Modeling Measurements

Unit 6

Mathematical Concepts

- Statistical models explain the contributions of chance to the variability observed in a sample. In the measurements of Unit 1 or the production processes of Unit 4, random sources of error produce variability so that not all measurements or products are the same.
- To explain the variability observed in measurement and production processes, a model represents the measurement or production process as a combination of signal, the true measurement or target value of the production process, and of noise, the variability produced by random errors of measurement or of production.
- The signal component of the measurement model is represented by a constant value, estimated by a statistic of central tendency, such as the mean or median.
- The noise component of the measurement model is represented by a collection of random devices, such as spinners or mixers, each of which describes one source of variability. For example, when moving the ruler to measure an arm-span, small gaps often occur just by chance. A spinner could describe the likelihood and the magnitude of the under-estimates produced by these gaps.
- The signal-noise measurement model simulates each observed value as the sum of the value of the statistic of central tendency and the outcomes of the random sources of error that produce variability about this central tendency. Repetition of the model produces one simulated sample of observations. For example, 40 repetitions produce 40 simulated measurements.
- The sampling distributions of measures of center and of variability of a large number of runs of the model provide a means for considering model fit. Statistics describing the real data, such as the real sample's median and the real sample's IQR, should be in the center clump of the model's sampling distributions for these statistics.
- Good models fit the data and are good explanations of the variability of the measurements.

Contents

Mathematical Concepts 1
Unit Overview 2
Materials & Preparation 4
Mathematical Background 5
Instruction
Building a Model 11
Model Revision 28
Model Extensions 31
Formative Assessment 34
Student Worksheets 36
Unit Quiz 41
Unit Quiz Scoring Guide 45

Unit Overview

Unit 6 introduces students to building and testing models of the process that produced the measurements of the arm-span (or other attribute) generated by the class. The measurement process has two components. The first is the true measure of the arm-span, estimated by a statistic such as the mean or median. Real measurements tend to clump because there is a true measure. The second component is random errors of measurement. The unit assumes that Tinkerplots is available for use.

Days 1 and 2: Building a Model
Accounting for signal. Students begin by building a spinner model that estimates the true measure of the arm span data previously collected, or some other measurement or production data (at the teacher's discretion). The model's outputs are compared to the sample data. This comparison suggests that the model captures part of the real sample's distribution (its center), but the model fails to represent variability.

Incorporating chance into the model. To improve the fit of the model, students identify sources of chance variability that influence measurement. For each source, students design a chance device to model the magnitude and likelihood of the errors resulting from that source. Signed arithmetic represents errors of overestimate (+) and underestimate (-).

Combining signal and error. The model represents each measurement as a sum of signal and error components. The measurements now vary because errors are random and so vary from observation to observation.

Days 3 and 4: Model Revision
Students run the model to simulate a sample of the same size as the one they generated, generally consisting of between 30 and 40 observations. Students first look at one run of the model, comparing the model's output to their real data. Here statistics are useful for comparison. Is the median of the simulated sample the same as, or close to, the median of the real sample? Is the IQR or average deviation of the simulated sample like that of the real sample? Is it OK if the simulated sample does not include the extreme values of the real sample? After these considerations, which often prompt revisions to the model, the model is run repeatedly and measures of its performance, such as the simulated median of each sample and the simulated IQR of each sample, are collected. These repetitions of the model-generated samples are like those of Unit 5, where we generated a sampling distribution of the behavior of one random device. Here we are generating a sampling distribution of the model's predictions about center

and precision. The real sample's statistics ought to fit somewhere in the center of the model's sampling distributions of these statistics. If not, this indicates that the model may not be describing the process very well when we consider sample to sample variability.

Day 5: Model Extensions
Students create poor fitting models that result in about the same center but very different shapes than the real data. Or, at the teacher's discretion, students build a model of a production process that they have generated.

Days 6 and 7: Formative Assessment
Students create models of a production process and judge the adequacy of the model. A formative assessment conversation compares students' models.

Materials & Preparation

Read

- ☐ **Unit 6**
  Start by reading the unit to learn the content and become familiar with the activities.
- ☐ **Mathematical Background**
  Reread the mathematical background carefully to help you think about the important mathematical ideas within the unit.
- ☐ **Sample Student Thinking and Classroom Talk**
  Reread the Student Thinking and Classroom Talk boxes to anticipate the kinds of ideas and discussions you will likely see during instruction.
- ☐ **Modeling Variability (MoV) Construct Map**
  Read the construct map and visit the website to help you recognize the mathematical elements in student thinking, and to order these elements in terms of their level of sophistication.

Gather

For the class

- ☐ Student worksheets
- ☐ TinkerPlots or hand-held spinners

What is a Model?

Statistical models are explanations of how the variability of a process can be attributed to chance. Not all models of variability are statistical. For example, temperature variation could be explained by a source of heat that has no chance component. But in the real world, some of the variability in nearly every process has some variation that is due to chance. Although there may be philosophical disagreements about whether or not chance is an apt description, whenever we cannot control nor predict a particular outcome, but can predict the long-run structure of outcomes, then we can assume that chance is at play.

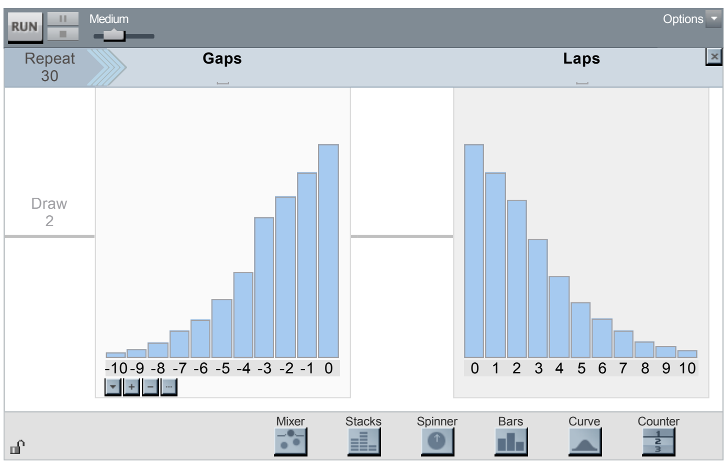
Modeling Signal-Noise Processes

The repeated measures (e.g., the measures of arm-span) and the production control contexts (e.g., the Toothpick Factory or the Rate Walk of Unit 4) both feature a repeated process that has two components: signal and noise. The first, signal component is the true measure for the measurement process or the target value for the control process. Measures tend to cluster around a central value, because the true measure is not shifting. Products tend to cluster around a central value, because the target value of the production process is not shifting. This component is estimated by statistics of center, such as the mean or median. If the process were ideal, there would be no variability. But for error, everyone would obtain the same measurement and all the productions would be exactly alike.

The second, noise component of each process is the variability that arises due to chance fluctuations in the measurement process or in the production process. These chance fluctuations are caused by, for example, random opportunities to inadvertently produce gaps or laps when measuring with the ruler in Unit 1, or to over- or under-estimate of the felt weight when packing the toothpicks in Unit 4.The sum of these chance fluctuations is estimated by statistics of variability, such as IQR and average deviation. In summary:
Observed Measure = True Measure + Random Error of Measure.
Observed Product = Target Value + Random Error of Production.

Analyzing and Modeling Sources of Variability

Analysis of variance refers to modeling the sources of random error so that the magnitude and likelihood of error approximates the errors observed in the world. For each source of error, a chance device represents the magnitude and likelihood of that error. For example, when students measure the length of an attribute with a 15-cm. ruler, they often notice gaps that result when the ruler is moved without measuring a portion of the length. Moving the ruler can also lead to inadvertent overlapping of the ruler. Students call this source of error, Gap-Lap. Gaps produce errors that are underestimates of the true length (the length is not measured), while laps produce errors that are overestimates of true length (the same length is measured twice). Small magnitudes of gaps and laps are much more likely than large ones, but all of these errors seem to happen just by chance. An example of a possible model of gap-lap error is depicted below using the TinkerPlots value bar random device. Notice that small errors are much more likely than large ones, and that the modelers believe that overestimates (+) and underestimates (-) are equally likely.

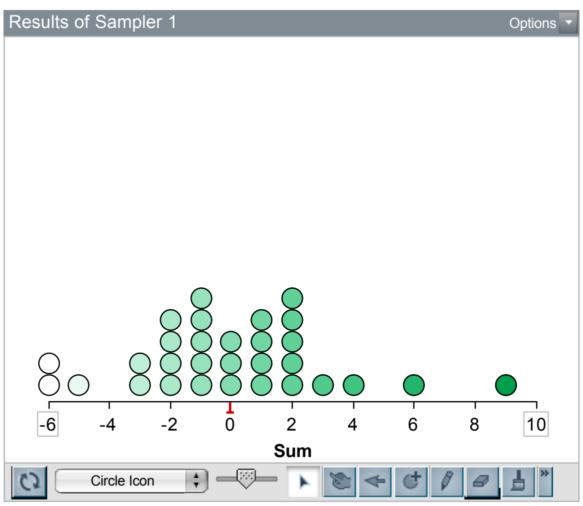


The results of 30 trials of the Gap-Lap model, obtained by checking the sum option under Results in TinkerPlots, are displayed in the table shown on the following page. Note that separate estimates of random error are produced for gaps (underestimates) and laps (overestimates), with the Gap-Lap error represented by the sum. For example, the 25th simulated measurement resulted in a Gap error of -3 cm. and a Lap error of 1 cm. for a total Gap-Lap error of -2 cm. This represents a measurer who left small spaces several times as he or she moved the ruler and who perhaps overlapped the ruler once and so underestimated the true measure by 2 cm.

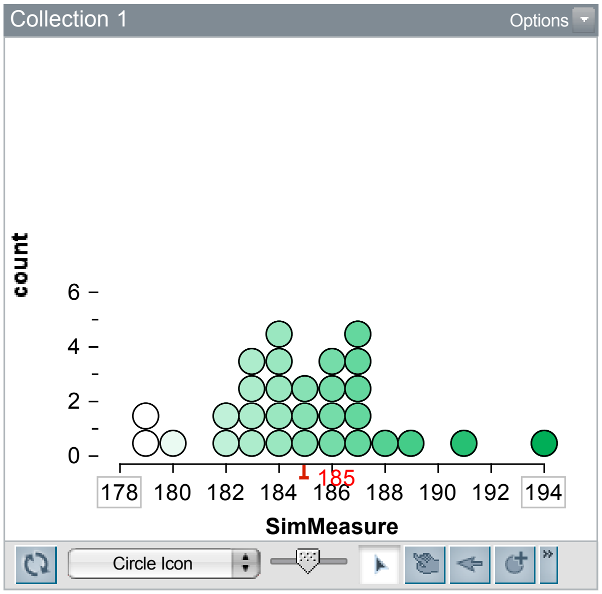
Results of Sampler 1 — Options

	Join	Gaps	Laps	Sum	<new>
19	-1,2	-1	2	1	
20	-2,0	-2	0	-2	
21	-1,1	-1	1	0	
22	-1,1	-1	1	0	
23	0,4	0	4	4	
24	-5,4	-5	4	-1	
25	-3,1	-3	1	-2	
26	0,2	0	2	2	
27	-7,1	-7	1	-6	
28	-3,0	-3	0	-3	
29	-2,0	-2	0	-2	
30	0,9	0	9	9	

The shape of the resulting errors is shown below, roughly symmetric about 0, as we would expect from the model.



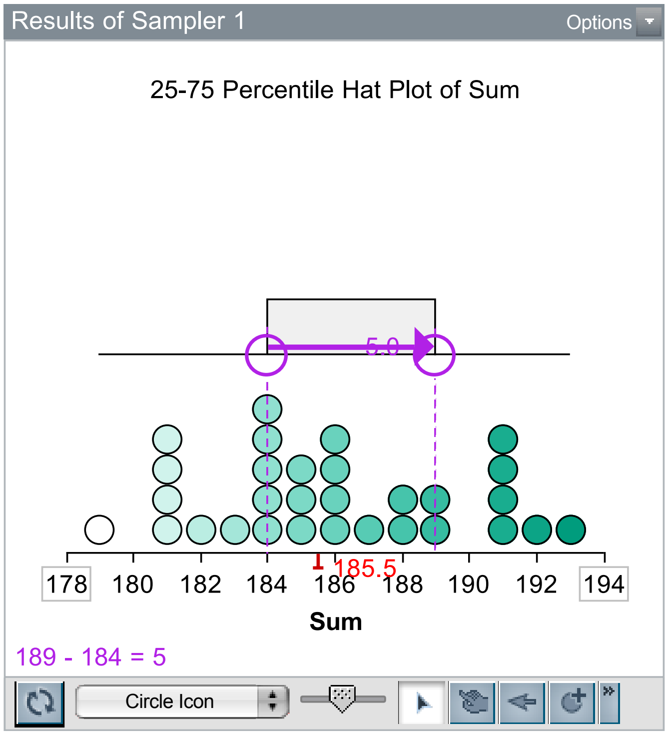
If we believed there were no other significant sources of error, we could form a model of the collection of measurements as the sum of the best guess of the true measure and the Gap-Lap error, as displayed on the following page, for a collection of measurements in which the median measurement was 185 cm.



Model Fit and Model Revision

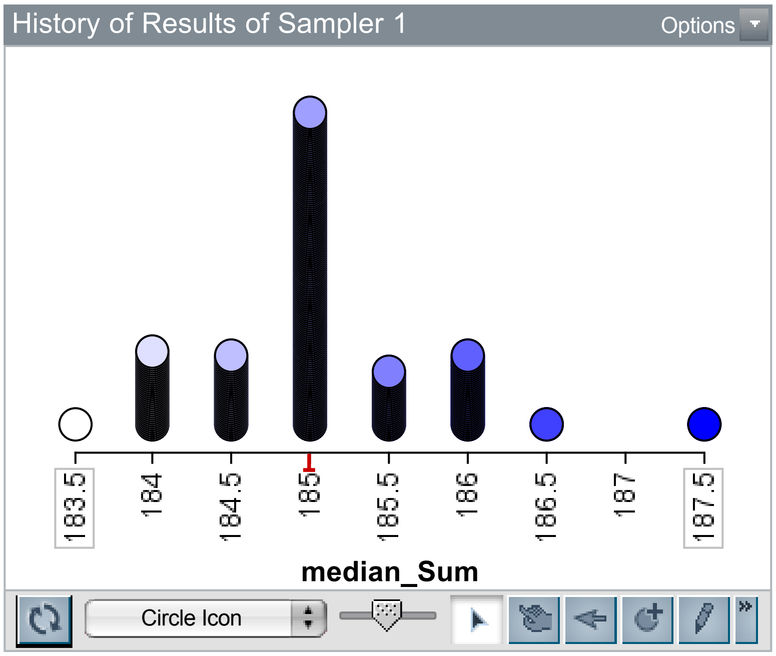
Models are intended to approximate real data. No model is ever a complete description of the process generating the real data, but we do expect that some models are better approximations than others. For example, one model might do a better job than another of approximating the shape of the real data. Two models both could describe the shape of the data equally well, but one might explain variability by showing its sources while the other does not.

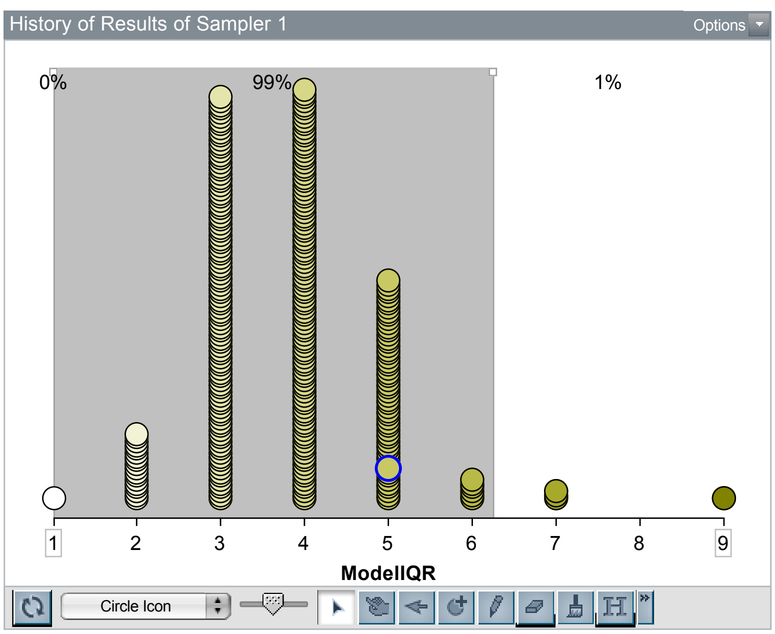
To consider model fit, we will consider a real sample of measurements with a median of 185 and an IQR of 8. So far, the model developed previously represents an observed measurement as a combination of the sample median (our best guess of the true length of the person's arm-span) and random gap-lap errors of measurement. The results of running the model again is displayed next, with the variable, Sum, describing the simulated measurements as the sum of the sample median and the random gap-lap error. Two statistics are shown, the model's estimate of the true measure (the median of the simulated sample) and its estimate of the precision of the measurements in the simulated sample, indicated by the mid-50 hat plot and the Ruler tool in TinkerPlots. Note that the median for this simulated sample of measurements is very close to the real median. The simulated IQR of nearly 6 is harder to interpret. Is it just due to this particular model run, or would other model runs have IQR's closer to the real sample's value of 8?



By running the model 300 more times, simulating a sample of 30 measurements each time, and collecting the model's median and IQR (or if we prefer, average deviation) for each sample, we can develop an approximation to the sampling distribution of the model's statistics, as displayed on the following page. These sampling distributions give a good sense of how well the model approximates the real sample's median and variability (the IQR of the real sample).

Recall that the real data sample had a median of 185 and an IQR of 8. Running the model 300 times, we see below that the model median (Median_Sum) did not vary much from sample to sample, and it was centered at the median of the sample of real data. Hence, the model does a good job accounting for the true value of the measurements. But, as the figure on the next page illustrates, the model did a poor job accounting for the variability in the real data. It suggests that in about 1% of the samples, we may obtain a (model) IQR greater than 6. Perhaps our sample is one of those extreme samples, but this is very unlikely (a 1 in 100 chance). It seems instead that we will need to revise the model to account for more sources of error and hence more variability. Such a model would produce more realistic estimates of the variability of the measurements. See TeacherNoteArmSpan.tp in the Materials section of the website for Unit 6 for an example of a model of a teacher's arm-span that describes more sources of variability. Moving back and forth between assumptions, data, and simulation is how models are generated and revised.





Building a Model

Accounting for Signal

In this activity, students observe a display of the class measurement data and use a measure of center as the best guess of the true measure. The class builds a model of the true measure and compares the model's outputs to the collection of real-world measurements. The goal of the activity is to demonstrate that if measurement were ideal, everyone would obtain the same value.

Whole Group

1. **Review the idea of finding a best guess of the real measure of (name-of-person)'s arm span.**

   Construct: MoV1(a) and MoV2(b)
   This task engages students in thinking about sources of variability.

   a. Make available the arm span displays students created in Unit 1 or supply the students with a new batch of measurement data.

   b. Provoke a discussion about what the real measure of (name-of-person)'s arm span is by asking questions like:

      Q: What do you think the real measure is?
      Q: What is the best guess for the real measure?
      Q: Why aren't all the measurements exactly the same?
      Q: Why didn't the measurers tend to agree?

   Note. Some students may estimate the real measure by looking at the center clump. This is a good way to go about answering the question and should be accepted. But to move the students' thinking, ask if any statistic could be used. Statistics of central tendency should be viewed as sensible. For the question about tendency to agree (the precision of measure, estimated by IQR or average deviation), students may again focus on the shape of the data. It would be good to remind them of ways that we have learned to measure variability, such as the interquartile range (IQR) and average or median deviation.

2. **Introduce the idea of using a spinner to model "real measure."**

   a. Display either a hand-held spinner or project a TinkerPlots spinner while posing questions like:

      Q: If we used a spinner, how could it be designed to model the real measurement? Models approximate what happens in the world. Our goal will be to build a model that explains why we did not all get the same measurement and that approximates our measurements. We are starting our model building by thinking about the real length of _____.

Q: What value should we use to represent the actual value of (name-of-person)'s arm span? (or of any other batch of measurements collected by the class)

Q: For each spin, what value will result?

Q: What will the plot of those values look like?

Q: If we run the model again and again, what will change?

Small Group

3. Distribute the 'What if Measurement Were Perfect?' worksheet to students.

a. Ask students to use the worksheet or TinkerPlots to draw a spinner that models perfect measurement by 30 measurers. If using TinkerPlots, ask your students to run the model a few times.

b. Give your students time to answer the questions on the worksheet.

Whole Group

4. Review the questions from the worksheet.

1. Compare model data to observed data.

a. Ask questions to elicit what students perceive as the strengths and weaknesses of the model.

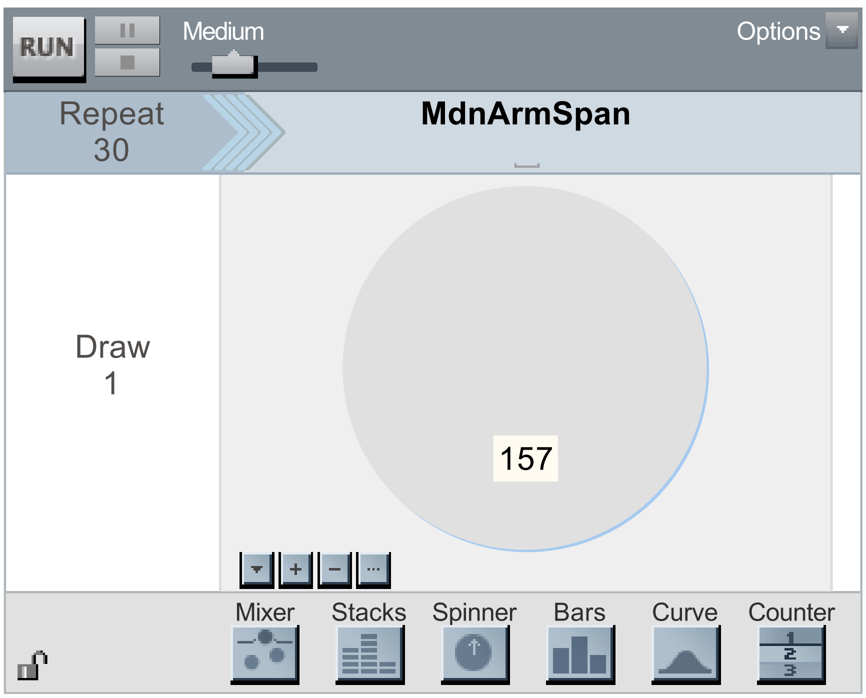
Q: Comparing the model of perfect measurement for 30 pretend measures to the actual measurements made by the data, what do you notice?

Q: What aspect of the observed measurements does the model capture? What aspect of the observed measurements suggests that it is a poor model of the measurement process?

Note. It is important for students to know what the different parts of the model represent. The Repeat (number of repetitions) represents the number of measurers— a class of 30. This can be adjusted if your class is smaller or larger (or it can be taken as an approximation to a class). Each run of the model represents an estimate of what would happen if the class measured again on another day, repeating the same process they used on the first day. Students often find it humorous to consider a variation of the movie *Ground Hog Day* in which they would repeat the measurement process every day.

The spinner model below displays measurement of the arm span of one teacher by 30 measurers. The value, 157 cm., was estimated by the sample median of the class of students who measured the length of their teacher's arm span. Repeated runs of the model yield identical results: There is no variability in either the sample or between samples.

The model of perfect measure does a good job showing the median of best guess of the real measure. But because it shows no variability (no differences among the measurements), it is a poor model of the measurement process.



Incorporating Chance Errors into the Model

Building a Model
Model Revision
Model Extensions
Formative Assessment

In this activity, students identify sources of variability that can not be completely eliminated even when we desire to do so. For each source of variability identified, students use a TinkerPlots chance device (if TinkerPlots is available) or a spinner (or value bars as in the example in the Mathematical Background) to portray the magnitudes of error from that source and the likelihood of each magnitude.

Whole Groups

Introduce the activity: identifying sources of variability.

Construct: MoV1(a) and MoV2(b)
This task engages students in thinking about sources of variability.

a. Remind students that when they measured (name-of-person)'s arm span, not all of their measurements were identical. Point out that many measurements were in a center clump, but many were over- or under-estimates.

b. Provoke a discussion about the sources of variability in measuring (name-of-person)'s arm span by asking questions like:

Q: Why aren't all the measurements exactly the same? Where did the spread come from?
Q: What might be some sources of error?
Q: For each source of error, about how much does that source usually contribute to overestimating the actual value? About how much for underestimating the actual value? What would be an unusually large amount of overestimation? What would be an unusually large amount of underestimation?

2. Distribute the "Thinking About Errors" worksheet to students.

a. Instruct students to examine the table provided on the worksheet. Explain that this is how another student filled out the worksheet when she did this activity.

b. Discuss the ideas of magnitude and likelihood by asking questions like:

Q: What do we mean by magnitude of error? Why are some negative and some positive? Why are some bigger or smaller?
Q: Why are some errors more or less likely? Do you think it is more or less likely to make big or small errors?

Note. Below is an example of how one student filled out her table for a source of error she identified as Ruler Reading Error.

Magnitude of Error	Likelihood of Error	Possible Measuring Behavior
0	Most likely	No error.
+ 1 cm.	Somewhat likely	The ruler reading was really something like 14.1 cm., but the person rounded up to 15 cm.
- 1 cm.	Somewhat likely	The ruler reading was really 14.9 cm., but the person rounded it down to 14 cm.
+ 2 cm.	Unlikely	A person would have had to misread more than once, overestimating each time.
- 2 cm.	Unlikely	A person would have had to misread more than once, underestimating each time.

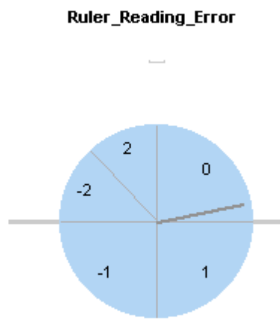
3. **Involve students in considering how to use a spinner as a chance device.**
   a. Use thought-revealing questions to generate discussion. For example:
      Q: For this ruler reading error, how might we use a spinner (or the bars device in Tinkerplots) to stand in for magnitude and the chance of making that kind of error?
      Q: Why is a spinner (or bars or other TinkerPlots devices) a good thing to use? What does it say about our errors (chance)?

Individual or Small Group

4. **Direct students back to the "Thinking About Errors" worksheet.**
   a. Ask students to create a spinner (or bar or some other chance device) model for each source of error they have identified.

Note. The example below shows a spinner model designed by one student for the ruler reading error.

Magnitude of Error	Likelihood of Error
0	25%
+1 cm.	25%
-1 cm.	25%
2 cm.	12.5%
-2 cm.	12.5%

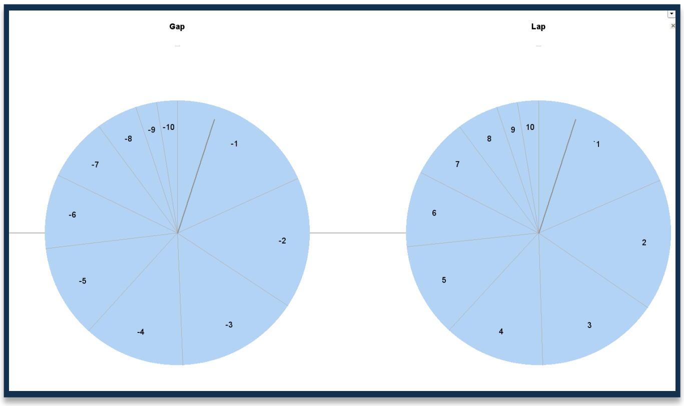


Building a Model
Model Revision
Model Extensions
Formative Assessment

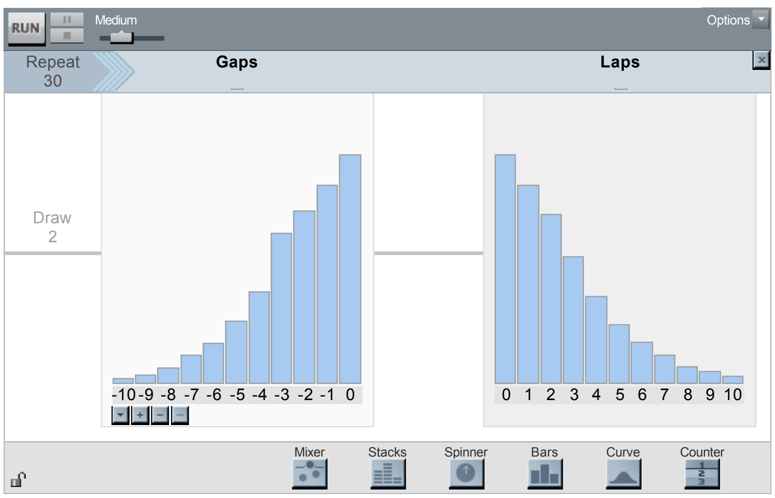
Try to emphasize the relationship between a source of error and the method used for measuring. For example, students may identify iteration errors with the ruler: overlapping the ruler and leaving gaps. Enact this source of error and decide on its range (overestimates of length for gaps and underestimates for overlaps). One tactic might be to have a just noticeable gap, measure it, and then multiply it by the number of iterations. This will be one way to model the effect of the less precise (15 cm. ruler—many gaps) and more precise (the meter stick—one gap) tools on measurement error. Then repeat the process by overlapping the ruler slightly. You might want to share this way of thinking about gaps and laps with the class.

In the following spinner models, the likelihood of smaller gap errors and smaller lap errors is greater than the likelihood of larger errors, because the two kinds of errors tend to "cancel out." The negative sign indicates underestimates (from gaps, when the length is not measured) and a positive sign indicates overestimates (when the laps result in space that is measured more than once). You might also ask students to predict the shape of the gap + lap error, and ask if making all the sectors the same area would matter.

Building a Model
Model Revision
Model Extensions
Formative Assessment



You can also use other TinkerPlots devices to model chance error. For example, a Value Bars representation of gaps and laps might be:



Student Thinking: Sources of chance error

Identifying and estimating the contributions of different sources of error. Usually, students readily identify sources of error that might account for the differences in their measurements of the teacher's arm-span. These typically include:

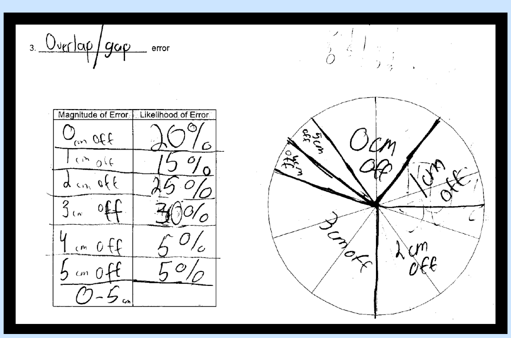
1. Reading error (misreading the ruler)
2. Space skipping (gapping) with ruler, resulting in underestimates
3. Overlapping with ruler, resulting in overestimates
4. Arm movement by the teacher
5. Rounding error (e.g., rounding 152.5 to 153)
6. Calculation error (making a mistake in addition, losing track of place)

After further discussion, students usually focus at least three major sources of error:

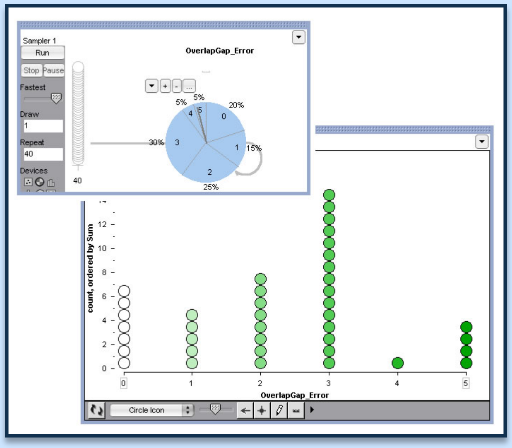
1. Gap-Lap error (gapping/overlapping)
2. Arm movement (arms up/arms down)
3. Rounding error (rounding measurements up/down)

Note. Estimating the magnitudes of the contributions of different sources of errors can be challenging. Have students enact the measurement process and then estimate the possible contributions of different sources of errors. For example, students might put two 15-cm. rulers side-by-side with a small gap and then estimate the maximum magnitude of underestimation. Students can overlap rulers in a similar manner and estimate maximum magnitude of overestimation. Then, because they know that these maximum values are unlikely, they can assign a small area of a spinner model (or just a few objects with an mixer model) to this magnitude. Often, students are less diligent about estimating magnitudes, but as long as they make some reasonable guesses, the models are likely to approximate the observed distribution of values.

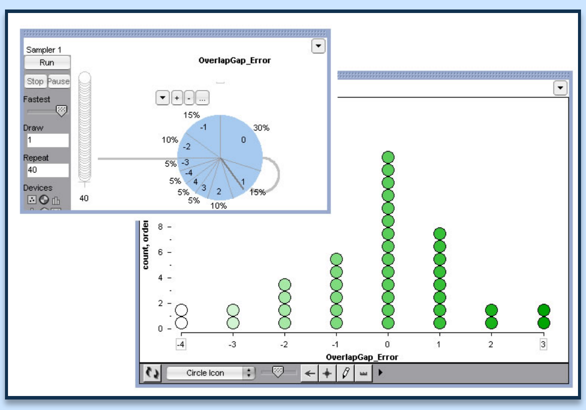
Designing spinners for each source of error. Students' initial spinner designs often focus on the magnitudes of a source of error, but ignore direction. For example, the image below shows a spinner designed by a group of fifth-grade students to represent ruler error (overlaps or gaps when iterating the ruler). Although students noted, "There are a lot of measurements in the middle, and some measurements a little over and some measurements a little under," the first spinner they designed represented smaller errors, such as 0-2 cm., as less likely than larger error, such as 3 cm. They also did not differentiate between over- and underestimates.



When students ran their spinner model with TinkerPlots (see below), they realized that overestimates were represented, but not underestimates. And, larger magnitude errors were more likely in their model, but not in the data.



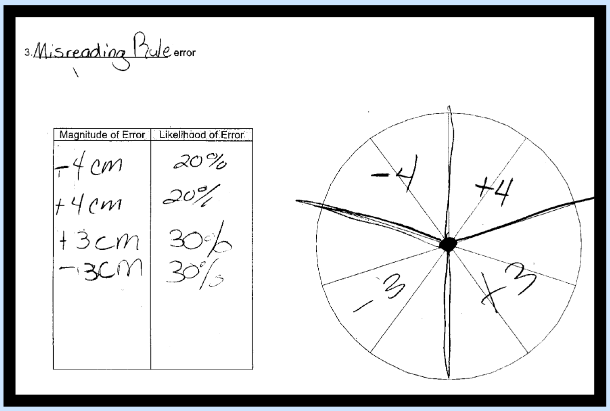
When students considered these factors, they redesigned their spinner, with the help of their teacher who suggested using signs to represent over (+ values) and under (- values) estimation. The following image displays their TinkerPlots implementation of a spinner with direction of error and with greater likelihoods of small (0,1) magnitudes of ruler error.



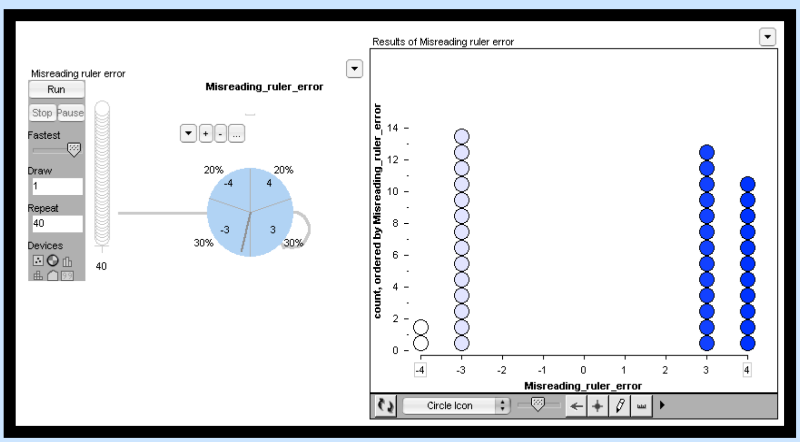
An overlap/gap error spinner that represents large proportions for small errors and small proportions for large errors

Building a Model
Model Revision
Model Extensions
Formative Assessment

Discrepancies between intended and actual representations of spinners are commonplace and running TinkerPlots models often helps students recognize these discrepancies. For example, another group created the following spinner model to represent misreading the ruler. Their model captures the distinction between over- and underestimates, but it does not represent their sense that often, there were very small or no reading errors. After running the model, students could see a gap in the display that suggested small errors were not represented at all in their spinner model.

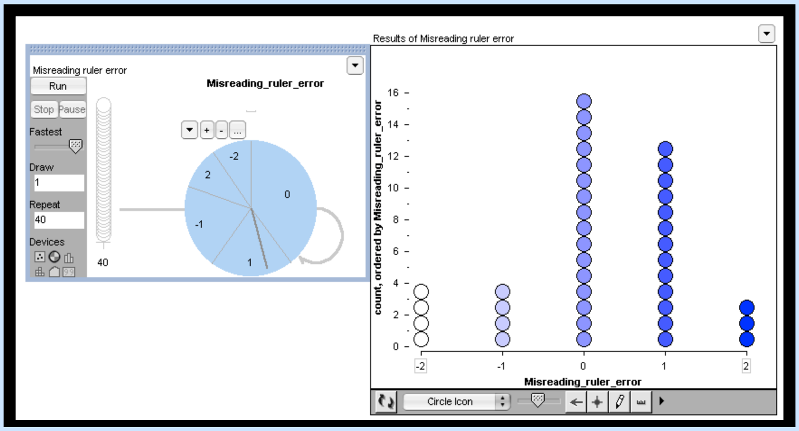


A misreading ruler error spinner designed by a group of 5th grade students.



The distribution of the outcomes by running the misreading error spinner 40 times

When students considered these factors, they redesigned their spinner, with the help of their teacher who suggested using signed values to represent zero error and small reading errors. The following figure displays their TinkerPlots implementation of a spinner with zero error and small reading errors.



Combining Signal and Error

In the previous activity, students built models for different components of variability. Now students develop models as combinations of the estimated true value and the chance errors due to factors such as ruler iteration and calculation errors.

Building a Model
Model Revision
Model Extensions
Formative Assessment

Constructs: MoV1, MoV2, MoV3, and MoV4

Whole Group

1. Introduce the activity: combining chance devices.

a. Remind students that in the last activity, they created a spinner for each source of error.

Use thought-revealing questions to generate discussion about how to combine the sources of error.

Q: Now that we have a spinner model for each source of error, how should we combine them?
Q: Do you think we will ever have a situation in which the combination of errors is zero? Why or why not?

Individual or Small Group

2. Use TinkerPlots to sum the error for each chance device.

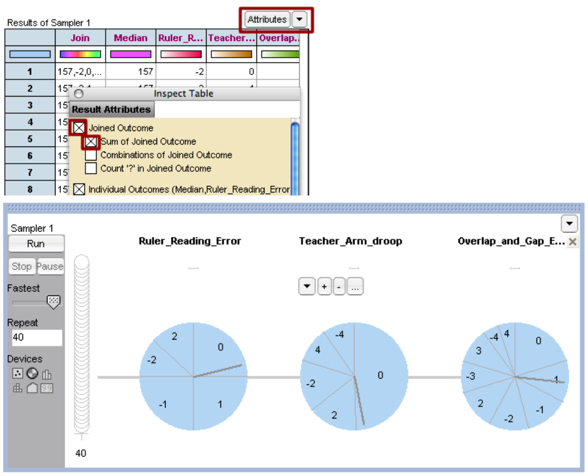
a. Use TinkerPlots Sampler to create a model with at least two different sources of error. Run the model and have TinkerPlots sum the outcomes. Have students look at the results of the Sampler. If TinkerPlots is not available, have students construct spinners and sum the outcomes on the "Model Measure" worksheet.

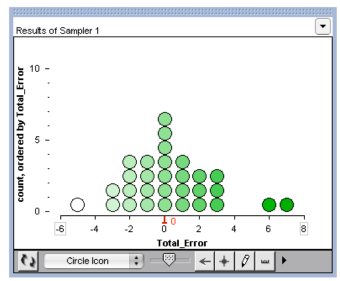
3. Circulate among groups and ask thought-revealing questions like:

Q: What did you notice when you looked at the sum of errors?
Q: What was the largest sum?
Q: What was the smallest sum?
Q: Did you ever get a sum of zero? How could that be?

Note. An easy and productive way of combining errors is to sum them. It is important to help students see that any one trial does not mimic the aggregate and that some trials may be very unusual. A TinkerPlots display of a model of the arm span errors is shown, along with the output obtained from 30 repetitions of the model. Notice that for some repetitions of the model the error was zero, and for others, large.

Building a Model
Model Revision
Model Extensions
Formative Assessment





Results of Sampler 1

	Ruler_R...	Teacher...	Overlap...	Outcome	Total_Er...	<new>
1	-2	0	-1	-2,0,-1	-3	
2	2	4	1	2,4,1	7	
3	2	0	-2	2,0,-2	0	
4	0	-2	3	0,-2,3	1	
5	0	-4	3	0,-4,3	-1	
6	1	4	1	1,4,1	6	
7	1	0	1	1,0,1	2	
8	1	0	-1	1,0,-1	0	
9	-1	-2	1	-1,-2,1	-2	
10	0	0	2	0,0,2	2	
11	-1	2	2	-1,2,2	3	
12	1	0	-4	1,0,-4	-3	
13	0	0	0	0,0,0	0	
14	1	0	0	1,0,0	1	
15	1	-4	3	1,-4,3	0	
16	-1	2	-2	-1,2,-2	-1	
17	1	-2	-4	1,-2,-4	-5	

Whole Group

Building a Model
Model Revision
Model Extensions
Formative Assessment

4. Provoke a discussion about how to model measurements.

a. Ask questions to understand what students are noticing about the error data.

Q: Now that we have a way of representing sources of error and how much each source typically contributes to a measurement error, how could we model our measurements?
Q: What should the shape of our data be?
Q: What is the shape of the error data? What is the shape of the observed measurements? How are they the same? How are they different?

Note. Let students know that one way to model the measurements is to think about the center of the collection, as determined by the actual length of (name-of-person)'s arm span. Although students do not know what that length really is, they can guess. So, one part of the model is the estimate for the actual length. (This is what students did when they modeled true measurement in the first activity.)

Then, students need to think about what causes the variability around the actual length—why the measurements are sometimes overestimates and sometimes underestimates. The variability comes from error—just by chance, sometimes measurements are greater than the real measure and sometimes measurements are less than the real length. This relationship is described below:

Observed Measurement = Best Guess of True Measure + Total Error

The error part of the equation can be broken into parts:

Total Error = Error-from-Source-1 + Error-from-Source-2+…….

For example:

Total Error = Gap-Lap Errors + Ruler Reading Errors + Counting Errors

Small Group

5. Have students run their model several more times.

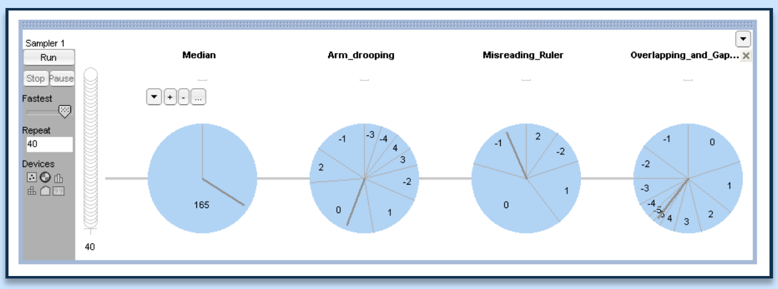
a. Ask groups to run their model once with 30 repetitions (each repetition represents a measurer).

b. After the initial run, ask the groups to run the models a few more times. During this time, pose questions to the groups, like:

Q: What do you notice?
Q: Does the model do a good job of approximating the real measurements?
Q: What is good about it?
Q: What needs improvement?
Q: Are all the runs exactly the same?
Q: What changes?
Q: What stays the same?

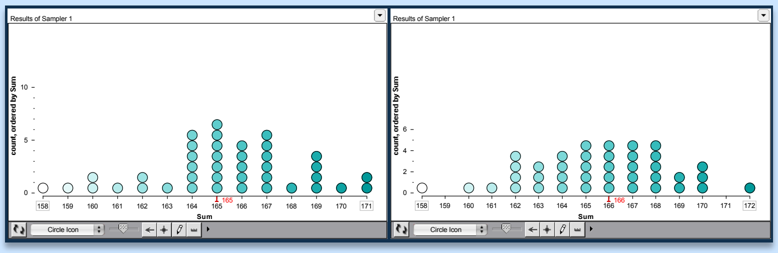
Student Thinking: Constructing models as a sum of components

Modeling measurement: putting components together. A group of 5th grade students designed their model of the measurements of their teacher's arm-span by combining their estimate of the real length (the sample median) with the other spinner models of error: arm-movement (droop), misreading the ruler, and ruler use (overlap & gap) as below.

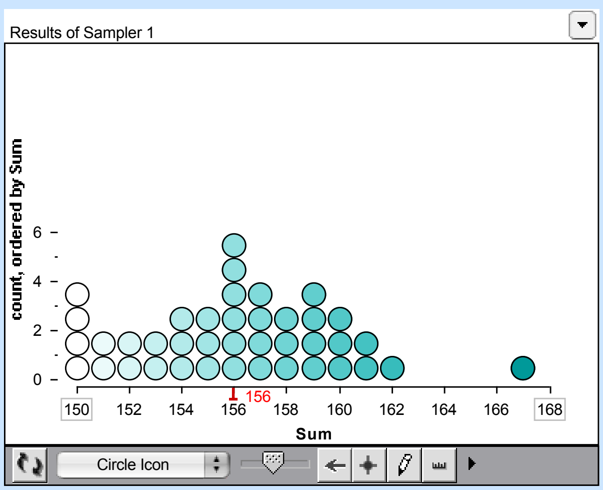


Students ran the model 40 times to simulate 40 measurers. Each measure was represented by the sum. When the teacher asked students to compare their real measurements to the modeled measurements, a student said, "The spinner model was pretty accurate because most of them were close together." The student compared the distribution of the real measurements and the distribution of modeled measurements based on the center clump. Different students used different criteria to compare the two distributions: median, shape, minimum and maximum measurements, etc.

Because measurements will vary from sample to sample (or from one time of measurement to the next), the class should run the model several times and observe how the distribution changes and stays the same from model run to model run. For example, here are the distributions of the first two 40 runs of the model above:

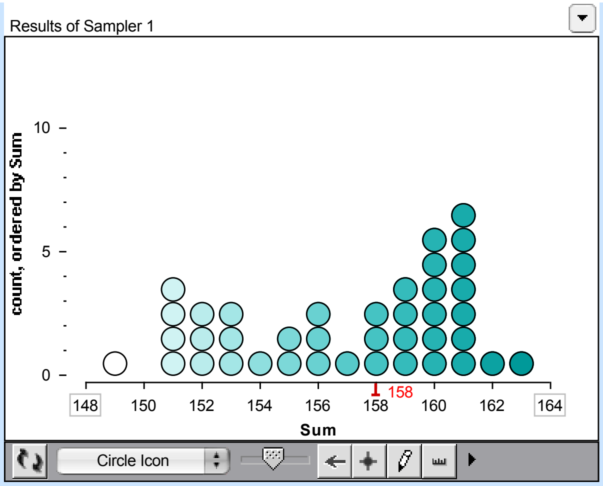


Running models. If students run the model several times, they will get some unusual runs and some more routine ones. Here are some unusual runs for a sample of arm span measurements with a median of 158 cm.:

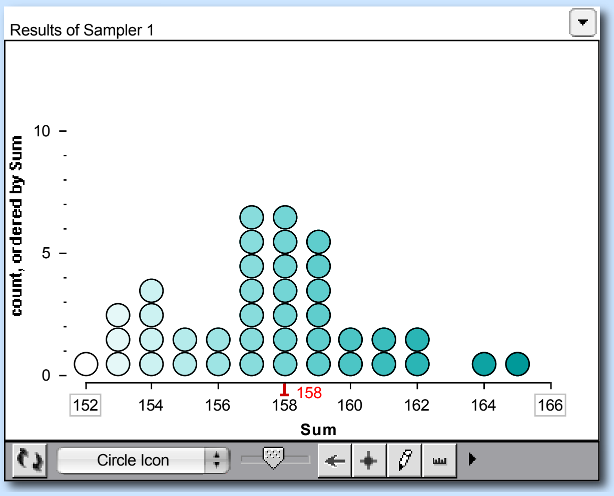


The model measurements are skewed to the left (above) and right (below)

Building a Model
Model Revision
Model Extensions
Formative Assessment



Here is a more likely sample run:



Because each model run represents a sample of measurements, and there is always sample-to-sample variability just due to chance, encourage thinking about how good a model is by looking at its behavior over many samples, as described in the next section about model fit.

Model Fit and Model Revision

In this activity, students compare their models' predictions about center and precision to those describing the real-world measurements in order to judge if the model is consistent with the data and therefore "good."

The aim of this activity is to get students to relate the output from the model to the data. The first step is to establish correspondences between the model results/predictions and the real measurements.

Whole Group

Constructs: MoV1, MoV2, MoV3, and MoV4

1. **Introduce the activity: model fit.**

   a. Use questions to generate discussion about model fit.

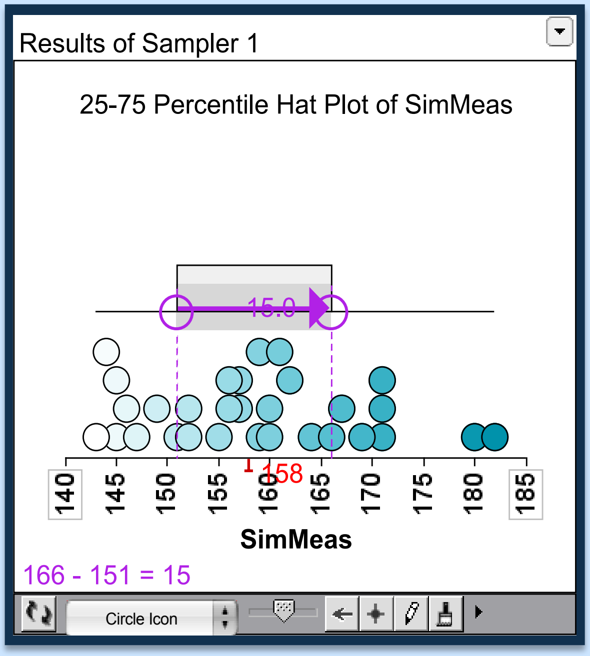
      Q: What happens when you run the model again and again?
      Q: What about these model runs is worth keeping track of?
      Q: Is the shape of the data about the same?
      Q: Is the center about the same?
      Q: What about the precision?
      Q: What are the model's strengths?
      Q: What are the model's weaknesses?

   Note. Real data samples vary from sample to sample, as we have seen in Units 4 and 5. No two samples of measurements of the same person are exactly alike, and we have seen that products vary as well (e.g., not all people can match a target rate, not all packages of toothpicks or candies match the target value). To mimic this characteristic of real samples, we can run our model repeatedly. Each model run produces a simulated sample. By keeping track of each sample's center (e.g., median) and variability (e.g., IQR), students can see if the estimates from their model tend to agree with their real-world measurements when sampling is taken into account. For example, is the average value of the medians in the model's sampling distribution close to the real-world median? Is the average value of the IQR's in the model's sampling distribution close to the real-world IQR?

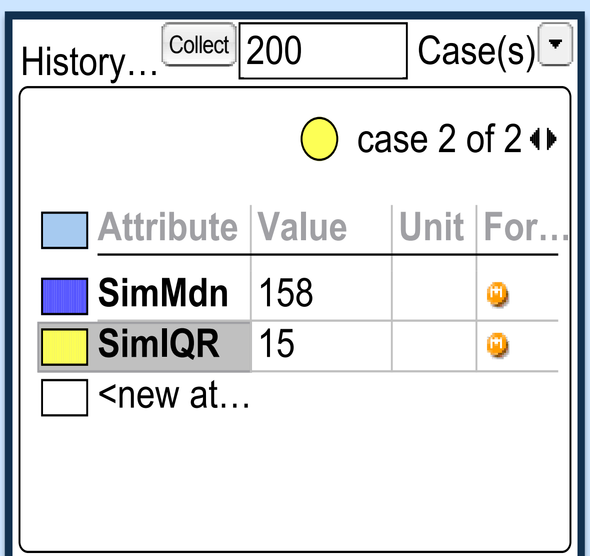
Building a Model
Model Revision
Model Extensions
Formative Assessment

Classroom Talk: Guiding model fit

Two measures (statistics) from one run of a model of the measurement of a teacher's arm-span: The model simulates the measurements made by 30 students. The median and IQR are found for this simulated sample.

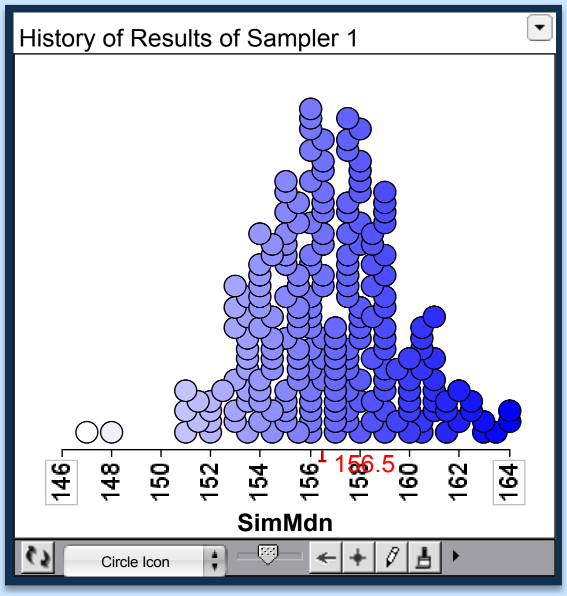


The model is run repeatedly and these statistics are collected.

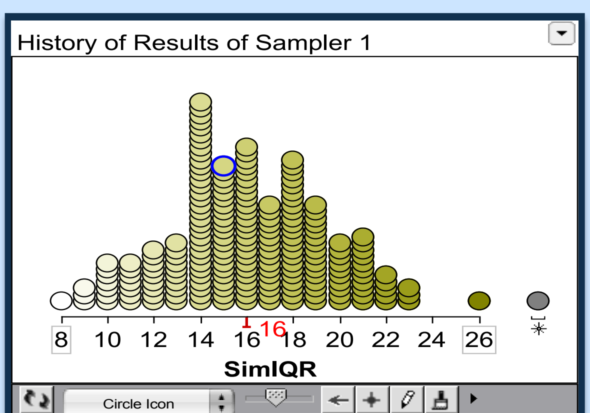


Building a Model
Model Revision
Model Extensions
Formative Assessment

The distribution of the medians of repeated samples of 30 is displayed. Be sure that students know that each value represents the median of the 30 simulated measurements. The medians ranged from about 147 to 164, and the average value was 156.5. This accorded well with the observed measurement median of 157 cm. So, the model was a good fit for the center.



The distribution of the IQR for each sample of size 30 is also displayed. The IQR estimates ranged from 8 to 26 and were centered at 16. This was also consistent with a real-world sample value of 17. One could expect 17 just by chance, if the model were true. So the model was a good fit for precision of the measurements as well.



Model Extensions

Building a Model
Model Revision
Model Extensions
Formative Assessment

In this activity, students are challenged to create different types of models, including "bad" models that result in about the same center, but different shapes than real-world data.

Whole Group

1. **Pose a challenge to students by asking them to make a "bad" model.**

   a. Ask students to make a "bad" model with a similar estimate for (name-of-person)'s arm span and a similar range of measurements.

   Note: Although it might seem like a step back from model fit, intentionally constructing a bad, poorly fitting model provides students further opportunity to learn about the relation between the likelihood of particular magnitudes of error and the resulting shape of the distribution. For example, if students design chance devices where the likelihoods of large errors are high and small errors low, the resulting distribution will have a U shape.

Individual or Small Group

2. **Allow students to make their bad models.**

3. **Ask students to share their models with the class.**

   a. Use questions to generate discussion. For example:

      Q: How does your "bad" model compare to the real data?
      Q: What did you have to do to your model to make it bad?
      Q: How do you know it's bad?

Whole Group

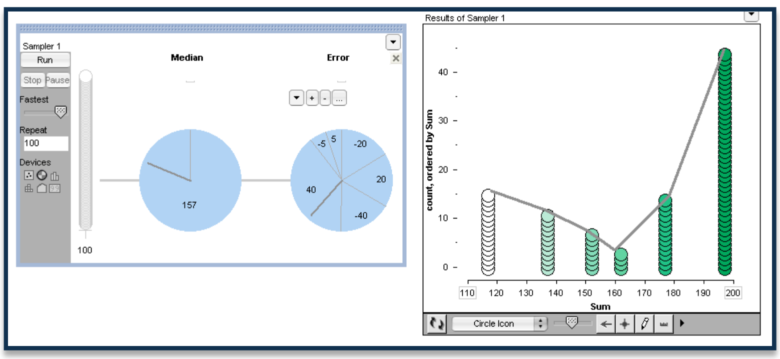
4. **Pose alternative challenges to students. For example**:

   a. Suppose overestimate errors were for some reason more common for each source of errors. What do you think the effect would be? Would the mean of the measurements stay about the same or would it change? Would the spread of measurements change? Would the shape of all the values change? Why do you think so? Create a model simulation to find out.

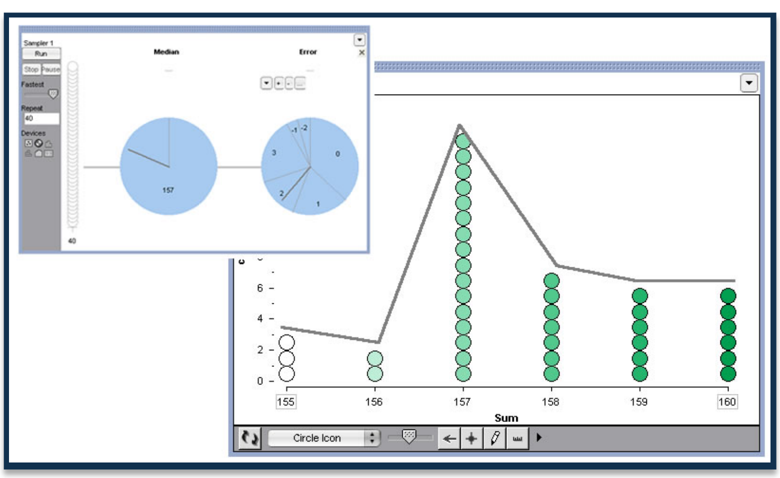
Note: The intention here is to help students learn more about the relation between the structure of a chance device and the empirical outcomes observed—another opportunity to strengthen the connection between empirical and theoretical probabilities. Here by controlling the structure of the chance devices, students can make models that are worse, instead of better. This may seem perverse, but it tends to push on student understanding in productive ways.

b. People often improve the accuracy of their measurements with practice. This means that they get closer to the real height (or whatever the measurement is). How might you model this result with spinners?

Note. Below, a bad model created by students kept the median but made larger errors more likely and smaller errors less likely.

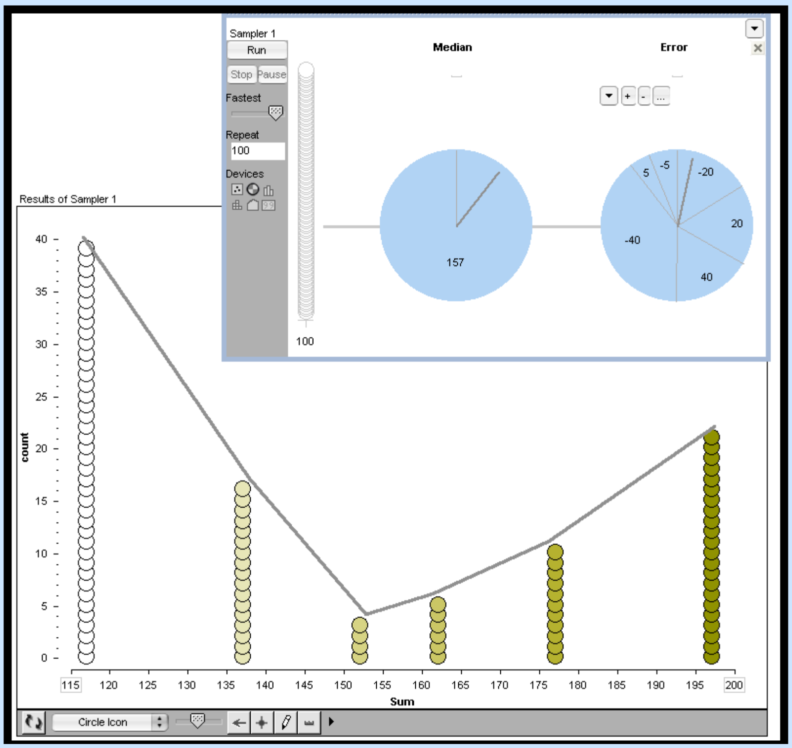


This symmetry of a distribution can be affected if the likelihood of an error of overestimate is greater than the likelihood of an underestimate. This is called skew.



Student Thinking: Bad models

Bad models. Fifth-grade students designed a bad model. The teacher asked students how they could design spinners so that the shape of the modeled measurements was different from the actual measurements. Here is an example of a bad model by a 5th grade student:



The student said, “I made 40 and -40 so big that they result in larger errors.” The class agreed that her model was pretty bad because the shape of the modeled measurements were “pushed down in the middle,” which was different from the shape of the original measurements, but “the center was about the same.”

Formative Assessment

1. **Administer the quiz.**

2. **Use the scoring guides to score student responses.**

3. **Use the Cookie Dough Scoop to generate a discussion of how to build a componential model of the process of making cookies.** One component is the target diameter of the cookies: 10 cm. The other component is random error due to chance variations in the process of scooping the cookie dough.

   a. **Select student responses to compare and contrast.**

   While scoring question 1, select 2-3 different responses to use in the conversation. One response should label the entire spinner with 10 and the other two should reflect other forms of thinking, such as confusing the number of cookies with the target diameter. While scoring question 2, select 3-4 different responses to use in the conversation. One type of response may simply create sectors corresponding to the different diameters represented in the display of the batch of cookie diameters. Another type of response recognizes that the model is composed of a combination of target size and error, and creates a range of positive (over) and negative (under) errors, with larger areas dedicated to smaller errors. Variations on this type of response may include equal area sectors or a lack of negative signed errors. While scoring question 3, select 4-5 responses that span the range of response levels in the scoring exemplar. It is particularly important to contrast student justifications of model fit, especially responses that rely upon (a) literal similarity of cases, or (b) similarity between a single model run's statistics and real data statistics (e.g., same median, close IQR) or (c) similarity between the sampling distribution's statistics over repeated runs of the model and those of the real sample. While scoring question 4, look for student responses that lead to improvements in the original model. Student responses for the fifth question are apt to range across all 3 multiple choices. Give students the opportunity to explain their reasoning (choice 2 is most likely).

   b. **Prepare questions to support and guide student thinking.**

   For example:

Q: If the cookie manufacturing process were perfect, what would the diameter of every cookie be?

Q: What would a display of the diameters of 15 cookies look like if the process had no error?

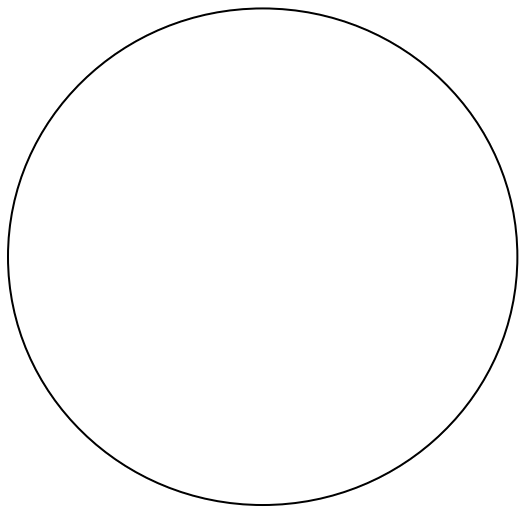
Q: What will happen when the model in question 2 is run one time? Many times?

Q: If the scoop process improves, what do you expect to change? Why do you think so?

c. **Use an Assessment Conversation to help students build and revise the model of the cookie dough manufacturing process.** To do so, start with the student responses to the first question and then consider each question in turn. Be sure to include less sophisticated responses for each question, highlighting what is helpful about these responses and then providing students with opportunities to build on that thinking.

What If Measurement Were Perfect?

Draw a spinner that models perfect measurement—measurement where the measurer always finds the real length of the arm-span.



1. Why did you draw the spinner that way?

2. If you run the spinner more than once, what happens? What does each spin represent?

3. Is the model of perfect measurement a good model? What does it do a good job showing? What does it do a poor job showing?

Thinking About Errors

We decided that the reason our measurements were not completely precise was because we made mistakes when we measured, no matter how hard we tried.

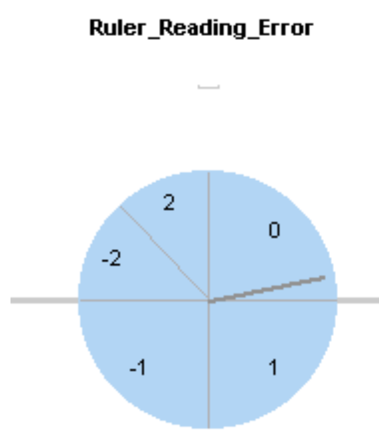
Sources of Error

1. Ruler reading error

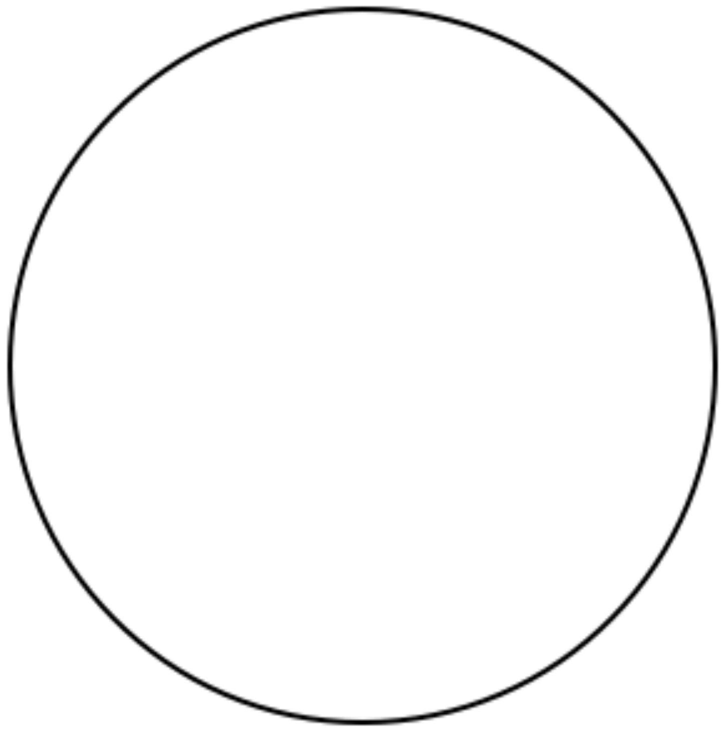
Magnitude of Error	Likelihood of Error	Possible Measuring Behavior
0	Most likely	No error.
+ 1 cm.	Somewhat likely	The ruler reading was really something like 14.1 cm., but the person rounded up to 15 cm.
- 1 cm.	Somewhat likely	The ruler reading was really 14.9 cm., but the person rounded it down to 14 cm.
+ 2 cm.	Unlikely	A person would have had to misread more than once, overestimating each time.
- 2 cm.	Unlikely	A person would have had to misread more than once, underestimating each time.

This is a spinner model designed by one student for the reading error:

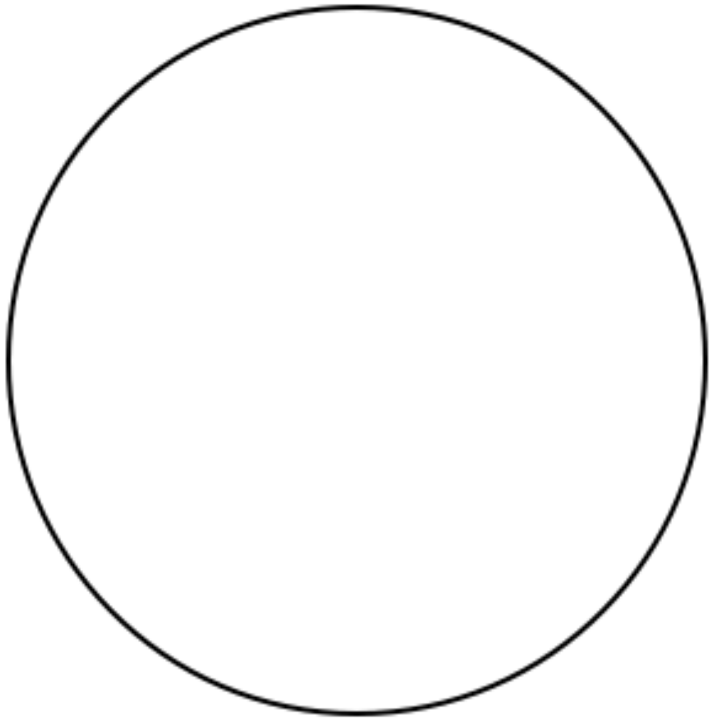
Magnitude of Error	Likelihood of Error
0	25%
+1 cm.	25%
-1 cm.	25%
2 cm.	12.5%
-2 cm.	12.5%



Next, identify TWO more sources of error and create a spinner model for each using the blank error worksheets.



Likelihood of Error	
Magnitude of Error	



Likelihood of Error	
Magnitude of Error	

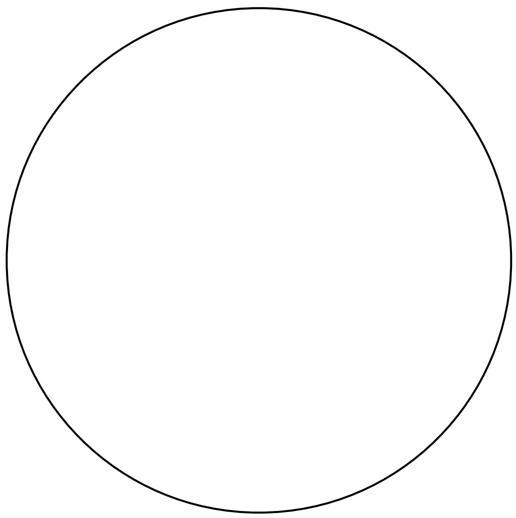
Model Measure

	Median	Ruler reading error	______error	______error	Modeled Measurements
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					

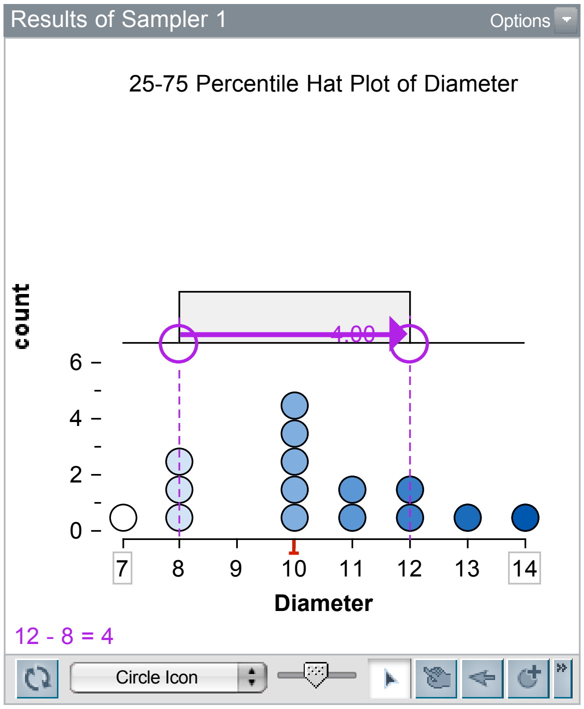
Unit Quiz

Angeline is learning how to make chocolate chip cookies from her mom. Her mom tells her that it is important that the size of the cookies is the same. The cookie dough scoop is supposed to make cookies of 10 cm in diameter. Her mom demonstrates how to use a cookie dough scoop. Angeline puts 15 scoops on a baking sheet and bakes them.

1. If Angeline is really good at scooping dough and her cookies are all the same size, how would you design a spinner to model her perfect scoop?



She measures the diameters of the 15 cookies. Here are the diameters of the 15 cookies.



2. Angeline and her mom think about why Angeline's cookies are all different sizes. Angeline figures that when she uses the scoop over and over again, the dough sticks to the scoop and that makes cookies smaller. Sometimes, she dips the scoop too hard and that makes cookies bigger. Angeline calls this scoop error. How would you design a spinner to model the scoop error?

Unit Quiz

3. Run your model. Do you think your model is a good one?

☐ Yes

Why?

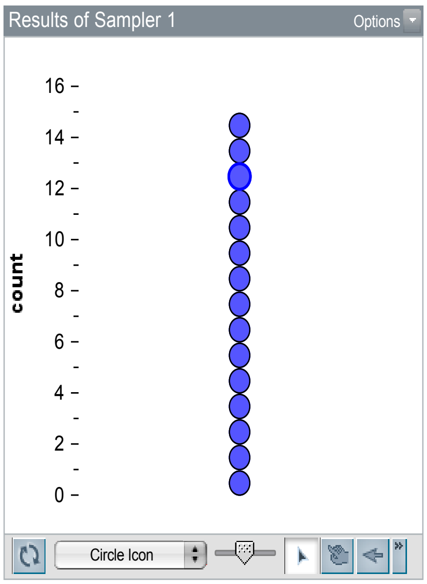
☐ No

Why not?

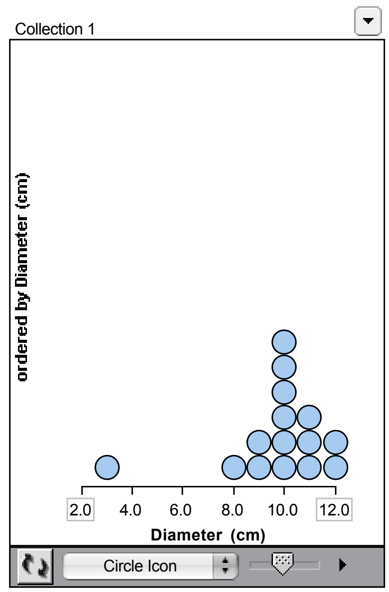
4. If not, how would you change your model?

5. Her mom gives Angeline a tip that Angeline should put the scoop in water so that dough does not stick on the scoop. Angeline follows her mom's advice. Which display of the resulting cookie diameters is closest to the results that you might expect if her mom's advice is good?

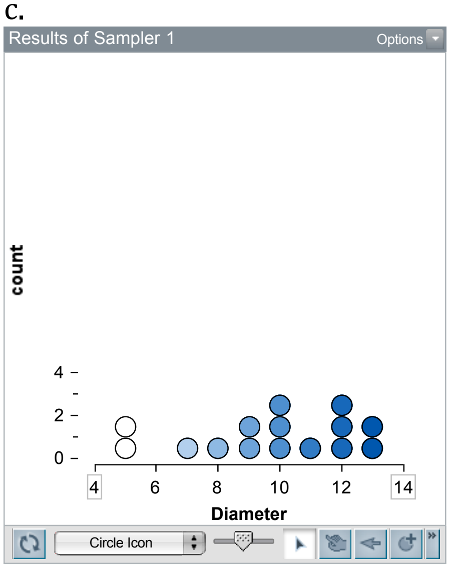
a.



b.



c.

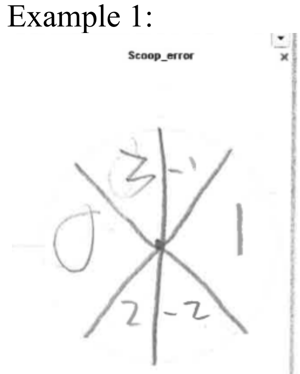


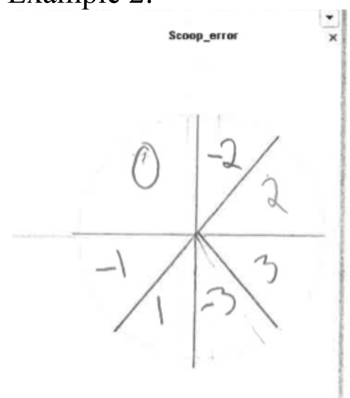
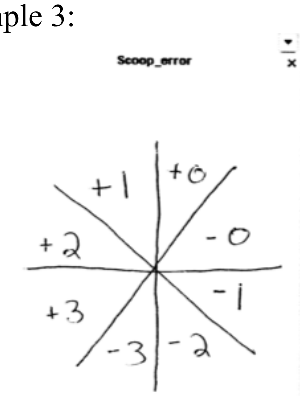
Why?

Modeling Cookie Dough

Question 1: Modeling Cookie Dough and Modeling Variability (MoV)		
Level	**Performance**	**Example**
MoV(3a)	**Use chance device to model uncertain outcomes.**	• Student labels the entire spinner "10" or "10 cm." • Student divides the spinner into sections, but all sections are labeled "10" or "10 cm."
NL(ii)	Response is relevant but unclear. Student may also answer yes or no without providing an explanation.	• Student treats the spinner as a literal cookie. • Student does not think that spinners can model perfect measurement. • Student draws sections on the spinner representing different measurements.
NL(i)	Response is irrelevant, unclear, or a restatement of given information. Student doesn't compare the real data to any attributes of the model. Includes responses where student does not attempt to design a spinner.	• "I'm not sure." • "?"
M	Missing response	

*Mock student responses

Question 2: Modeling Cookie Dough and Modeling Variability (MoV)		
Level	**Performance**	**Example**
MoV(4b)	**Create model of total variability as a composition of chance (and perhaps, constant) devices.** Students construct a spinner where each section represents a different "scoop error".	• Example 1: 

		• Example 2:  • Example 3: 
MoV(3a)	**Use chance device to model uncertain outcomes.** Student uses the blank spinner to model the entire system.	• All sections of the spinner represent different measurements (as opposed to error). For example, the spinner might have sections for 9, 10, or 11 cm.
NL(ii)	Response is relevant but unclear. Student may also answer yes or no without providing an explanation.	• Student treats the spinner as a literal cookie. • Student does not think that spinners can model perfect measurement. • Student draws sections on the spinner representing different measurements.
NL(i)	Response is irrelevant, unclear, or a restatement of given information. Student doesn't compare the real data to any attributes of the model. Includes responses where student does not answer whether the	• "I'm not sure." • "?" • "27"

	model is good.	
M	Missing response	

*Mock student responses

Question 3: Modeling Cookie Dough and Modeling Variability (MoV)		
Level	**Performance**	**Example**
MoV(5a)	**Judge model fit in light of variability across repeated simulation with the same model.**	• "It's a bad model because if we ran it a bunch of times, we would keep getting the outlier even though it's not likely."* • It is a good model, because even though the mean changes with each run it is mostly very close to 10.*
MoV(4c)	**Compare model output to data and judge adequacy.**	• I ran it once and its shape was a lot like the data, so it is a good model. • The model median and the data median match. • The model IQR is pretty close to the observed IQR.
MoV(3b)	**Evaluate fit of chance device by appealing to relations between simulated and observed values.** Student compares the values or the shape of the modeled data to the display of the original data and emphasizes resemblance.	• "Yes, because the results are almost alike." • "Yes, because it looks like the original one." • "Yes, you don't know if you're getting it right above or below the measurement. Just like with the measurement."
MoV(3b-)	**Evaluate fit of chance device by appealing to relations between simulated and observed values.** Student looks for literal resemblance in values OR students look for the structure of the spinner to literally resemble the original data	• "Yes, all the numbers are the same." • "Yes, because it has the same numbers as the outcome." • "No, because it doesn't show all the answers." • "Yes, it shows exactly the measurements." • "Yes, he has more on 158 because there are a lot of 158s on the spinner." • "No, he needs more 158 pieces on the

		spinner."*
MoV(3a)	**Use chance device to model uncertain outcomes.**	• "Yes, because you used a spinner to show the chances of each measurement"* • "Yes because it gives each number a probability of being landed on." • "It has an equal chance of landing on each of them."
NL(ii)	Response is relevant but unclear. Student may also answer yes or no without providing an explanation.	• "Spinners are good models." * • "No because there are not even."
NL(i)	Response is irrelevant, unclear, or a restatement of given information. Student doesn't compare the real data to any attributes of the model. Includes responses where student does not answer whether the model is good.	• "Maybe. I'm not sure." • "?" • "27"
M	Missing response	

Question 4:

Students' responses to question 4 are likely to be highly variable, and are dependent on the decisions they made in questions two or three. In general, look for and encourage students to look for ways to improve their model. Here are few likely improvements students might think about:

- Changing from using one spinner to using multiple spinners to account for variability
- Changing the signs of the errors to account for over and under estimates
- Changing the size of the sectors to account for probability structure (giving smaller errors the larger sectors since they are the most likely)

Question 5: Modeling Cookie Dough and Modeling Variability (MoV)		
Level	**Performance**	**Example**
MoV(2a)	**Describe how a process or a change in process affects variability.** **Student describes that the center would likely stay the same, but the variability would decrease.**	• "b, because it is a clump in the middle, but the clump is tighter than the others." • "not a, because 10 is still the most, and it is still not going to be perfect. So it must be b or c. I think it's b."
NL(ii)	Response is relevant but unclear. Student may also answer yes or no without providing an explanation.	• "Spinners are good models." * • "No because there are not even."
NL(i)	Response is irrelevant, unclear, or a restatement of given information. Student doesn't compare the real data to any attributes of the model. Includes responses where student does not answer whether the model is good.	• "I'm not sure." • "?" • "27"
M	Missing response	