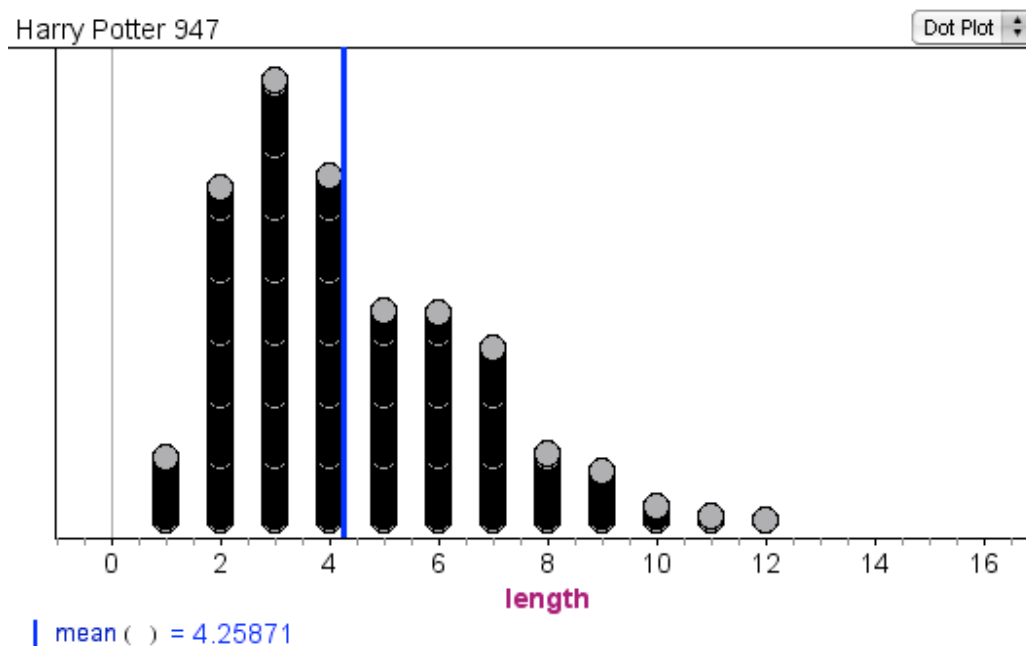
    Here is a dot plot of the lengths of the 947 words on the page. And the mean is about 4.26.

- What do you learn from seeing the three *distributions* that you don't get if we just knew the means?

    The overlap, and possibly the difference in spread.

Students may argue whether randomness is "fair." After all, Sylvie got a lot of short words and a mean of 2.8. Marco got a lot of long ones and a mean of 6! If we choose the words ourselves we can even it out. It's true that with only 10 in each sample, sometimes that will happen. Ten is really too small. But our *class* actually sampled 300 (or whatever)—and look at the class answer: it's better. In the long run, *randomness keeps us from making biased choices*.

Extensions and Alternatives

Numbers versus Graphs. When students come to the board to post their data *points*, have them list their data *values* as well. That is, each student puts a dot on the dot plot (at their mean) and also writes the mean in a list.

Then, in the discussion, ask about which is more useful, the graph or the list of numbers? The answer, as usual: it depends. If you want to calculate the mean of the means, say, you need the numbers. But mostly, the plot will give you the best understanding of the data.

Box Plots. Box plots (Grade 6, but students still may need practice) are a terrific way to summarize a distribution. Make sure students recorded the numbers, as suggested above in *Numbers versus Graphs*. Then, after you've put up all three methods, have the students make a display that features three parallel box plots, one for each method.

This requires three axes (number lines), all parallel, with the same *range* and *scale*.

Don't remember how to make a box plot?
http://math.serpmedia.org/diagnostic_teaching/mathematics-q/center-and-spread/

Want to see where box plots fit in the panoply of plots?
http://math.serpmedia.org/diagnostic_teaching/mathematics-q/dot-plots-box-plots-and-histograms/

Español. Do you think the average Spanish word is longer, shorter, or about the same as the average English word? Do an experiment to find out. But should you just translate individual words or look at some text? If text, should it be the same meaning?

Reading Level. You can actually calculate the reading level of the passage using the Flesch-Kincaid Grade Level test. Here is the scoop from Wikipedia:

This test rates text on a U.S. school grade level. For example, a score of 8.0 means that an eighth grader can understand the document. The formula is:

Grade level = (.39 × *ASL*) + (11.8 × *ASW*) − 15.59

where:

- *ASL* = average sentence length (the number of words divided by the number of sentences)

- *ASW* = average number of syllables per word (the number of syllables divided by the number of words)

Students can calculate *ASL* easily since they know that there are 947 words on the page. But they will need to sample to estimate *ASW*.

Estimating the Crowd

In this activity, students get a relatively open-ended task: to estimate the size of a crowd from a photograph. The picture at right is a small version. The plan is to impose a grid and sample.

One straightforward plan is to make a 10-by-10 grid and use 10-sided dice to choose sections at random. But there are many ways to address this task.

Because this is so open-ended, it's probably best to save this for last. It would help students to have done the "Harry Potter" activity (*Words on a Page*) first.

Alternatively, you could *start* the unit by introducing this task and letting students start talking about it, then return to it later when they have more skills and can see how useful sampling can be.

One Agenda

- Present the problem and form groups (if you do this in groups)

- At least 15 minutes of opening work time

- Get together to discuss challenges—especially deciding on sampling

- More work time (First class period may end during this phase)

- Groups make posters (or reports, written or "vodcast")

- Selected groups present to the class; teacher facilitates discussions about sampling and accuracy.

Materials

- Student handouts (especially the picture)

- Rulers

- Random-number generators, e.g., 10-sided dice, random digit tables, suitable software.

Estimating the Crowd • Student Pages

There was a big protest in Christchurch, New Zealand, in 2012. The next page has a picture of the crowd.

Your job is to estimate how many people are in the crowd. It's too hard to count them all, so you need to sample.

1. Divide the page (or part of the page) up into smaller sections—sections that are small enough to count. A grid makes sense.

2. Sample from those sections, counting all the people in each section of your sample.

3. Finally, perform some sort of calculation to determine the total number.

Make a report (or poster, whatever your teacher tells you to do) that explains what you did and why, what answer you got, and how accurate you think your answer is.

Your report should answer questions such as:

- What problems did you have deciding how many people were in a section? How did you solve these problems?

- How did you choose which sections were in your sample? Why does your procedure make sense?

- How did you choose what part of the picture to put in your grid?

- How did you make your calculation?

- What are the smallest and largest numbers you think the total number could be?
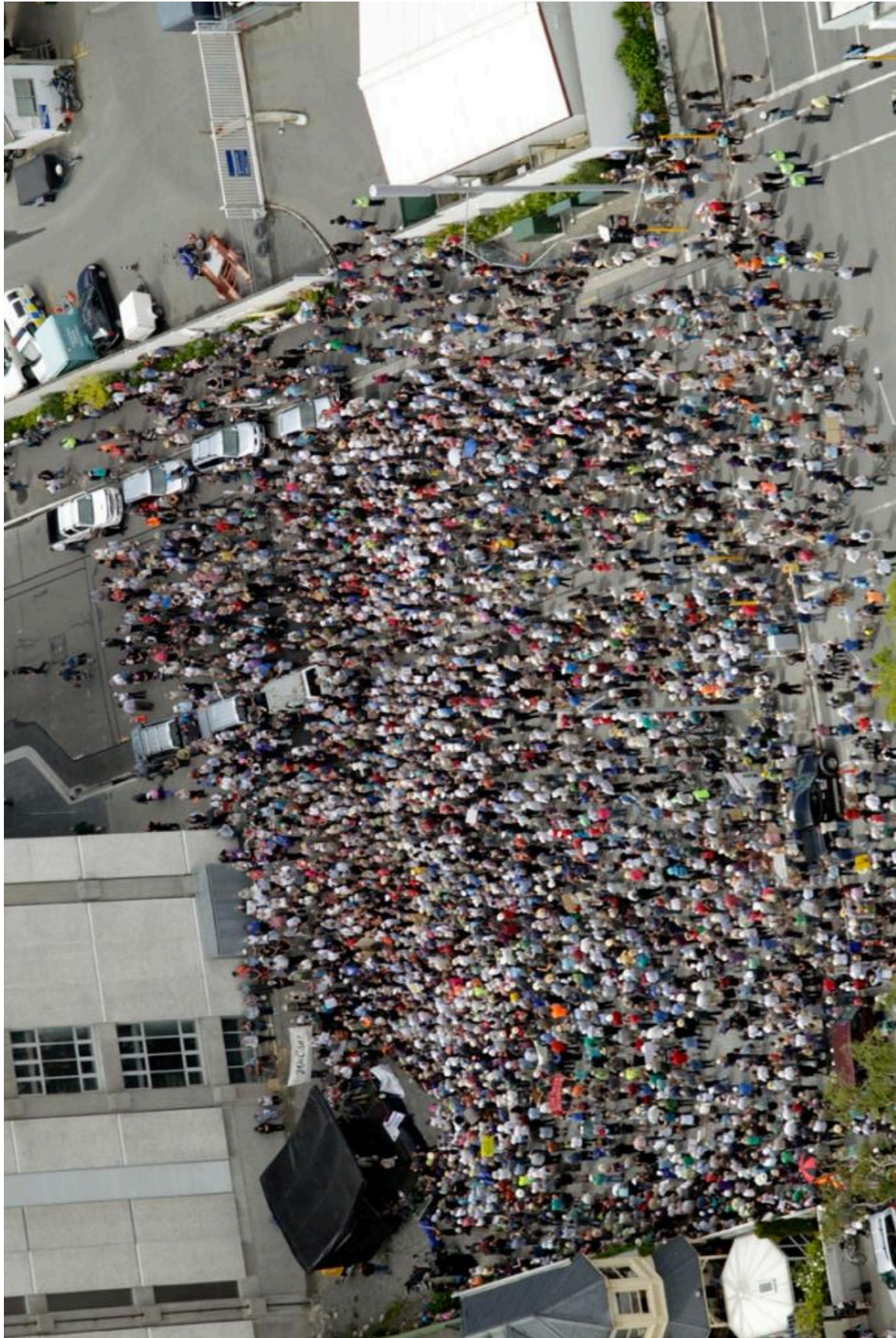
Estimating the Crowd • Teacher Notes

There are all sorts of challenges in this task. Here are some, with some comments:

- If you impose a grid, some people will be on the line

  Right. So you have to develop a scheme for dealing with it. Maybe you count the person if their head is in the grid cell. Or if most of the person is in. Or maybe you count half-people. Of course, the count is inherently inaccurate, so it may not even matter as long as you don't do something obviously biased, such as requiring the whole person to be "in."

- It's blurry! How do you tell where a person is?

  From farther away, it looks as if you'll be able to pick out individuals easily. But you can't. It helps to back off a bit to get more perspective, but even then it's tough. Therefore, the counts will be inaccurate.

- How do we deal with counts that we're not sure of?

  Try to be consistent. But also, consider having more than one person count each cell. Maybe the average of your counts will be more accurate. If you do, be sure you agree on what counts as being "in."

- What about cells that have no people in them? Do they count?

  Suppose you have 100 cells, and sample ten. You would expect to multiply the sum of the ten cells by 10 to get the total for 100.

  You have two choices to deal with zeros. The first is to have them count as zero. Just add up the people in the sample cells (one or more may be 0) and multiply by 10.

  The next choice is to eliminate the zero cells at the beginning. Suppose you have 9 cells that are empty. We'll call them "dead." If you choose one of those cells, it's a do-over until you get a "live" cell. You sample 10 and get 330 people. But now it's 10 out of 91, not 10 out of 100. So instead of multiplying by 10 to get the whole crowd, you multiply by 9.1. This comes from a direct proportion: the sample size (10) is to the population size (91) as our count (330) is to the actual total (x). You can also reason that you multiply by a smaller number because you have a bigger sum (since you avoided the dead cells).

- What about places where they are probably people but you can't see them?

  Then you have to do your best to estimate.

- I can't see my lines! Where are the edges?

  Make better lines. But if that doesn't work, make a "sampling frame." Cut a hole in a piece of paper the size of your cells. (They're all the same size, right??) Or make a rectangular frame the right size using paper and tape. Then place it over the picture in the right place.

With something like this, there are no right answers. But there are better and worse ones! Accept as many student ideas as you can. Strange ideas are good topics for discussion with the class. The point is to make a fair estimate; anything that uses sampling sensibly, and "scales up" the count you get to the whole population, is fair game.

Materials and Other Resources

Beans
Dried beans are cheap and durable. In the workshop, I brought four pounds: two pounds of pintos and one each of white and pink. I picked them to be about the same size so it was harder to distinguish the beans by "feel." Why? Because then there could be bias in sampling. I really wanted to use black beans, but with the brand I was picking, they were too small.

Ten-sided dice
Learning Resources via Amazon. $33.10 for 72 (double) dice.
http://www.amazon.com/Learning-Resources-Sided-Dice-Set/dp/B0018M0ILK

Direct from Learning Resources (http://www.learningresources.com) $36.99

Alternatives to ten-sided dice
You can make spinners practically for free. Make a circle on a piece of paper and divide it into ten equal parts (36°, right?). Label them 0 to 9. Now partially-unbend a paper clip to make a pointer. Stab a pencil-point through the loop and point it at the center of the circle. Tweak to spin.

You will probably need to do this, given students' abilities with protractors. But if you put four on a page, you can easily photocopy enough for the class.

Coping with Randomness in the Classroom
Whenever you use randomness, there's a chance that it won't turn out "right." We try to design things so it's unlikely, but it can happen. What should you do?

One answer: cheat. But it may be a better lesson to accept the real fluctuations that accompany randomness. Some advice from SERP:
http://math.serpmedia.org/diagnostic_teaching/mathematics-q/variability/