Generating Random Data—Cat Factory

Social scientists are increasingly using simulation methods to help them understand the social processes they study. One method they use, called Monte Carlo simulation, is to generate many samples from a specified population or model. Then they can study the patterns that emerge from these samples. In this activity you will learn how to set up a defined model (population) using TinkerPlots.

Exploring a Pre-Built Sampler

Before building your own data factory, you will explore a pre-built sampler that simulates data about cats.

Open the file *Simulating Data.tp*. This is a built-in file that comes with TinkerPlots™. From the File menu select Open Sample Document > Tutorials > Simulating Data.tp.

Take a few moments to understand what is in this document. The sampler at the top of the screen can be used to generate a data set of cats with four attributes: *Gender, Name, Length,* and *EyeColor*. The case table and plot at the bottom of the screen show 500 cases—in this case, cats—that were generated by the sampler.

Now, you will create a new data set using the built-in sampler.

- At the top left corner of the sampler, click the Repeat value, *500*, and change it to *5* to generate data for five cats.
- Now, click the Run button at the top left corner of the sampler.
- Watch as the sampler generates data for five cats. As the data for a cat is completed, a new case appears in the results table, and a new case icon appears on the plot.

1. Describe the data generating process (i.e., how are the simulated data being produced; describe what is happening).

2. What characteristics are dependent on the cat's sex?

3. What characteristics are independent of the cat's sex?

4. Examine the plot. What characteristics are being plotted?

- Click a case icon in the graph. Notice that the cat is also highlighted in the results table.
- Now click on a cat in the results table (click on the cat's row number). The cat's case icon will also be highlighted.

Building a Data Factory

Now you will build your own cat factory. You may want to leave the sample document open for reference.

- Open a new document in TinkerPlots™ by selecting File > New.
- Drag a new Sampler from the object toolbar into your blank document.
- Click and drag the samplers lower right-hand corner to make it larger.

At the bottom of the sampler, you will see six sampling devices that can be used to generate attributes.

Mixers and **Stacks** draw from a set of discrete elements. For example, the Name attribute in the cat factory was chosen from a mixer. If you have many repeats of the same value, such as choosing from a set of 30 boys and 45 girls, stacks are a better option than a mixer.

Spinners and **Bars** draw from discrete elements, but can have different probabilities for each value. The Gender and EyeColor attributes in the cat factors were determined by spinners, and the Length attribute was determined by bars.

Curves draw from a continuous range of numerical values, which can have different probabilities.

Counters select values systematically, rather than randomly.

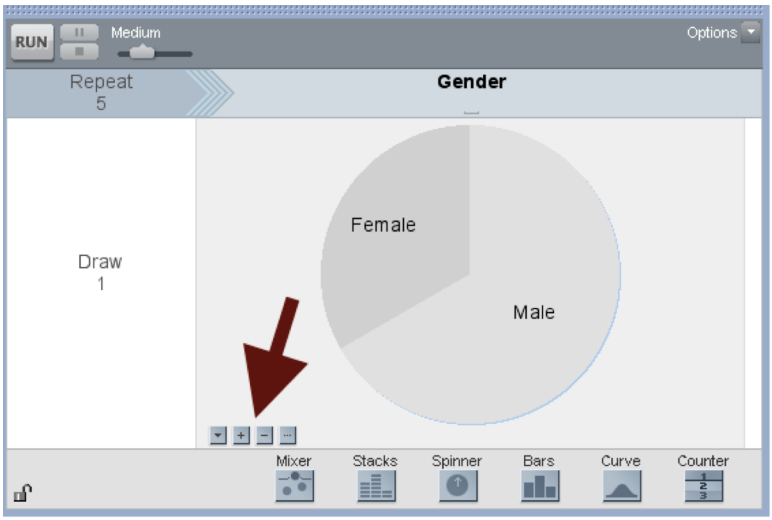
To model a cat factory, you will need to model both categorical (discrete) attributes, such as gender, and numerical (continuous) attributes, such as length.

Modeling Cat Genders: Spinner

The first attribute is *Gender*. Gender can be modeled best using a mixer or a spinner. Although the default device given in a new sampler is a mixer, we are going to use a spinner.

- Change the mixer to a spinner by dragging a spinner from the sampler's bottom toolbar into the sampler, and releasing it above the pink dot that appears in the center of the current mixer. (Pink dots show places where you can drop the new device. A black rectangle also highlights where you can drop the new device.)
- Select the text *Attr1* above the spinner and relabel it *Gender*.
- Click the *a* text in the spinner and change it to *Male*. Click the *b* text and change it to *Female*.
- Change the Draw value from *2* to *1*. The draw value will indicate the number of sampling devices included in your sampler. Currently there is only one sampling device, a spinner.

Note the four buttons in the lower left corner of the spinner device.



Clicking the first icon shows the Device options menu; clicking the + and – buttons adds and subtracts values to the device; and clicking the ... button allows you to enter a range of values into the device. Click each button to see the result.

Currently one of the genders is more likely than the other. To make female cats as likely as male cats, you'll need to change the position of the divider in the spinner.

- Click the Device options menu and choose Show Percent.
- Change the percentages so that each is *50%*.

Modeling Cat Names: Mixer

The next attribute we want to model is *Name*. To model the cat names we will use a mixer. Furthermore, because male and female cats tend to have different names, you will use two mixers, one for male cat names and one for female cat names.

- Drag a mixer from the lower sampler toolbar into the sampler, and drop it on the pink dot to the right of the Gender device. (A black rectangle will highlight when you're in a position to drop the mixer.)
- Change the attribute name from *Attr2* to *Name*.
- Drag a second mixer into the sampler but do not release it yet. You will see four pink dots that represent locations where you can place the mixer. If you drop the mixer on the pink dot in the Gender spinner or in the first mixer, it will replace that device. Because you want this to be a second Name mixer, drop it on the pink dot attached to the Gender mixer, and directly below the first Name mixer. Notice the Male and Female labels on the lines connecting the Gender device to the Name devices.

Now we need to add the potential names into each mixer. First we will add the male names. The 10 male names we will add into the mixer are:

Charlie	Shadow	Spot	Jack	Max
Smokey	Oliver	Buddy	Simbar	Tiger

- Click the + (add element) button below the Male mixer. This will add an element called *a* into the sampling device.
- Change the name of the element from *a* to *Charlie*.
- Add nine more elements to the mixer, changing their names to each of the nine other male cat names. When you have finished, there should be 10 male cat names in the Male mixer.

Next we will add the 10 potential female cat names. Rather than have you enter the data manually, this time, you will use **Copy and Paste** to input the data.

- Open the file f*emale-cat-names.tp*.
- Highlight the Female attribute by clicking on the column header.
- Copy the list of names by selecting Edit > Copy Attribute (or use appropriate keyboard shortcuts).
- Go back to your cat factory document. Select the female name mixer by clicking on it, and paste the values by choosing Edit > Paste Cases (or use appropriate keyboard shortcuts).

Modeling Cat Length: Bars

The next attribute we want to generate data for is the cats' *Length*. This is a numeric attribute with varying probability for each value, so you can use a bars or curve device. We will use the bars device in this activity.

- Drag a bars device into the sampler and drop it just to the right of the Male Name mixer.
- Then drag a second bars device and drop it to the right of the Female Name mixer. (Males and females have different distributions of potential lengths, so their lengths will be chosen from different devices.)
- Change the name from *Attr3* to *Length*.

Cats typically have lengths between 10 and 30 inches. There are several ways to specify a range.

- Hold your mouse over the top bars device, and click the + (add element) button. Then change *a* to *10*.
- Click the + (add element) button again. Notice that it automatically adds the next number to the device (e.g., 11). You can continue to click the + (add element) button until you have all the values from 10 to 30, but there is a faster way.
- Click the ... button and enter the range "*10 to 30*."
- Enter the same range in the lower bars device.

Since all the bars are the same height, this indicates that each length, from 10 to 30, is equally likely.

5. Are all these lengths really equally likely? For example, are 30-inch-long cats just as common as 17-inch long cats?

It is more likely that cats have lengths in the middle of this range, and that lengths at the extreme values in this range are less likely. To model this, we will shape the *Length* distribution so it has a bump (is higher) in the middle. (remember, the height indicates the relative probability of each element, in this case, cat length).

- Position your cursor at the top of the far-left bar in the *Length* device for male cats (a plus-sign should appear). Click and drag the cursor over the bars in the shape of the desired distribution.
- Repeat for the Length of female cats. Since female cats tend to be a little shorter than male cats, you should place your bump a little to the left of the bump in the distribution for male cats.

Modeling Cat Eye Color: Spinner

The final attribute we want to generate is *Eye Color*. Because this is a categorical (discrete) attribute, we can use either a spinner or stacks. Here, we will again use a spinner.

- Drag a spinner into the sampler and drop it to the right of the bars device for male cats.
- Drag a second spinner into the sampler and drop it to the right of the bars device for female cats.

Since eye color does not vary for male and female cats, we do not need a separate device for males and females. To join these two devices together:

- Hold your mouse over the upper spinner, and click the Device options menu (down arrow). Choose Merge Device > Merge with Device Below.
- Change *Attr4* to *EyeColor*. (Note that attribute names cannot contain spaces so we use bumpyCase.)
- Click the + (add element) button to add three eye colors to the mixer. Label the values *Yellow*, *Green*, and *Blue*.
- Blue-eyed cats are less common than yellow or green-eyed cats, so make the Blue section smaller than the other two by changing the percentages.

You should now have a cat factory that resembles the one in *Simulating Data.tp*. Click Run to generate five cats with randomized attributes. Notice that a results table automatically appears and is filled in.

6. Save the TinkerPlots™ document and email it to all of your group members so they have a copy.

Practice 1: Population of Students

7. How would you set up a sampler to generate data for 25 students from a population of students, where 40% of the population are freshmen, 30% are sophomores, 15% are juniors, and 15% are seniors? Sketch a picture of the sampler below. Don't forget to indicate both the Draw and Repeat values in your sketch.

8. Open a new TinkerPlots™ document and implement the sampler you just sketched.

9. Generate data for the 25 students by clicking Run. How many of the 25 students generated were seniors? What percentage is that?

10. Generate data for another set of 25 students by again clicking Run. How many of the 25 students generated were seniors? What percentage is that?

11. If you run the simulation many times, will there always be exactly 15% of the generated data that are seniors? Explain.

12. Save the TinkerPlots™ document and email it to all of your group members so they have a copy.

Practice 2: Random Band Members

13. Consider the following eight students: *John, Paul, Ringo, George, Mick, Keith, Charlie,* and *Ronnie*. How would you set up a sampler to randomly choose three of them to be in a rock band? Sketch a picture of the sampler below. Don't forget to indicate both the Draw and Repeat values in your sketch.

14. Open a new TinkerPlots™ document and implement the sampler you just sketched.

15. Generate data for the members of the rock band by clicking Run. Carry out the simulation several times. Do you ever get the same person multiple times in the same band? Why does this happen?

16. Try to change the sampler so that a band member is never generated more than once in each trial of the simulation.

17. Add another sampling device to your model so that after a band member is selected, that person is randomly assigned one of the following four instruments: *kazoo*, *bass*, *cowbell*, and *turntable*. Sketch a picture of the entire sampler below.

18. Generate 25 simulated bands by clicking the Run button 25 times. How many of the 25 simulated bands included a cowbell player?

19. Generate another 25 simulated bands. How many of the 25 simulated bands included a cowbell player named Ronnie?

20. Save the TinkerPlots™ document and email it to all of your group members so they have a copy.