Introduction to Monte Carlo Simulation

Carsey and Harden[1] define Monte Carlo simulation as,

> any computational algorithm that randomly generates multiple samples of data from a defined population based on an assumed data generating process (DGP). The DGP is the mechanism that characterizes the population from which simulated samples of data are drawn. Then the researcher explores patterns that emerge across those simulated samples.

In this activity you will learn the process of carrying out a Monte Carlo simulation and how to do so using TinkerPlots™.

Model—Simulate—Evaluate

Looking back at the definition of a Mote Carlo simulation above, the process encompasses (1) defining a population or model, (2) randomly generating several samples of data from the population or model, and (3) exploring the patterns that emerge across the simulated samples. In simpler terms, (1) model, (2) simulate, and (3) evaluate.

[1] Carsey, T. M., & Harden, J. J. (2014). *Monte Carlo simulation and resampling methods for social science.* Thousand Oaks, CA: Sage.

In the previous course activity, you created several models using TinkerPlots™ and used them to randomly generate data. The key to Monte Carlo simulation is to generate many, many randomly generated samples. The catch is that we need to collect some information from each of these samples so that we can examine this information across the many samples. The information we collect is often a quantifiable summarization of the sample, for example the mean value, a count, or a proportion. The summary we choose is based on our research question.

Monte Carlo Simulation 1: Coin Flips

In the first Monte Carlo simulation you will be exploring the following questions:

If we flip a "fair" coin 10 times, how many heads would we expect? How much variation in the results would we expect if we did this many times?

Questions 1–3 are asking for your intuitions. You do not have to calculate exact values. We will explore these questions in more detail later in this activity.

1. Imagine that you flip a fair coin ten times and count the number of heads. How many heads would we expect to see on average? Why?

2. Now, imagine repeating this process 100 time: flipping that fair coin ten times, and counting the number of heads. Would you expect to see the same result in all 100 trials?

3. How variable would the results be? What do you think the smallest and largest number of heads would be? What do you think the range would be for most results?

Modeling and Simulating

To save time and to gather data quickly, you will use TinkerPlots™ to model tossing a coin 10 times.

4. How would you set up a sampler to toss a fair coin 10 times? Sketch a picture of the sampler below. Don't forget to indicate both the Draw and Repeat values in your sketch.

- Open a new TinkerPlots™ document and implement the sampler you just sketched.
- After you have set up the model, click the Run button.
- A *case table* displaying the 10 outcomes for the "coin flips" should have been produced.
- Plot the 10 outcomes. Fully separate the cases and vertically stack them.
- With the plot highlighted, click the Case Count (N) icon in the upper toolbar. This should display counts of the number of heads and tails in the plot.

5. Record the **number of heads** from your randomly generated data below.

In a simulation, each time the model is used to produce a sample of data, it is referred to as a **trial**. A trial can consist of one or many outcomes depending on the simulation. In this simulation, the trial consisted of 10 outcomes (flips). In TinkerPlots™, the summary, or how we quantify the sample is referred to as the trial's **result**. In this simulation, the trial result would be the number of heads. In order to study any patterns that might emerge, we need to generate many trials and record the result from each of them.

6. Re-click the Run button in the sampler to generate another random sample of data. Since the plot and the case counts are linked to the outcomes from the sampler, these should update automatically. Record the **number of heads** from this new sample below.

7. Generate 23 more samples. For each sample generated, record the number of heads below.

Evaluating the Results from Many Trials

At this point, we have completed two of the three parts of the Monte Carlo simulation process, namely (1) model and (2) simulate. In order to study any patterns in the trials' results, we need to plot the results from the 25 samples of data you generated.

- Open a new TinkerPlots™ document.
- Drag a new Case Table from the object toolbar into your blank document.
- Click on <new> to change the attribute name. Rename this attribute *Results*.
- Enter the results from the 25 trials into the results column in the case table.
- Plot the 25 results. Fully separate the case icons in the plot and vertically stack the cases.

8. Sketch the plot below.

9. Based on the plot of the simulation results, what was a typical number of heads from 10 flips? Explain how you decided this from the plot.

10. Based on the plot of the simulation results, how variable are the results? What are the smallest and largest number of heads that you observed? What is the range of values where most of results lie?

11. Based on the plot of the simulation results, would two heads out of 10 flips be a likely or unlikely result? Explain.

12. Based on the plot of the simulation results, would seven heads out of 10 flips be a likely or unlikely result? Explain.

13. Save the two TinkerPlots™ documents and email them to all of your group members so they have a copy.

Monte Carlo Simulation 2: Generating a Sample of Students

In the previous course activity, 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.

In this Monte Carlo simulation you will be exploring the following questions:

> In a class of 25 randomly selected students, how many juniors would we expect? How much variation in the the number of juniors would we expect if we generated many samples?

Questions 14 and 15 are asking for your intuitions. You do not have to calculate exact values. We will explore these questions in more detail later in this activity.

14. Imagine generating 100 random samples of 25 students from the defined population.What do you think the typical number of juniors in a class of 25 would be? Explain your reasoning.

15. How variable would the results be? What do you think the smallest and largest number of juniors would be? What do you think the range would be for most classes?

Modeling and Simulating

- Open the saved TinkerPlots™ document from the previous activity where you set up this model. If you didn't save the TinkerPlots™ sampler fro the previous activity, re-create the sampler.
- After you have set up the model, click the Run button.
- A *case table* displaying the 25 outcomes for the first trial.
- Plot the 25 outcomes. Fully separate the cases and vertically stack them.
- With the plot highlighted, click the Case Count (N) icon in the upper toolbar.

16. Record the **number of juniors** from your randomly generated data below.

17. Generate 24 more samples. For each sample generated, record the number of juniors below.

Evaluating the Results from Many Trials

- Open a new TinkerPlots™ document.
- Enter the 25 results into a Case Table.
- Plot the 25 results. Fully separate the case icons in the plot and vertically stack the cases.

18. Sketch the plot below.

19. Based on the plot of the simulation results, what was a typical number of juniors in a class of 25 students? Explain how you decided this from the plot.

20. Based on the plot of the simulation results, how variable are the results? What are the smallest and largest number of juniors that you observed? What is the range of values where most of results lie?

21. Based on the plot of the simulation results, would 10 juniors in a class be a likely or unlikely result? Explain.

22. Save all TinkerPlots™ documents and email them to all of your group members so they have a copy.

Monte Carlo Simulation 3: Generating Cat Lengths

In the previous course activity, you set up a sampler to generate data for female cat lengths. In this Monte Carlo simulation you will be exploring the following questions:

> What is the average cat length for a sample of 40 female cats? How much variation in the average length would we expect if we generated many samples of females?

Question 23 is asking for your intuition. You do not have to calculate exact values. We will explore these questions in more detail later in this activity.

23. Imagine generating 100 random samples of 40 female cats lengths from the defined population (the population was defined in the last course activity) and computing the average for each of these samples. How variable would the averages be? What do you think the smallest and largest averages would be? What do you think the range would be for most averages

Modeling and Simulating

Although you previously set up a sampler to generate female cat lengths in the previous activity, it isn't as useful for this simulation since it is connected to other samplers. It will be better to re-create the female cat lengths sampler in a new TinkerPlots™ document. (Plus, it is good practice.)

- Open a new TinkerPlots™ document.
- Set up a sampler to model female cat lengths. If you've forgotten how to do this, look at the instructions in the previous activity. You will need to modify the sampler so it generates data for 40 cats.
- After you have set up the model, click the Run button.
- A *case table* displaying the 40 outcomes for the first trial.
- Plot the 40 outcomes. Fully separate the cases and vertically stack them.
- With the plot highlighted, click the Averages (Mean) icon (triangle) in the upper toolbar. Then click on the Averages Options icon (down-facing triangle) in the upper toolbar, and select Show Numeric Value(s). This should display the value of the mean into the plot.

24. Record the **mean female cat length** for the sample.

25. Generate 24 more samples. For each sample generated, record the mean female cat length below.

Evaluating the Results from Many Trials

- Open a new TinkerPlots™ document.
- Enter the 25 results into a Case Table.
- Plot the 25 results. Fully separate the case icons in the plot and vertically stack the cases.

26. Sketch the plot below.

27. Based on the plot of the simulation results, what was a typical average female cat length? Explain how you decided this from the plot.

28. Based on the plot of the simulation results, how variable are the results? What are the smallest and largest mean values that you observed? What is the range of values where most of results lie?

29. Based on the plot of the simulation results, would an average of 17 inches be a likely or unlikely result? Explain.

30. Save any TinkerPlots™ documents and email them to all of your group members so they have a copy.