Exploring Sampling—With or Without Replacement

You will need
• **WithWithout.ftm**

In this activity you'll use Fathom's sampling capabilities to see if there is a difference between sampling with replacement and sampling without replacement.
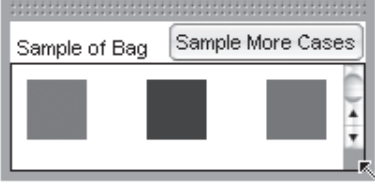
GENERATE DATA

1. Open the Fathom document **WithWithout.ftm.** In this document there is a collection called Bag, with six cases. Each case (colored square) represents an equal-sized slip of paper.

2. Add a new attribute, *Name,* to the case table. Enter six different names. Make yours one of them.

3. Select the Bag collection and choose **Collection | Sample Cases.** A new collection appears named Sample of Bag. Make a case table for that collection.

Q1 How many names are in the case table for the sample collection? Does Fathom appear to be sampling with or without replacement? Explain.

INVESTIGATE

Bag Simulation

4. Double-click the sample collection to show its inspector. Go to the **Sample** panel and change the number of cases to 5.

You can also click **Sample More Cases** in the inspector.

5. Drag the lower-right corner of the sample collection until you can see the **Sample More Cases** button. Click **Sample More Cases** repeatedly.

Q2 Describe what happens as you sample. Does Fathom appear to be sampling with or without replacement? Explain.

6. In the inspector, change the number of cases to 3. Click **Sample More Cases** again. Count the number of times you have to click **Sample More Cases** until you see the same person appear twice in the same sample of 3.

Q3 How many times did you have to sample? Try the experiment a few more times. Were your results about the same?

Now you will sample *without replacement.*

7. In the inspector, uncheck With replacement. Click **Sample More Cases** repeatedly.
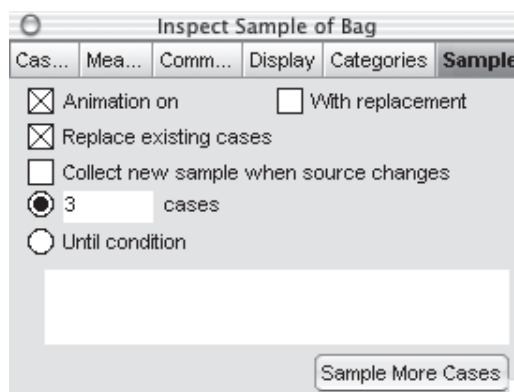

Q4 Will the same name ever appear twice in the same sample of 3? Explain. What do you think would happen if you changed the sample size to 6?

8. Check your prediction by changing the number of cases to 6—the number of names in your collection. Click **Sample More Cases** repeatedly.

Q5 What happens in the case table as you sample?

9. Now, increase the number of cases to 10. Watch the inspector closely when you click **Sample More Cases.**

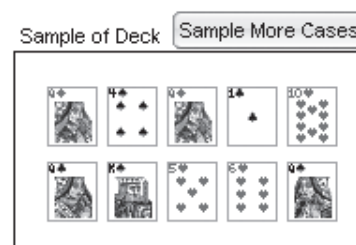**Q6** Describe what happens as you sample.

Deck Simulation

Now you'll use what you've learned about sampling to simulate picking cards from a deck of cards.

10. Scroll down in the document until you see the Deck collection. Double-click this collection to show the inspector.

Q7 How many cases are in the collection? What are the attributes?

11. Select the collection and choose **Collection | Sample Cases.** A new collection appears, called Sample of Deck. Open the collection by dragging the lower-right corner.

Q8 Click **Sample More Cases** several times. Describe what happens.

Now you'll make a simulation to help you determine the chance of getting a pair when you draw two cards. A pair occurs when the *Number* values of the two cases are the same.
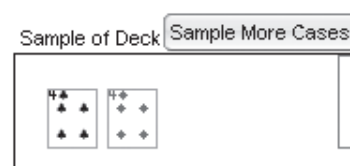
12. Change the number sampled from 10 (the default) to 2. Uncheck Animation on. When you click **Sample More Cases** (whether in the inspector or in the collection), you now get two cards.

Three functions that might help are first(), last(), and uniqueValues(). They're all in the Statistical section in the Functions list.

13. To keep track of the number of pairs, you can record the results in a *measure.* Go to the **Measures** panel in the inspector. Create a new measure, *Pair,* with a formula that will be true if the two cards in the collection are a pair and false if they are not.
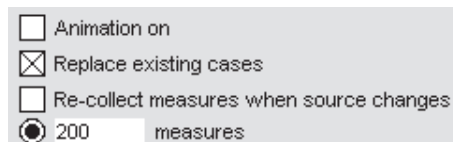
Q9 What is your formula?

Now you'll collect some different samples—different sets of two cards.

14. With the Sample of Deck collection selected, choose **Collection | Collect Measures.** A new collection appears, called Measures from Sample of Deck. Show this collection's inspector and go to the **Cases** panel.

Q10 What attributes does this collection have? How many cases does it have?

15. Go to the **Collect Measures** panel and enter these settings. Click **Collect More Measures.** Be patient. You are collecting 200 samples of two cards.

Q11 Make a bar chart of *Pair.* How many of your 200 samples were pairs? How do you know?

Q12 In the entire class, how many pairs were there? Out of how many samples?

16. Show the inspector for the Sample of Deck collection—not the measures collection. On the **Sample** panel, uncheck With replacement. You will now sample *without replacement.*

17. Show the inspector for the Measures from Sample of Deck collection, and go to the **Collect Measures** panel. Make sure Replace existing cases is checked so that you'll get 200 new samples. Click **Collect Measures.** You'll (slowly) get the data on 200 samples.

Q13 How many of those samples were pairs? In the entire class, how many pairs were there? Out of how many samples?
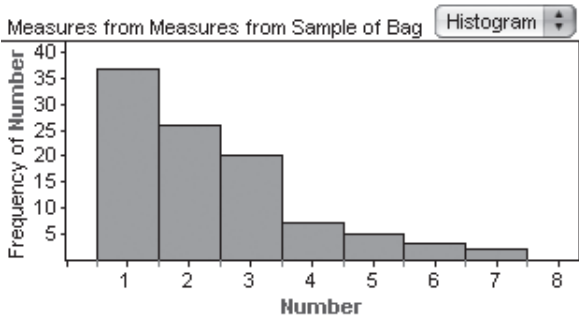
Q14 What's the empirical probability of getting a pair when drawing from a 52-card deck *with replacement*? What's the probability *without replacement*? Explain, in words, why one probability is greater than the other.

EXPLORE MORE

1. Find the theoretical probabilities for drawing a pair from a 52-card deck with replacement and without replacement.

2. Modify the Deck simulation to compute the probability of drawing two cards and having one be an ace and the other a king, queen, jack, or ten (getting a blackjack).

The uniqueValues function will help.

3. Modify the Deck simulation to compute the probability of getting only one pair in a five-card sample (a pair in poker).

The uniqueValues function will help. The histogram also shows a hint.

4. Design a Bag simulation that counts the number of samples it takes to get the same name to appear twice in a sample of 3. Then generate a distribution of the number of samples it takes to get the same name to appear twice in a sample of 3 and find its center and spread.

5. Open the document **Westvaco.ftm.** The Westvaco Corporation, which makes paper products, decided to downsize. They laid off several members of their engineering department and Bob Martin was one of those who lost their job. He claimed he had been laid off because of his age. Westvaco's management went through five rounds of planning for a reduction in workforce. In the first two rounds, 20 people were fired, 16 of whom were 50 or older.

Using what you have learned in this activity, design a simulation to calculate the probability of firing 20 people and getting 16 or more people who are 50 or older.

Exploring Sampling— With or Without Replacement

Objectives

- Using simulation to calculate empirical probabilities
- Creating a model of a chance process
- Using a sampling distribution to find the probability of a given compound event by sampling with replacement and without it

Activity Time: 30–45 minutes

Setting: Paired/Individual Activity or Whole-Class Presentation (use **WithWithout.ftm** for either)

Optional Documents: Westvaco.ftm (Explore More 5) and **Westvaco2.ftm** (Explore More 5 solution)

Statistics Prerequisites

- Definition of probability
- Familiarity with the definition of independent events, $P(A \mid B) = P(A)$
- Familiarity with the conditional $P(A \mid B)$
- Familiarity with sampling

Statistics Skills

- Multiplication rule
- Working with the definition of conditional probability
- Working with the definition of independence
- Calculating probabilities with graphs and two-way tables
- Sampling with and without replacement and how that affects probabilities and independence
- *Optional:* Collecting measures of measures (Explore More 4)

AP Course Topic Outline: Part III A (1–3, 5), B (1), D (6)

Fathom Prerequisites: Students should be able to make case tables and bar charts, create attributes, and use the formula editor.

Fathom Skills: Students sample with and without replacement, change the properties of a sample, use an inspector to create a measure, and create a collection of measures.

General Notes: Sampling is a central idea in statistics. This activity explores the differences when sampling with or without replacement and then uses sampling to investigate probability. Using Fathom allows students to quickly take many samples of different types and easily build sampling distributions.

Procedure: The Bag Simulation section simulates putting 6 names on equal-sized slips of paper—one name on each slip—then putting the slips of paper in a bag, shaking it up, then drawing out a sample from the bag, one name at a time. In steps 3–6, the slip of paper would be returned to the bag and the slips well mixed before choosing another name, whereas in steps 7–9, the slips are not returned to the bag. You might want to demonstrate each of these with a real bag and six pieces of paper—just to reinforce the method. You can also suggest that students resize the sample collection so that they can see the whole sample. Some students will find it easier and faster to spot repeats by looking for the repeating colors rather than names.

In the Deck Simulation section, students need to find out how many pairs were drawn in all the samples from the entire class. You will need to facilitate this data collection in any way that is practical. This is done to increase the number of samples and to therefore decrease the sampling error of the probability.
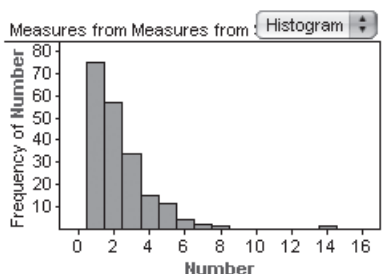
GENERATE DATA

Q1 There are 10 names by default, so Fathom is sampling with replacement. This is evident because there are only 6 names in the bag, but 10 in the sample. Some must be repeating.

INVESTIGATE

Q2 Each time **Sample More Cases** is clicked, a new set of 5 names appears in the case table. With only 6 names to choose from, most likely some will repeat.

Q3 Here is one set of results for 100 trials where the sampling was continued until the first sample was

drawn that had a repeated name. The mean for these 100 trials was 2.33 but as you can see below, 2 trials took 14 samples before a name repeated in a sample of 3. (See Explore More 4 for how this simulation was set up.)

Q4–Q5 The same name will never appear twice in the sample. If the sample size is changed to 6, all 6 names will come up every time (but the order will change).

Q6 When the **Sample More Cases** button is clicked, the inspector changes the number of cases to 6. This is the maximum number of cases for sampling without replacement.

Q7 There are 52 cases in the collection. The attributes are *Suit, Number, Name,* and *CardID.*

Q8 Each time **Sample More Cases** is clicked, a new set of 10 cards is drawn from the deck. Cards may repeat.

Q9 The activity doesn't tell students how to make the formula for *Pair*—the Boolean formula that assesses whether the two cards in the sample are a pair. Any formula that accomplishes the task is acceptable. Here are two possibilities:

first(Number)=last(Number)
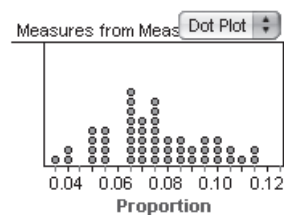
uniqueValues(Number)=1

Students are more likely to come up with the first formula, which works nicely. If they go on to the Explore More section, you should introduce them to the second formula. For the question on getting a pair in a five-card sample, the formula is uniqueValues(Number)=4. (The equals sign—as opposed to ≤—is correct. If

uniqueValues(Number)=3, for example, that means the sample has two pair or three of a kind.)

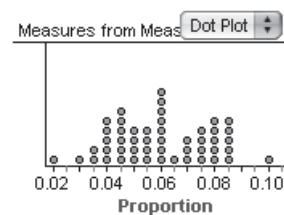**Q10** There should be 5 cases if the sample size is set to its default of 5. This collection has attributes *CardSize* and *Pair.*

Q11 The bar chart in the activity has 18 pairs. You can put the cursor over the bar and look in the lower-left corner of the Fathom window.

Q12 Make sure that you have recorded class values before moving on. The next part of the activity will wipe out the previous data.

Q14 Make sure students know that empirical probability is the same as experimental probability. Use the class results to calculate these answers. Sample answer: The probability would be 18/200, or 0.09, for sampling with replacement and 13/200, or 0.065, for sampling without replacement.

Here are 50 runs of 200 samples of two cards. The distribution of proportions from sampling with replacement (top) is centered a bit more to the right (at 0.75) than the distribution of the proportions from sampling without replacement (centered at 0.6), as it should be.

With Replacement

Without Replacement

EXPLORE MORE

1. With replacement: 4/52; without replacement: 3/51.

2. Note that the value of *Number* for an ace is 1. You need to test the first and last (second) cards in the sample to see whether they are aces or ten cards. (Face cards have a value of 10.) One way to do this is to create new measures in the original collection with these formulas:

   HasAce=(first(Number)=1) or (last(Number)=1)
   HasTen=(first(Number)>9) or (last(Number)>9)
   Blackjack=(HasAce) and (HasTen)

   The theoretical probability for getting blackjack without replacement is
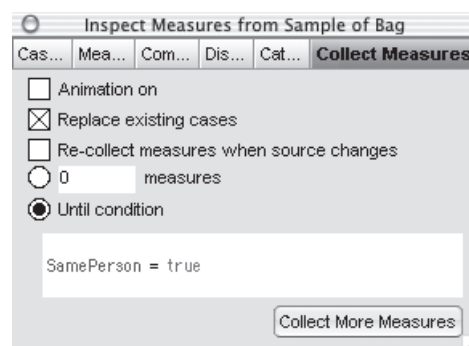
   $$\frac{4}{52} \cdot \frac{16}{51} + \frac{16}{52} \cdot \frac{4}{51} \approx 0.048$$

   With replacement, the probability is

   $$\frac{4}{52} \cdot \frac{16}{52} + \frac{16}{52} \cdot \frac{4}{52} \approx 0.047$$

3. Change the number of cases for the sample collection to 5. To get a pair, use the formula uniqueValues(Number)=4. This will return the cases where there are only four unique numbers, indicating that two cards must match.

4. In the sample collection, define the measure *SamePerson* with the formula uniqueValues(Name)<3. In the measures collection, create a measure *Number* defined by the formula count(). On the **Collect Measures** panel, change the panel to sample until SamePerson=true as shown. Then select the measures collection, and choose **Collection | Collect Measures.** The new collection, Measures from Measures from Sample of Bag, counts the number of samples that were taken before a repeating name popped up. See the histogram and mean in the answer to Q3 for possible means.

5. See the document **Westvaco2.ftm** for one solution.