OVERVIEW

In this activity, students build models of real-life situations and use data from the models to estimate probabilities. Students should already have some prior experience creating models in TinkerPlots, as in the activity "Building a Data Factory," because they are not told how to do this in the worksheet. Students should also have prior experience using simulation data to estimate probabilities. This activity can be used as a "capstone" activity, allowing students to apply skills and concepts they learned in prior activities.

Activity Time: One to two class periods

Objectives

- Design and build samplers in TinkerPlots to model real-life situations.
- Use data from simulations to estimate probabilities.
- Develop an understanding that frequencies generated from simulations allow you to give a range for an unknown probability, and that these ranges are narrower for larger sample sizes.

Common Core Standards Addressed

Approximate the probability of a chance event by collecting data on the chance process that produces it and observing its long-run relative frequency, and predict the approximate relative frequency given the probability.

Grade 7, Statistics and Probability Standard 6

Develop a probability model (which may not be uniform) by observing frequencies in data generated from a chance process.

Grade 7, Statistics and Probability Standard 7b

Design and use a simulation to generate frequencies for compound events.

Grade 7, Statistics and Probability Standard 8c

Prerequisites

- Experience building samplers in TinkerPlots
- Understanding of how to estimate probabilities using simulation data (could be learned by doing the activity "Wink, Blink, and Stare")
- Understanding of how to use sample spaces to calculate theoretical probabilities (could be learned by doing the activity "Sum of Two Dice")

Materials

- Modeling Challenges worksheet (one copy per student)

LESSON PLAN

Introduction (10 minutes)

Review the process of building a TinkerPlots model by referring to other activities students have already done. Tell them that the challenge of estimating probabilities from any model is that each run of the sampler generates a different result.

What then do you give for the probability of the event?

Discuss this and encourage students to give the answer in range form, rather than as a precise number. In the **Letters** model, for example, students might state that the probability is somewhere between 1% and 3%, because that's about how often they'll get the word *CAT* when the model is repeatedly sampled. Refer to an activity students have already completed in which they have estimated probabilities from simulation data, and how they came up with those probability estimates.

Another main idea is that the larger the sample, the less variability there is in the results from trial to trial. This means that you can give estimates with a narrower range for the unknown probability as the sample size is increased. You'll want to encourage students to draw larger samples, as many students will not believe that the sample size makes a difference, and will want to use a consistent sample size of, for example, 100.

One way to motivate this is to let your students use sample sizes that make sense to them, and then have students discuss why they chose the sizes they did. Comparing the range of results students got when sampling with different sizes (for example, 100 versus 1000) will help them see that they get less variability in results with the larger sample sizes.

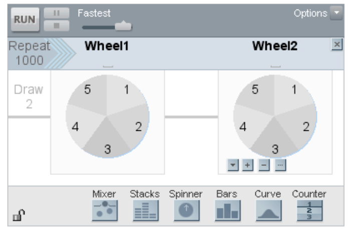
Student Work at Computers (20–80 minutes)

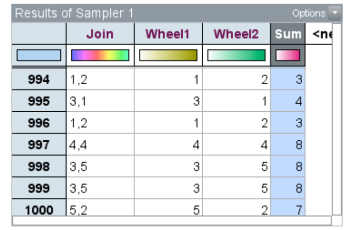
This activity can be modified in many ways. You can assign students just one or two of the four problems, or you can let them choose one or two. Alternatively, you can assign different problems to different students, or if you have time, have everyone make all four models. You might also choose to use one of the four as an example to do with the whole class, and then assign the others or let students choose.

The next four sections each describe a problem in detail, including the answers. You may not expect your students to compute theoretical probabilities for such complex situations, but they are provided here for your benefit.

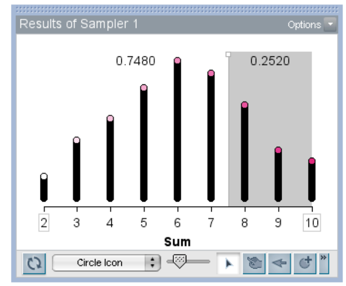
Carnival Game

This situation is very similar to that of finding the sum of two dice, which should be straightforward for most students to model. Estimates for the probability of a sum of 8 or more should range from 21% to 27% for sample sizes of 500.

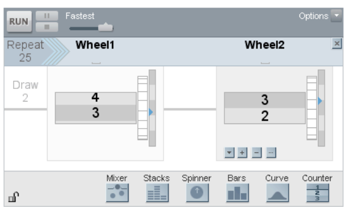


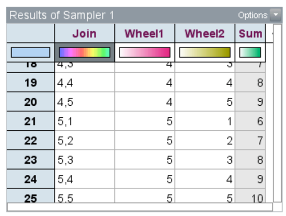

Results of Sampler 1

	Join	Wheel1	Wheel2	Sum
994	1,2	1	2	3
995	3,1	3	1	4
996	1,2	1	2	3
997	4,4	4	4	8
998	3,5	3	5	8
999	3,5	3	5	8
1000	5,2	5	2	7

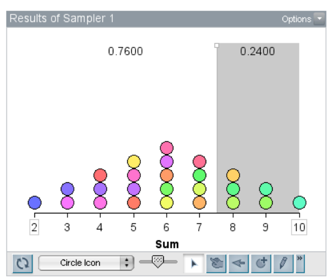


Students could also figure out the theoretical probability if they have previous experience creating organized lists or tables, as in the activity "Four-Child Families." You can also generate the sample space in TinkerPlots using the **Counter** device.




Results of Sampler 1

	Join	Wheel1	Wheel2	Sum
19	4,4	4	4	8
20	4,5	4	5	9
21	5,1	5	1	6
22	5,2	5	2	7
23	5,3	5	3	8
24	5,4	5	4	9
25	5,5	5	5	10



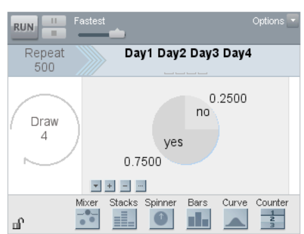
As shown at left, you can replace spinners with counters. The results will start with the outcome 1, 1 and end with 5, 5, for a total of 25 possible outcomes.

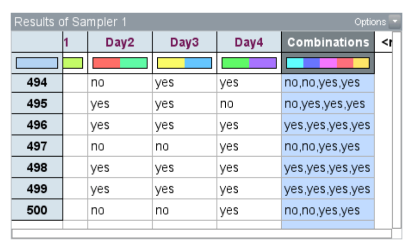
Looking at the plot of this sample space, students can see that 6 of the 25 equally likely possibilities give a sum of 8 or more, so the theoretical probability is

$$\frac{6}{25} = 24\%.$$

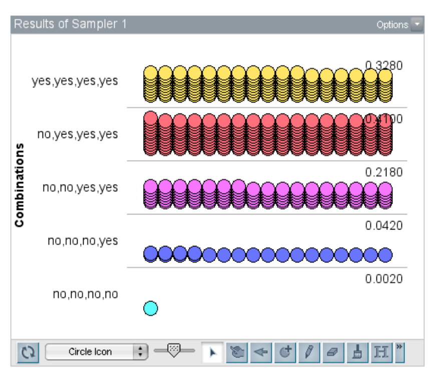
Rain

Estimates for samples of 500 should range from 28% to 34%. The model shown here uses a *Combinations* attribute that combines all of the sequential variations of the same outcome.




Results of Sampler 1

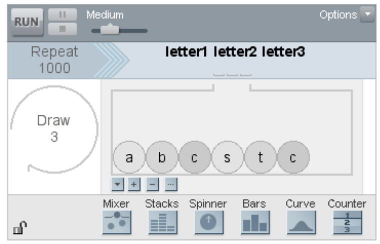
	Day1	Day2	Day3	Day4	Combinations
494		no	yes	yes	no,no,yes,yes
495		yes	yes	no	no,yes,yes,yes
496		yes	yes	yes	yes,yes,yes,yes
497		no	no	yes	no,no,yes,yes
498		yes	yes	yes	yes,yes,yes,yes
499		yes	yes	yes	yes,yes,yes,yes
500		no	no	yes	no,no,yes,yes

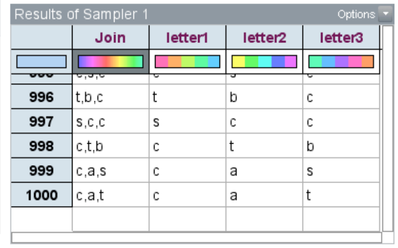


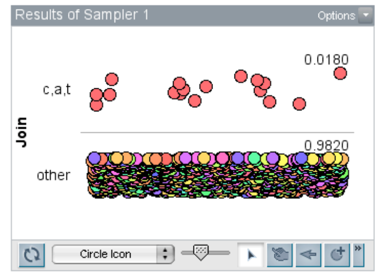
For example, all combinations without rain on only one of the four days are combined into the category "no, yes, yes, yes." Another option is to search for cases with one day with rain, leaving all other events in an "other" category.

The theoretical probability is $(0.75)^4$, or about 32%.

Letters

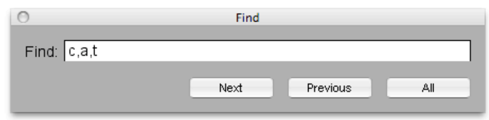


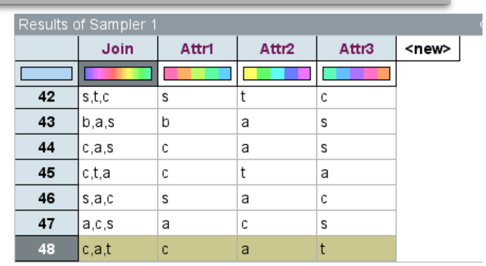


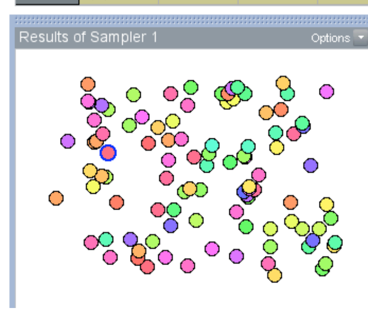


At right is an example of a model students might make. If students run the model and collect 1000 cases, estimates should range from about 1% to 3%. If they run it only 100 times, the results will be more variable. Encourage students to run the model several times at the same sample size to get a sense of the variability of the results for that size sample.

Many students will initially sample "with replacement." In this case "c, a, t" is harder to get, and common rates for 1000 trials will range from 0.4% to 1.5%.







Making a plot like the one here with the outcome "c, a, t" and the "other" category can be tricky. First, locate an instance of "c, a, t" after you run the simulation. One way to do this is to scan the results table until you find an instance. Clicking that row will highlight that case in the plot. (They may need to first click the light blue rectangle in the upper left of the results table to decolorize the data points first, so that they can see highlighting.). Now, if they remember which point that is, they can click the *Join* attribute in the results table, then click and drag the desired point to separate "c, a, t" from the other *Join* values.

A quicker way to do this is to select the *Join* attribute to color the cases. Then go to the **Data** menu and choose **Find Case.** Typing `c,a,t` (quotation marks are not necessary) into the dialog box will highlight all of the cases. The colors can make it difficult to spot the highlighted cases in the plot. Look for a thick dark blue outline.

Click one of the highlighted cases in the plot and drag it up. This will bring that case and all other c, a, t cases into their own group. If there are no highlighted cases, continue sampling until you get at least one. Now each time you click the **RUN** button, the results will be displayed in this plot.

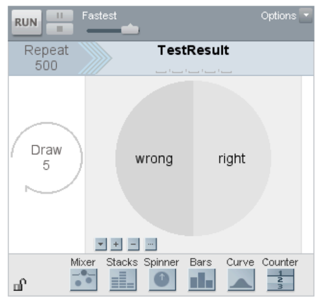
The theoretical probability is $\frac{1}{60}$, or 1.7%.

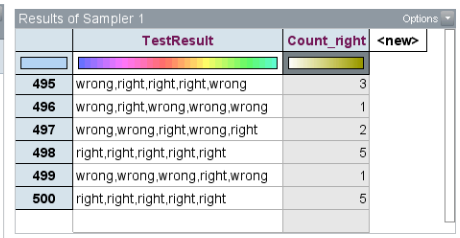
True or False

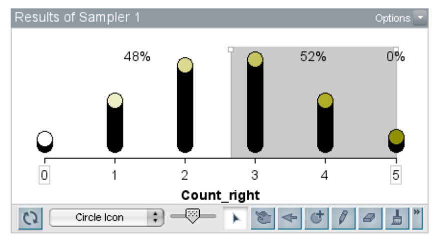
This is a particularly challenging problem to model. Many students will want to build a test with correct answers and then generate a set of random answers, which is what happens in the real-world event. For example, they might build a sampler that produces random answers (either "T" or "F") and then imagine a particular scoring key (T, F, T, T, T).

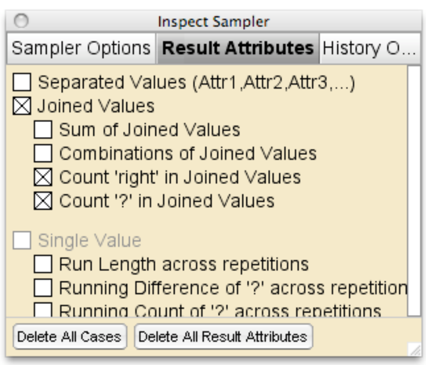
The problem with this approach is how to score all the results. Writing a formula that will do that is challenging.

Rather than focusing on "True" and "False," the sampler at right is based on the fact the Mike has a 50% chance of being right on each question, which eliminates the need for a scoring key.









To get the total number of correct answers, this model uses the option "Count '?' in Joined Values," as shown at left. Choose **Result Attributes** from the **Sampler Options** menu. In the dialog box, enter "`right`," which then counts the number of "right" values in each row.

Another way to do this would be to put zero (0) in the spinner instead of "wrong," 1 instead of "right," and sum them in the results table. The sum is the score on the test.

Estimates based on this model should vary between 40% and 60%, depending on sample size. Students can determine that the theoretical probability is 50% by the following logic. The probability of getting zero correct is equal to that of getting five correct, and the same is true for the probabilities of getting one or four correct, and for getting two or three correct. Thus the distribution is symmetric around its center, resulting in a 50% chance of getting three or more correct.

Wrap-Up (10 minutes)

Depending on how you use this activity, you might have student pairs present their models and answers to the class, followed by a class discussion that addresses some of the issues described in the preceding notes.

In addition to discussing that these simulations provide ranges for probability estimates rather than exact values, students should also understand these ranges can be narrower, or more accurate, with larger sample sizes.