# Statistical Thinking

4.3

## A Simulation Approach to Modeling Uncertainty

Catalysts for Change

C ATALYSTS FOR CHANGE

# Statistical Thinking

## A Simulation Approach to Modeling Uncertainty

C ATALYST PRESS

This book is a modified version of:

Zieffler, A., & Catalysts for Change. (2021). *Statistical Thinking: A simulation approach to uncertainty* (4.3th ed.). Minneapolis, MN: Catalyst Press.



PUBLISHED BY CATALYST PRESS



The work to create the material appearing in the book was made possible by the National Science Foundation (DUE–0814433).

Printed in the United States of America

ISBN 978-0615691305

Catalyst Press Minneapolis,
MN 55455

http://catalystsumn.blogspot.com

*Fourth (4.3) printing, August 2021*

# Colophon

The primary font used in the lab manual is Libre Baskerville. Headings are set in Neutra Display, and TinkerPlots™ commands are printed using Linux Libertine Mono.

Graphics in the book (not listed below) were in the public domain:

- Course activity icons from Open Clip Art Library.
- Dog icons used in the Matching Dogs to Owners visualization by Sara Quintana from the Noun Project. (CC License)
- TinkerPlots™ icons and screenshots are used by permission from Key Curriculum Press.

# *Introduction*

The materials in this lab manual and on the accompanying website ( https://RaoVNV.github.io/statistical-thinking/) will introduce you to the seminal ideas underlying the discipline of statistics. In addition, they have been designed with your learning in mind. For example, many of the in-class activities were developed using pedagogical principles, such as small group activities and discussion, that have been shown in research to improve student learning.

**Bring a copy of this lab manual (either physically or electronically) with you to class each day.**

## Course Readings

The course readings (available at https://RaoVNV.github.io/statistical-thinking) should be completed outside of class and are intended to help you learn and extend the ideas, skills, and concepts you learn in the classroom.

## TinkerPlots™ Software

Much of the material presented in the lab manual requires the use of TinkerPlots™. This software can be downloaded (for Mac or PC), and a license can be purchased from http://www.tinkerplots.com/. Mac users who are running an OS later than 10.15 (e.g., Catalina, Big Sur) will need to install the TinkerPlots 3 beta version ( link to this is on the web page).

## Data and Other Resources for the Book

The data sets used in the materials, as well as other materials that accompany the lab manual are available at https://github.com/RaoVNV/statistical-thinking/blob/master/data.zip?raw=true.
Clicking this link will download a ZIP file to your computer. Double-click on the ZIP file to view all the materials.

## Participation in the Learning Process

The lab manual, instructors, and teaching assistants are all resources that are at your disposal to help you learn the material. In the end, however, you will have to do all of the hard work associated with actually learning that material. To successfully navigate this process, it is vital that you be an active participant in the learning process. Coming to class, participating in the activities and discussions, reading, completing the assignments, and asking questions are essential to successful learning.

Learning anything new takes time and effort and this is especially true of learning statistics, as you are not just learning a set of methods, but rather a disciplined way of thinking about the world. Changing your habits of mind will take continual practice. It will also take a great deal of patience and persistence.

As you engage in and use the skills, concepts and ideas introduced in the material, you will find yourself thinking about data and evidence in a different way. This may lead you to make different decisions or choices. But, even if this course does not change your world overnight, you will at the very least be able to critically think about inferences and conclusions drawn from data.

# *Building a Sampler*

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 *pet-factory.tp3*. From the File menu in TinkerPlots™, select Open and navigate to and select the *pet-factory.tp3* data.

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 pets with three attributes: *AnimalType*, *Name*, and *EyeColor*.
- ✓ The case table, to the right of the sampler, shows 500 cases—in this case, pets—that were generated by the sampler.
- ✓ There is also a plot that is showing the number of pets of each type for the 500 pets sampled.

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, currently set at *500*, and change it to *5* to generate data for five pets.
- Now, click the Run button at the top left corner of the sampler.
- Watch as the sampler generates data for five pets. As the data for a pet 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. Is the name for the pet dependent on the animal type? Or independent of the animal type? Explain.

3. Examine the plot. What characteristic(s) are being plotted?

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

## Building a Pet Factory

Now you will build your own pet 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** draw from a set of discrete elements. For example, the *Name* attribute in the pet factory was chosen from a mixer.

**Stacks** draw from a set of discrete elements. For example, the *EyeColor* attribute in the pet factory was chosen from a stacks. If you have many repeats of the same value, such as choosing from a set of 30 dogs and 45 cats, stacks are a better option than a mixer. The height of the stacks indicates how likely each element is. For example, in the pet factory green-eyed pets are three times as likely as yellow-eyed pets.

**Spinners** and **Bars** also draw from discrete elements. Each element can have different probabilities that we set via proportion or percentages. The *AnimalType* attribute in the pet factory were determined by spinners.

**Curves** draw from a continuous range of numerical values, which can have different probabilities. (We will not use this sampling device in EPsy 3264.)

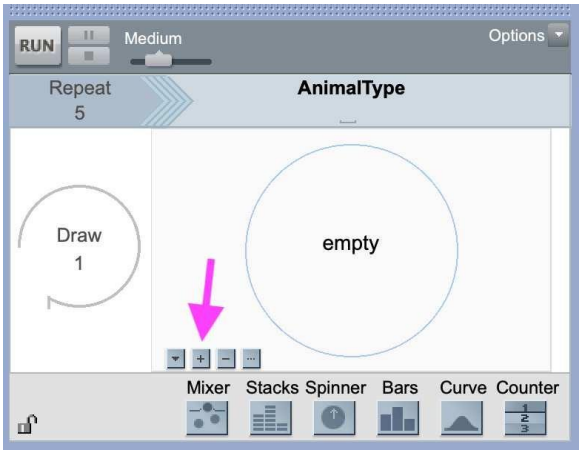
**Counters** select values systematically, rather than randomly. (This is another device that we will not use much in EPsy 3264.)

The values in these sampling devices can be either categorical (e.g., names) or quantitative (e.g., numbers). Some devices take only categorical values, others take only quantitative values, and some take either type. All the attributes in your pet factory will be categorical attributes.

## Modeling Animal Type: Spinner

The first attribute is *AnimalType* . Animal type can be modeled using several different devices. 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.)
- 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.
- Select the text *Attr1* above the spinner and relabel it *AnimalType* . (Note that attribute names cannot contain spaces so we use bumpyCase.) At this point, your pet factory should look similar to the figure below.




We need to add elements to the spinner, one for each animal type we will include in the pet factory. Note the four buttons in the lower left corner of the spinner device (where the pink arrow is pointing in the figure).

Clicking the first icon shows the Device options menu; clicking the + and – buttons changes the number of elements in the sampling device; and clicking the ... button allows you to enter a range of values into the device.

- Add two or three other animal types to your spinner. The original pet factory also included *guinea pigs* and *birds*, but feel free to include whatever kind of pets you want!
- Change the names of the elements (the text *a, b, etc.*) to indicate the type of animals you want to include in your pet factory (e.g., *dog*, *cat*).

Notice that all animal types are equally likely. To make some animal types more (or less) likely, you'll need to change the position of the divider in the spinner.

- Click the Device options menu and choose Show Percent.
- Hover over the divider between the different animal types in the spinner (a rotating arrow will appear). Click and drag the divider line to the desired percentage. ( You can also change the percentage by typing over the value of the percentage.)
- Change the percentages so that some animal types are more likely than others.

## Modeling Pet Names: Mixer

The next attribute we want to model is *Name* . To model the pet names we will use a mixer.

- Drag a mixer from the lower sampler toolbar into the sampler, and drop it on the pink dot to the right of the *AnimalType* device. (A black rectangle will highlight when you're in a position to drop the mixer.)
- Change the attribute name from *Attr2* to *Name* .

Now we need to add the potential names into each mixer.

- Click the + (add element) button below the *Name* mixer. This will add an element called *a* into the sampling device.
- Change the name of the element from *a* to *Hypatia*.

- Add nine more elements to the mixer, changing their names to a potential pet name. When you have finished, there should be 10 names in the *Name* mixer.

## Modeling Cat Eye Color: Stacks

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

- Drag a stacks device into the sampler and drop it to the right of the *Names* device.
- Change *Attr3* to *EyeColor*.
- Click the + (add element) button to add three (or more) eye colors to the stacks. Label the values for each eye color. At the very least include blue, green, and yellow.

Change the counts in each of the eye colors to reflect the following: Blue- eyed pets are less common than yellow or green-eyed pets.

- To do this, click the Device options menu and choose Show Counts. Type the count number by editing the value *1* that appears over each stack.

You should now have a pet factory that resembles the one in the *pet-factory* data set. Click Run to generate five pets with randomized attributes. Notice that a results table automatically appears and is filled in.

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

# *Generating Random Data*

In the last activity, you created a TinkerPlots model to be used in Monte Carlo simulation. In this activity you will learn how to enact simulation using TinkerPlots.

## Practice 1: Population of Students

1. How would you set up a sampler to generate data for 25 students from a population of students, where 40% of the population are first years, 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.

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

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

4. 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?

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

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

## Practice 2: Random Band Members

7. Consider the following eight students: *Maria, Nushi, Mohammed, Wei, John, Ana, Michael, Aleksandr*. 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.

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

9. 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?

10. 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.

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

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

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

# *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

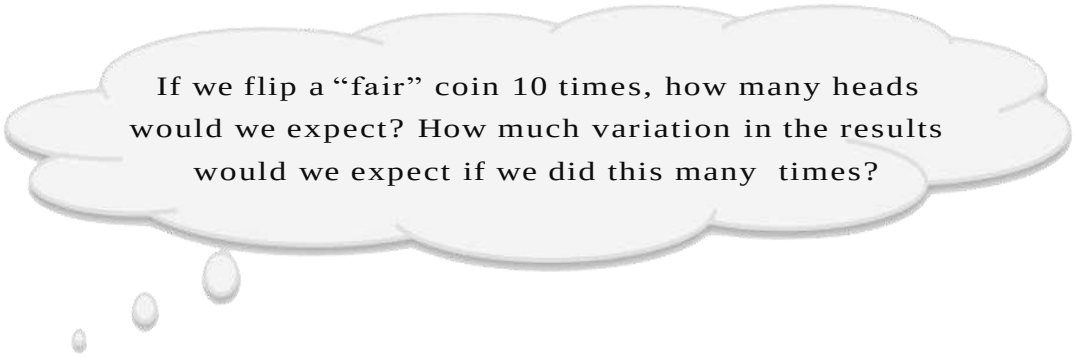
Looking back at the definition of a Mote Carlo simulation above, the process encompasses (1) defining a population or model, (2) randomly generating several samples of data from the population or model, and (3) exploring the patterns that emerge across the simulated samples. In simpler terms, (1) model, (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, called a *statistic*. For example, the mean value, a count, or a proportion are all statistics. The statistic 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:




*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 you expect to see on average? Why?

2. Now, imagine repeating this process 100 times: 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 statistic, 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.

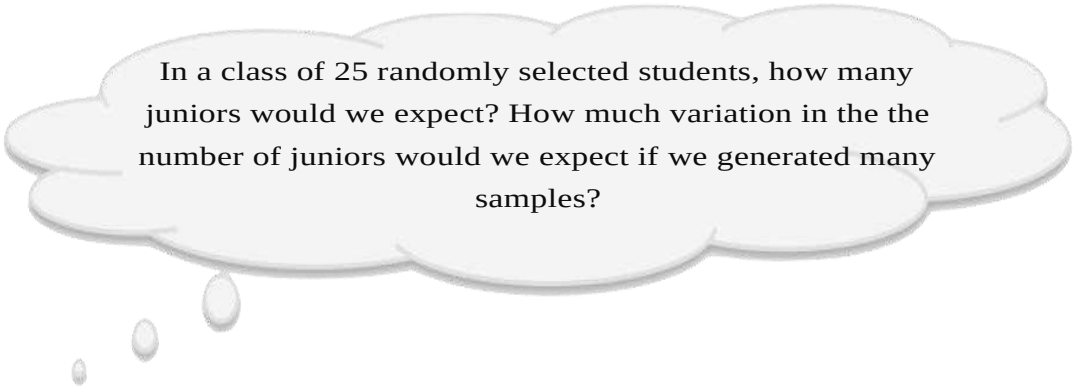
12. Based on the plot of the simulation results, would seven heads out of 10 flips be a likely or unlikely result? Explain.

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:




*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.

# *Automating the Simulation Process*

In previous activities and assignments, you have learned how to set up a model to run a simulation experiment using TinkerPlots™. In these simulations, you ran many trials from which you collected a particular outcome (e.g., the number of heads when flipping a coin 10 times). You also learned how to create a case table to collect the results from each trial into, and how to plot those results.

In this activity, you are going to be introduced to the Collect function in TinkerPlots™. This will automate the collecting of trial results in a simulation. It will also make carrying out several trials easier.

## Modeling Coin Flips

Recall in the previous activity you modeled flipping a coin 10 times. Suppose you wanted to simulate 100 more trials of 10 flips In TinkerPlots™.

- Set up a model to simulate tossing a single coin 10 times.
- After you have set up the model, click the Run button.

## Automating the Collection of Trial Results

Rather than having you record the number of heads that occurred in the 10 flips, we will automate this using TinkerPlots™. The general process for having TinkerPlots™ record and collect the trial results is: (1) plot the outcomes from the trial, and (2) collect the numerical result you are using to summarize the trial.

- Plot the 10 outcomes from the trial. Fully separate the outcomes and stack them vertically.
- Highlight the plot of the trial outcomes and click on the Case Counts (N) button in the upper plot toolbar. This should display a count of the number of heads and tails in the trial.

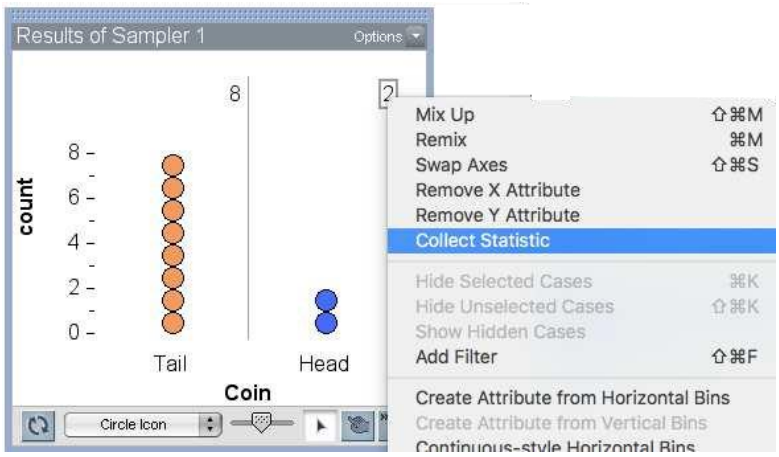
> Note that Case Counts (N) and Case Counts (%) will count the number of cases **within each section of a plot** . If there are not multiple sections (no bin lines), the number of total cases in the plot will be displayed. This is why we need to fully separate the cases when we plot them.

### *Collecting the Results from Many Trials*

You can also use TinkerPlots™ to automatically collect the summarized result from the trial into a case table.

- Use TinkerPlots™ to automatically collect the result from your simulated trial into a case table (see instructions below and figure on next page).

> **Collecting the Results from a Trial**
>
> - Right-click the statistic in your plot.
> - Select Collect Statistic.




It is important that you right-click on the *actual value of the result* in the plot since you want TinkerPlots™ to collect the value. For example, in the plot displayed above, you would right-click on the value *2* to collect the number of heads.

The result is then collected in a new case table. This case table, which is called *History of Results* , has a single row with the collected result, in this case two, displayed in a new attribute. The window next to the Collect button indicates the number of results that were collected, in this case one result was collected. This value can be changed to add the results of additional trials into the case table. In this case, the result collected from each trial is stored in a row of the *History of Results* case table.

- Change the value in the *History of Results* case table to *99* to add the results from an additional 99 trials of the simulation (see figure below).
- Click the Collect button.


*Change the value to 99 in the History of Results table.*

Use the data you collected from the 100 trials of the simulation to answer each of the following questions.

1. Record the result from the 87 $^{th}$ trial.

2. Plot the results from your 100 simulated trials. Don't forget to fully separate the case icons, and vertically stack them.

3. 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.

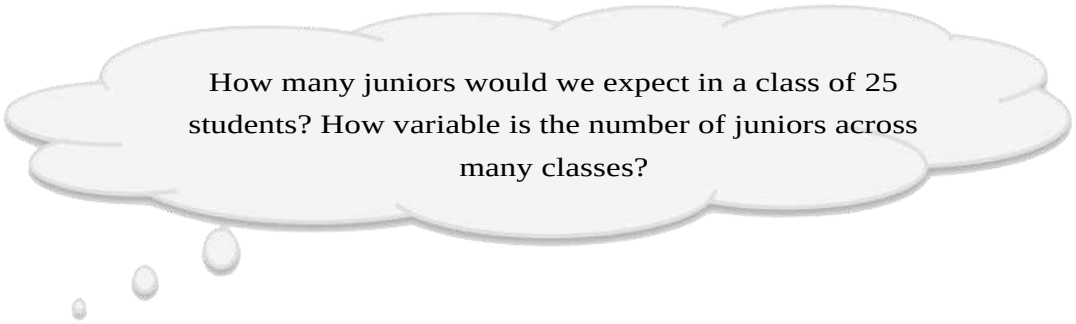
4. Based on the plot of the simulation results, how variable are the results? What are the smallest and largest number of heads that you observed? What is the range of values where most of results lie?

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

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

## Number of Juniors

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.



- Open a new document in TinkerPlots™.
- Set up the model of sampling students.
- Carry out a single trial of the simulation.
- Plot the 25 outcomes from the simulated trial.
- Stack and separate the cases into groups.
- Use Case Counts (N) to summarize the number of cases in each group.
- Collect the number of juniors from the trial into a *History of Results* case table.
- Carry out an additional 99 trials.
- Plot the results from your 100 simulated trials.

Use the plot of the results from your 100 simulated trials to answer each of the following questions.

7. Sketch a plot of the results. Be sure to label the axis.

8. 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.

9. 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?

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

# *Monday Breakups*

Facebook is a social networking website. One piece of data that members of Facebook often report is their relationship status: single, in a relationship, married, it's complicated, etc.

With the help of Lee Byron of Facebook, David McCandless—a London-based author, writer, and designer—examined changes in peoples' relationship status, in particular, breakups. A plot of the results showed that there were repeated peaks on Mondays. Based on this initial examination of data, McCandless speculated that breakups are reported at a higher frequency on Mondays. This is his **research hypothesis** .

To test this research hypothesis, McCandless collected a random sample of 50 breakups reported on Facebook within the last year. Of these sampled breakups, 13 occurred on a Monday.

In this activity, you will be exploring the following research question:

Is 13 (out of 50) breakups reported on Mondays consistent with the model where breakups are equally likely during the week? Or is it more consistent with a model where Mondays have a higher frequency of breakups?

## Discuss the Following Questions

1. What does the observed data ( 13 out of 50 breakups on Monday) suggest about the answer to the research question? Explain.

## 'Equally Likely' Model

Suppose for the moment that the researchers' conjecture is wrong, and breakups *are not* reported on Monday more than any other day. In other words, breakups are reported uniformly throughout the week. This, is a statistical hypothesis. Namely,

> *Statistical Hypothesis:* Breakups are reported at the same frequency/ percentage on each day of the week.

This statistical hypothesis specifies an "equally likely" breakup reporting model for each day of the week. We can use TinkerPlots™ to create this model and generate random outcomes.

**Statistical Hypothesis**

A statistical hypothesis is a statement **specifying a model** that explains variation in a particular outcome.

Wait a minute… if McCandless believes that breakups are reported more frequently on Mondays, why wouldn't that be his statistical hypothesis? It could be, but remember, the statistical hypothesis has to specify a model that can be used to generate outcomes. One such model that fits his research hypothesis is that 30% of breakups are reported on Mondays. Another is that 50% of breakups are reported on Mondays. How many others are there? In order to answer his research question, McCandless would have to examine the results from every single one of those models to see if 13 out of 50 breakups is consistent with one of those models.

**#protip:** Rather than examine each of those models, statisticians often use the 'equally likely' model. Mathematically, this model acts as a "lower bound" for all the models where Mondays has more reported breakups than other days. If 13 out of 50 breakups is an extreme (high) result given this model, we can rule out this model and suggest that the observed data are inconsistent with the model. This points toward a model in which the reported frequency of breakups is higher on Mondays, without having to know the exact percentage of breakups that occur on Monday!

2. Draw a picture of the sampler (model) that you will use to generate outcomes from the model specified in the statistical hypothesis. In the picture, be sure to (1) indicate the type of sampling device used (mixer, spinner, etc.); (2) label all the elements in your sampling device; (3) label the probability associated with each element; and (4) indicate the Repeat and Draw values you will use.

- Set up the model/sampler in TinkerPlots™.

## Simulating the Data

- Carry out a single trial of the simulation in TinkerPlots™.
- Plot the outcomes from the trial.

3. Sketch a plot of the outcomes from this trial. Add all labels and statistics (counts, percentage, etc.) to your plot.

Remember what we are ultimately interested in is the number of breakups (out of 50) that are reported on Mondays.

- Collect the appropriate statistic
- Carry out 499 more trials (500 trials total) of the simulation in TinkerPlots™.

## Evaluating the Hypothesized Model

- Plot the results from the simulation.

4. Sketch a plot of the results below.

5. The TinkerPlots model, and its results, are all reflective of the statistical hypothesis. The TinkerPlots model represents the statistical hypothesis, and its results represent what you would expect to actually happen based on that hypothesis. Based on the statistical hypothesis, interpret the results from the simulation in terms of your expectation for the number of breakups out of 50 to have occurred on Monday. Give your answer in the form of a range of numbers.

6. Now, reconsider the observed data; the data that McCandless actually observed initially. He observed 13 breakups (out of 50) that were reported on Monday. Given the plot of results, is 13 a value that is consistent with the model's results or not? Explain.

7. What does this consistency/inconsistency suggest about the answer the research question? Explain.

# Modeling the Monday Breakups Problem

MODEL

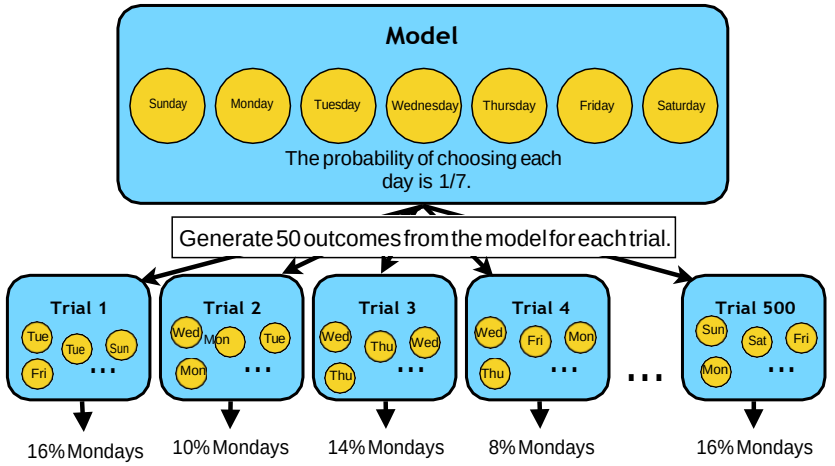
The model has seven outcomes–one for each day of the week. Each is equally likely.

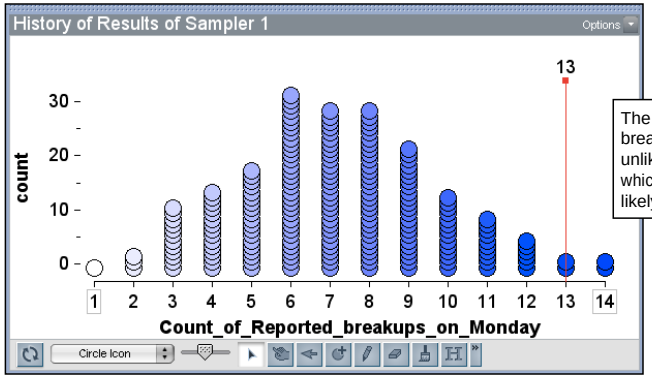
SIMULATE

A trial ends when 50 outcomes have been randomly generated from the model For each trial, compute the percentage of breakups "reported on Monday". Generate many trials.

EVALUATE

Compile all of the numerical summary measures into a single distribution, Answer the research question by evaluating whether the empirical evidence (observed result) tis consistent with the "equally likely" model.



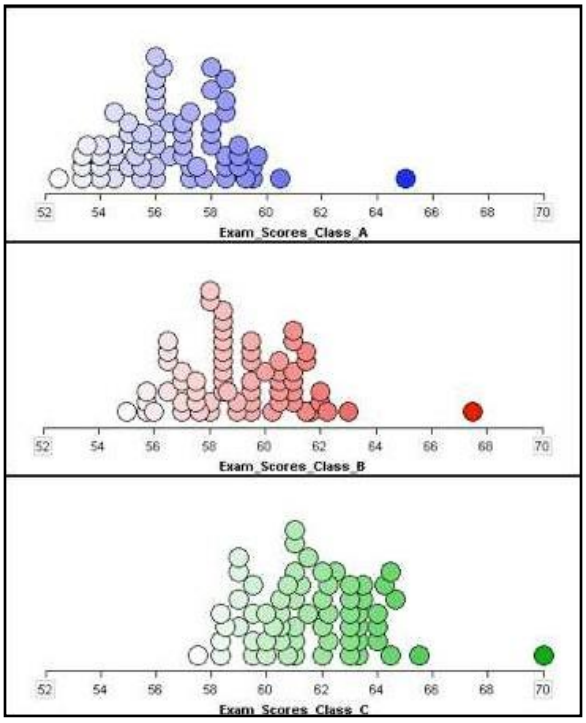




The observed result (13 breakups on Monday) is unlikely given the model in which breakups are equally likely each day of the week.

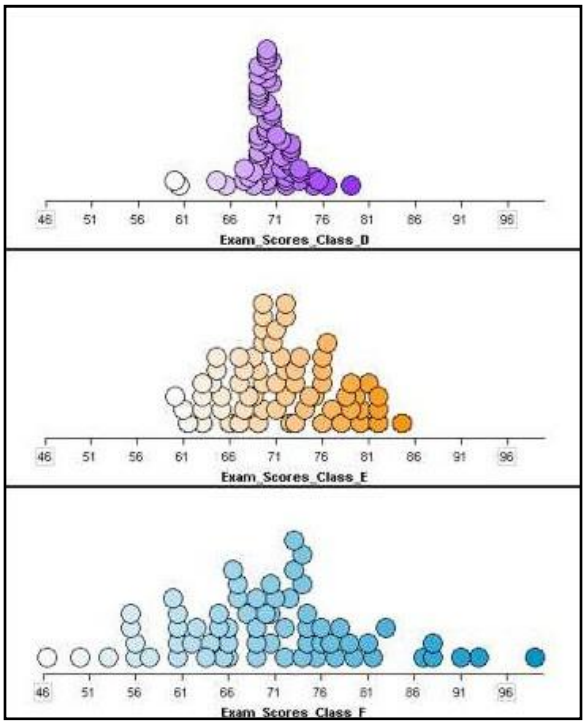
# *Features of Distributions*

Imagine multiple sections of the same college course, taught by different instructors. Below are a series of plots that depict the distributions of hypothetical exam scores in various sections.

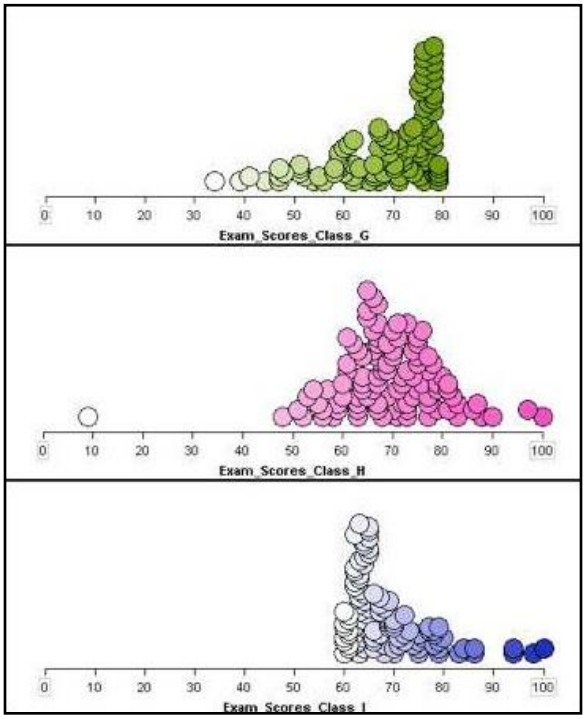
1. Examine the three distributions of exam scores for classes *A*, *B*, and *C*. What are the primary differences between these three distributions? What are potential factors that might explain the differences?



2. Examine the three distributions of exam scores for classes *D*, *E*, and *F*. What are the primary differences between these three distributions? What are potential factors that might explain the differences?

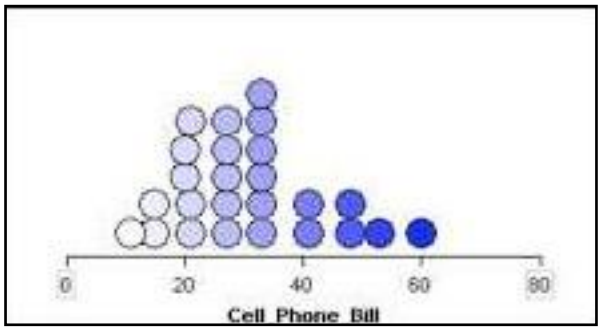


3. Examine the three distributions of exam scores for classes *G*, *H*, and *I*. What are the primary differences between these three distributions? What are potential factors that might explain the differences?



## Cell Phone Bills

Consider a survey study conducted on a random sample of 25 University of Minnesota students. One survey item asked students to self-report the amount of his or her last cell phone bill (in dollars). The plot of the bill amounts is shown below.




4. If you wanted to tell someone the amount of a "typical" cell phone bill for these students, what would you say?

5. How would you describe (quantify) the *overall* amount of variation in the distribution (i.e., for all 25 cell phone bills)?

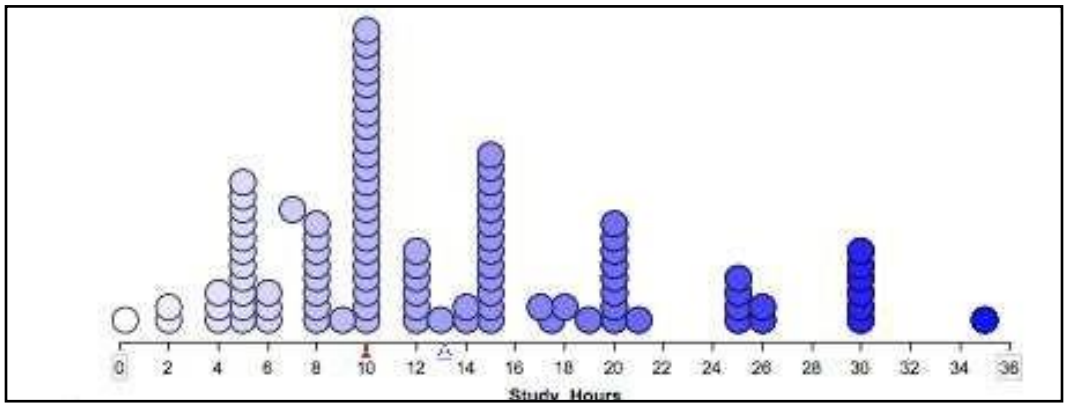
6. How far do cases typically vary from the value that you identified in Question 4?

7. What is a potential factor(s) that might explain the variation in these bills?

8. Using the typical cell phone bill you identified previously as a reference point, consider the amount of variation in the distribution on both sides of this point. Is the variation roughly the same on the left- and right-hand side of this point? Is there more or less variation on either side of this value?

## Number of Hours Studied

The plot below contains responses from 100 EPsy 3264 students who responded to the survey question: " How many hours per week do you typically study?" These students' responses are a random sample from all responses obtained from all classroom sections of EPsy 3264 taught from 2004–2010.
Examine the plot of these data.



9. What does each dot ( i.e., case) in the distribution represent?

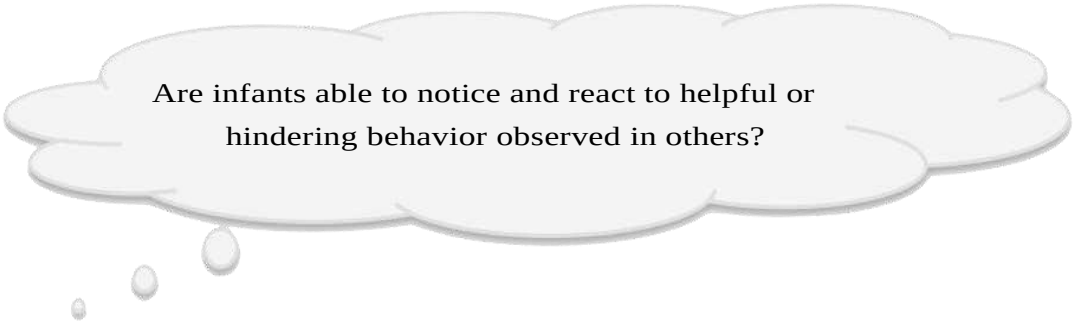
10. Summarize the features of the distribution. Be sure to identify the **"typical"** amount of time spent studying and the **variation** in the amount of studying. ( When describing the variation, you should quantify the "average" amount of deviation from the typical value.) You should also indicate the **shape** of the distribution.

11. What is a potential factor(s) that might explain the variation in these data?

# *Helper or Hinderer*

Most college students recognize the difference between naughty and nice, right? What about children less than a year old—do they recognize the difference and show a preference for nice over naughty? In a study reported in the November 2007 issue of *Nature*[1], researchers investigated whether infants take into account an individual's actions towards others in evaluating that individual as appealing or aversive, perhaps laying for the foundation for social interaction. In one component of the study, 10-month-old infants were shown a "climber" character (a piece of wood with "google" eyes glued onto it) that could not make it up a hill in two tries. Then they were alternately shown two scenarios for the climber's next try, one where the climber was pushed to the top of the hill by another character (*helper*) and one where the climber was pushed back down the hill by another character (*hinderer*). The infant was alternately shown these two scenarios several times. Then the child was presented with both pieces of wood (the helper and the hinderer) and asked to pick one to play with. The researchers found that 14 of the 16 infants chose the helper over the hinderer.

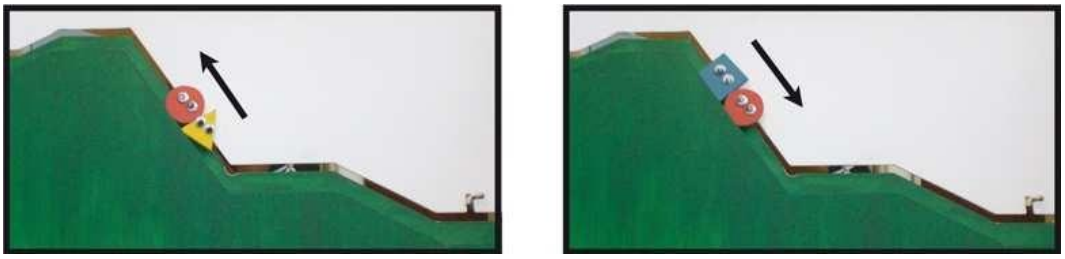
In this activity, you will be exploring the following research question:




[1] J. K. Hamlin, K. Wynn, & P. Bloom. (2007). Social evaluation by preverbal infants. *Nature, 450*, 557–559.

Before you begin, we would like you to watch the videos that were shown to the infants in the experiment. You can view them here:

- http://campuspress.yale.edu/infantlab/media/



*Helping and hindering habituation events. On each trial, the climber (red circle) attempts to climb the hill twice, each time falling back to the bottom of the hill. On the third attempt, the climber is either bumped up the hill by the helper (left panel) or bumped down the hill by the hinderer (right panel).*

## Discuss the Following Questions

1. What proportion of the infants in the observed data chose the helper toy?

2. What does that suggest about the answer to the research question? Explain.

Suppose for the moment that the researchers' conjecture is wrong, and infants *do not* really show any preference for either type of toy. In other words, infants just randomly pick one toy or the other, without any regard for whether it was the helper toy or the hinderer.

3. Write the statistical hypothesis for the "no preference" model.

4. If the no preference model is "truth", how likely do you think it is that 14 out of 16 infants would choose the helper toy? Explain.

## The "Just-by-Chance" (No Preference) Model

Consider the argument that that infants have no preference for either the helper or the hinderer. If that argument is true, then you would expect that any infants selecting the helper is only because of random chance—not due to any underlying psychological tendency of infants to prefer helpers to hinderers.

The good news is that this "just-by-chance" process (i.e., random chance) can be modeled using the same chance devices that you have been using in the course thus far. Under the assumption of "just-by-chance", the process of infants selecting toys can be modeled by randomly selecting either a helper or a hinderer. After the random selection, it can be determined how many infants "chose" the helper. This process can be repeated a large number of times to simulate the percentage of infants selecting the helper under this "no preference" model.

## Summary of the Simulation Process

The key to answering the research question in this activity is to determine how likely the observed result (14 of 16 infants choosing the helper) is under the assumption that infants have no preference for either the helper or the hinderer. The "no preference" model is again the "just-by-chance" model—infants randomly select either the helper or hinderer.

To determine this likelihood, you will model the process of 16 hypothetical infants making their selections using random chance. Then, you can count how many of these "infants" choose the helper toy. This process can be repeated many times to obtain a distribution of results that would be expected under the "no preference" or "just-by-chance" model.

The observed result of 14 of 16 infants choosing the helper can then be evaluated in light of this distribution to determine how likely it would be to obtain such a result (or a more extreme result) under the assumption of random chance. As such, the observed result can provide evidence to help answer the research question.

5. Draw a picture of the sampler (model) that you will use to generate outcomes from the "just-by-chance" model. In the picture, be sure to (1) indicate the type of sampling device used (mixer, spinner, etc.); (2) label all the elements in your sampling device; (3) label the probability associated with each element; and (4) indicate the Repeat and Draw values you will use.

## Simulating Data from the Hypothesized Model

In this study, a trial represents each of the 16 infants choosing a toy. The trial ends when 16 toys have been chosen randomly.

- Carry out a single trial of the simulation in TinkerPlots™.
- Plot the outcomes from the trial.

6. What is the statistic from the plot that you will be **collecting** ?

- Collect the appropriate statistic.
- Carry out 499 more trials (500 trials total) of the simulation in TinkerPlots™.
-

## Evaluating the Hypothesized Model

- Plot the results from the simulation.

7. Sketch a plot of the results below.

8. Describe the shape, center, and variation for the distribution of results.

9. Is the observed result from the original experiment likely or unlikely under the hypothesized model? Explain.

10. What does your answer to Question #9 suggest about whether infants are able to notice and react to helpful or hindering behavior observed in others? Explain.

# Helper or Hinderer

**MODEL**

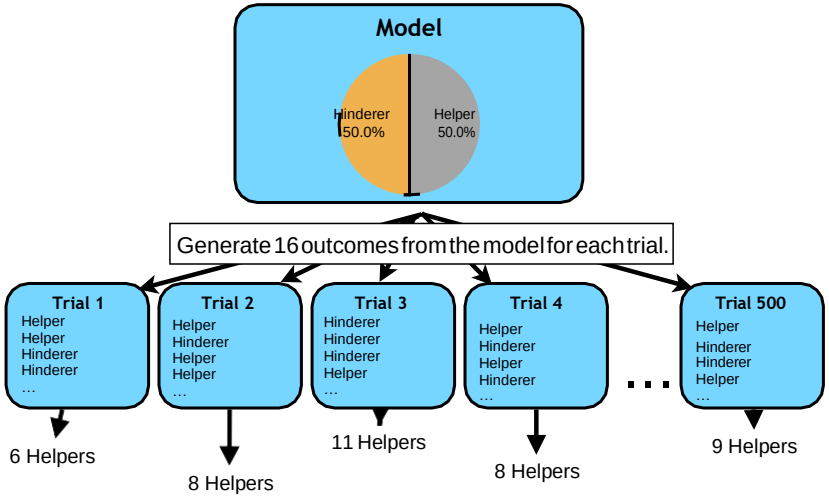
The model has two outcomes– one for Helper and one for Hinderer. Each is equally likely.

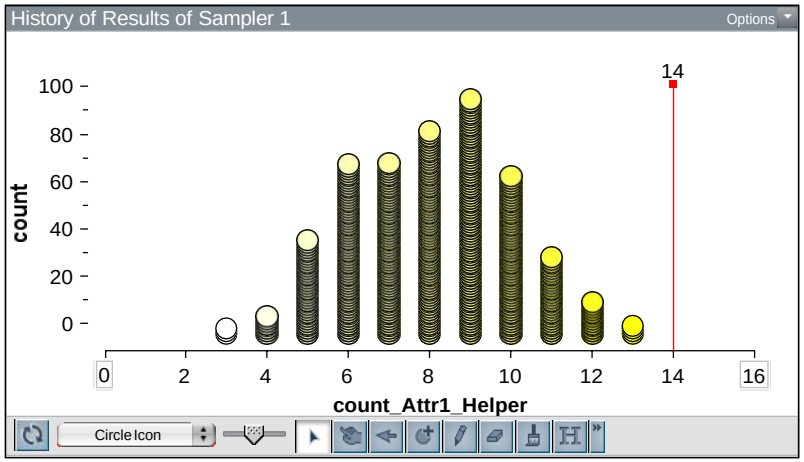
**SIMULATE**

Randomly generate 16 outcomes from the model. For each trial, compute the number of Helpers. Generate many trials.

**EVALUATE**

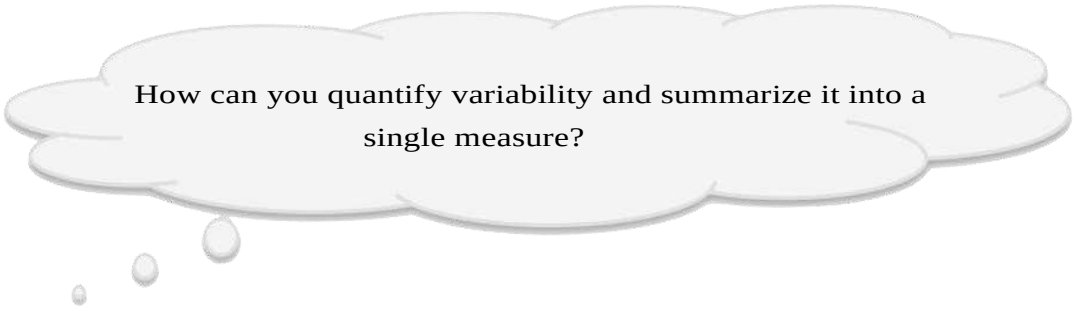
Compile all of the numerical summary measures into a single distribution, Evaluate the initial questions by deciding whether the observed result of 14 Helpers is likely under the hypothesized model.



# *Comparing Hand Spans*

In this activity, you will learn about the standard deviation, a common measure of variability. You will not be using TinkerPlots for this activity. Use a sheet of paper.




1. Measure and record the hand span for each person in your group.

2. Enter the data into a case table. Create a plot of the hand spans for your group. Sketch the plot below. Be sure to appropriately label the *x*-axis.

3. Compute the mean hand span for your group. Record the mean.

## The Standard Deviation

Recall that the mean is a single number that can be used to summarize the data. In this context, it is a description of the typical hand span measurement for your group. Of course, not every student in the sample is at the typical value (in fact all of them might be different from the typical value). Thus, it is also useful to have a single number description of how different the data tends to be from this typical value.

One single number description of the variability in a sample of data is called the **standard deviation** or *SD*. If the word "typical" is substituted for the word "standard" in its name, the name standard deviation (typical deviation) makes more sense. This measure quantifies variability by determining how far data cases typically deviate from the mean value.

- Create a new attribute in the table, called *Deviations*, that contains the difference between the observed data (hand spans) and the mean of your group members' hand spans. Compute these by subtracting the mean from each observation.
- Create a plot of the *Deviations* attribute.

4. How would you interpret the values of the *Deviations* attribute?

5. Sketch the plot below. Make sure to label the *x*-axis. Circle the mean and record its value.

6. How does the distribution of deviations compare with the distribution of hand span lengths you created in Question 2?

One thing you may not have known about the mean is that it is the value that "balances" the data. In other words, the mean is the value that gets the deviations to sum to zero. This is useful when describing a typical value of the data (it is the "closest" point to all of the cases, on average). If you try to average these deviations, however, you will always get zero. This is not very useful in summarizing variation in a data set, nor in comparing the variation between two data sets. One way to alleviate this problem is to square each of the deviations before you add them together.

- Create another attribute, *SquaredDeviations* , which contains the squared values of the deviations.
- Create a plot of the *SquaredDeviations* attribute.

7. Sketch the plot below. Make sure to label the *x*-axis.

8. How does the distribution of the squared deviations compare to the distributions of the original hand spans? How does it compare to the distribution of *deviations*?

9. Compute the mean of the squared distributions and record its value below. Because the deviations have been squared, this value represents the **typical squared deviation**.

10. Find the square root of the mean value and record its value below.

11. What does this new value represent (i.e., interpret its value)?

## Interpreting the Standard Deviation of a Plot of Results

12. Describe the shape, center, and variation for the distribution of results. This time, rather than giving a more informal description of the variation, compute the standard deviation.

13. From statistical theory, we know that most observations in a distribution are within two standard deviations of the mean. Add and subtract two standard deviations from the mean to complete the following sentence:

Most simulated means will be between _____ and _____.

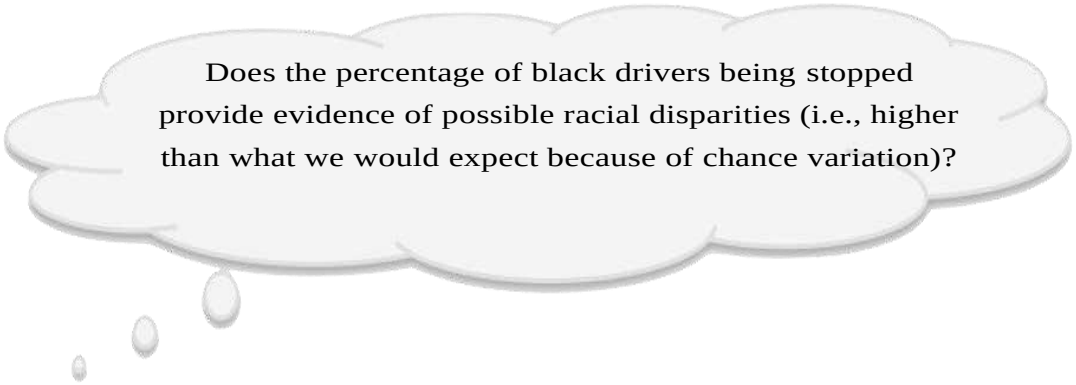
Most statisticians define *likely* results as those that are within two standard deviations of the mean. Anything more than two standard deviations from the mean would be called *unlikely*.

# *Racial Disparities in Police Stops*

In roughly a one year period, beginning in October of 2016, the Minneapolis Police Department recorded 50,950 traffic stops in their public database. In this activity, we will use these data to analyze whether black drivers were more likely than others in the population to be targeted by police for traffic stops.

On Halloween (October 31) 2017, the Minneapolis Police Department made 105 traffic stop. Of those, 39 of the drivers were black. At that time Minneapolis' population was 18.6% black.

In this activity, you will be exploring the following research question:




## Discuss the Following Questions

1. What percentage of the drivers that were stopped were black?

2. What does that suggest about the answer to the research question? Explain

Suppose for the moment that there are *no* racial disparities in police stops. In other words, the percentage of blacks stopped by the police should reflect the percentage of blacks in the population.

3. What percentage of the drivers that were stopped would you expect to be black, given the population?

In both the *Helper or Hinderer* course activity and the *Monday Breakups* course activity, all of the elements in the simulation model had the same probability. When the probabilities of each element in the model are exactly the same, we say the "just-by-chance" model is a **uniform probability model**.

The "just-by-chance" model does not have to be a uniform probability model. The elements can have differing probabilities. As long as the elements being selected are still random, this is still a "just-by-chance" model.

4. Using the value from the previous question, write the hypothesis for the "just-by-chance" model.

5. Draw a picture of the sampler (model) that you will use to generate outcomes from the "just-by-chance" model. In the picture, be sure to (1) indicate the type of sampling device used (mixer, spinner, etc.); (2) label all the elements in your sampling device; (3) label the probability associated with each element; and (4) indicate the Repeat and Draw values you will use.

- Set up the model/sampler in TinkerPlots™.

## Simulating the Data

- Carry out a single trial of the simulation in TinkerPlots™.
- Plot the outcomes from the trial.

6. What is the statistic from the plot that you will be collecting?

- Collect the appropriate statistic from the plot of your first trial.
- Carry out 499 more trials (500 trials total) of the simulation in TinkerPlots™.

## Evaluating the Hypothesized Model

- Plot the results from the simulation.

7. Sketch a plot of the results below.

### *Computing the Standard Deviation using TinkerPlots™*

Use TinkerPlots™ to find the standard deviation of the original data directly.

- In the results from the simulation table where you collected the appropriate statistic from 500 trials, create a new attribute called standardDeviation.
- Use the Formula Editor to compute the standard deviation of the hand spans by using the stdDev() function.

8. Describe the shape of the distribution. Also compute the mean and standard deviation.

9. Use the mean and standard deviation to compute the range of *likely* results.

10. Given the range of *likely* results you just computed, how compatible is the observed result with the results from the hypothesized model? Explain.

11. Use the evidence from the simulation to answer the research question.

# Racial Disparities in Police Stops?

**MODEL**

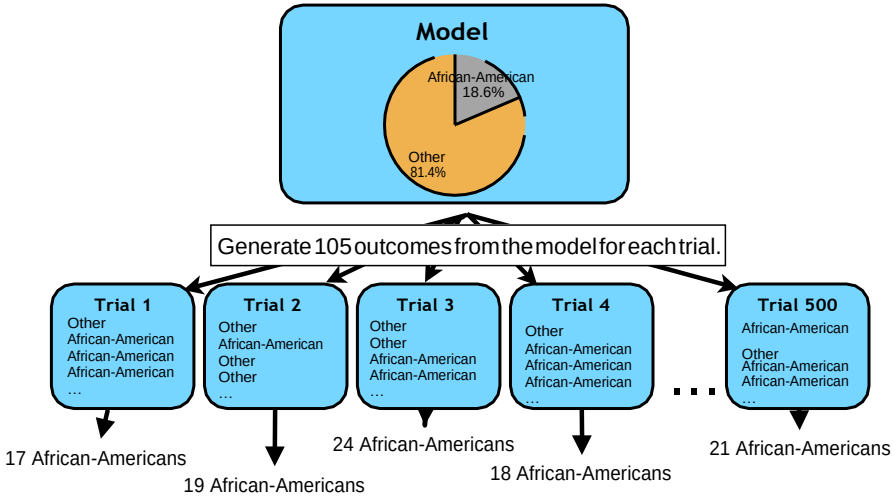
The model has two outcomes–one for African-Americans and one for other races. Each is **not** equally likely.

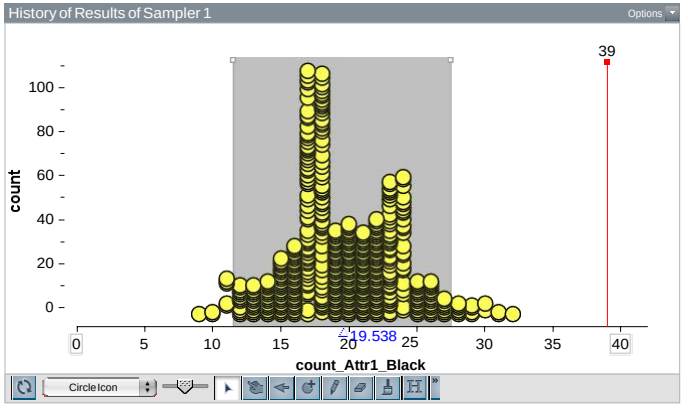
**SIMULATE**

Randomly generate 105 outcomes from the model. For each trial, compute the number of African-Americans. Generate many trials.

**EVALUATE**

Compile all of the numerical summary measures into a single distribution, Evaluate the initial questions by computing the range of likely values under the hypothesized model–two standard deviations from the mean. Determine whether or not the result from the observed data is within the range of likely results.



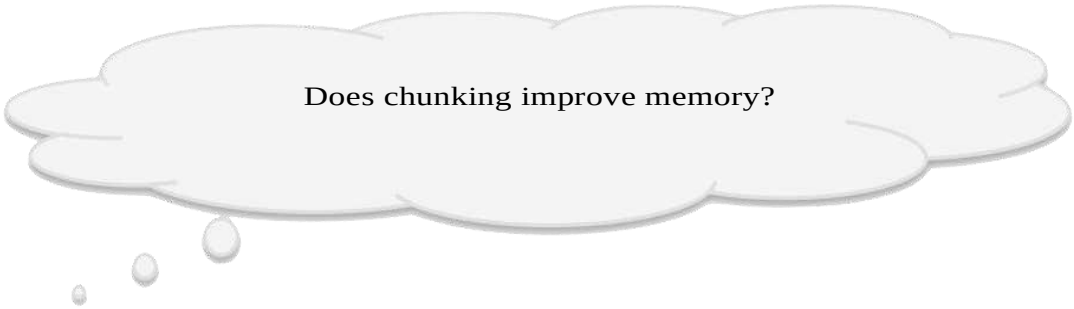



# *Memorization*

Many times during the semester, you may feel like your brain just cannot hold all of the information you are learning in classes. Are there ways to improve our memories so that we can comprehend even more information? Research in cognitive psychology has suggested that the answer to that question is a resounding " yes". This literature has suggested several strategies to improve memory, enhance recall and increase retention of information.

One of the strategies identified by cognitive psychologists is that of chunking. Chunking refers to the process of taking individual units of information and grouping them into larger units (chunks). One common example of chunking occurs when we write and recall phone numbers. For example, a sequence of digits in a phone number, say 8-6-7-5-3-0-9, would be chunked into 867-5309.

In this activity, you will be exploring the following research question:




To examine this research question, you will use the data collected from the memory experiment your class just partook in.

## Examining the Observed Data

The first part of any analysis is to examine the observed data. These are the *data that are actually observed* in the research study. In this study we have data on two attributes for each participant in the study.

- The first attribute we have information about is the participant's score (i.e., the number of letters recalled) from the memory experiment. This is called the **response variable** since it contains data on the subjects' responses to the experiment.
- The second attribute we have information about indicate the treatment condition that the subject was assigned to. This is called a **treatment variable**. In this research study the two *levels of the treatment variable* (the two conditions) are the **experimental condition** (chunking) and the **control condition** (no chunking).

1. Based on the scores, does it seem like there is an effect of chunking? In other words, does it seem like the scores are higher for the chunking group than for the non-chunking group? Explain.

## Summarizing the Difference Between the Two Conditions

In order to answer the research question, you need to summarize the difference between the treatment and control conditions into a single number.

When the response variable is quantitative, it is conventional to do this by finding the mean value of the response variable for each condition, and then *compute the difference between the two means*. The difference in means satisfies the need for a single number statistic. It also has another very nice quality, and that is the difference in means is interpretable. The difference in means indicates *how much better* the typical subject in the experimental condition does than a typical subject in the control condition.

2. Compute and record the mean score for each of the two conditions.

3. Compute the difference in means by subtracting the mean score for the non-chunking condition from the mean score for the chunking condition.

*Note that this difference is the difference in means for the observed data because we used the observed data (the data from our study) to compute it.*

4. Interpret this difference using the context of the memory study.

5. Does the difference you found in the observed data *suggest there is an effect* of chunking on memory or not? Explain.

## Considering Experimental Variation as an Explanation for the Difference in Means

Before you conclude that chunking has an effect on memory, consider another alternative: *the difference in means you saw in the observed data is solely attributable to experimental (chance) variation*. Under this model, the difference in means is not because chunking works, but rather because **the random assignment to conditions/groups introduces variation into the results**.

6. If there is *not* an effect of chunking on memory, what would you expect the difference in means to be? Explain.

### *The No-Effect Model*

To examine whether a result obtained in the observed data is solely due to chance (i.e., all the variation is due to the random assignment), one approach is to imagine the *scenario under which the chunking had no effect*, whatsoever. Under this assumption or scenario, evidence would be collected to determine if the difference in means that was observed in the data is too large to probabilistically believe that there is no effect of chunking. This statement or assumption of no effect of chunking is our statistical hypothesis, which is also called the **null hypothesis** and is written as,

$H_0$: There is no difference in the mean number of letters recalled between the control and experimental conditions.

If the null hypothesis is true, then chunking is truly ineffective, and each subject's score on the memory test is only a function of that person and not a function of anything systematic, such as the chunking. The implication of this is that, had a subject been assigned to the other condition (through a different random assignment), their score on the memory test would have been identical since, in a sense, both conditions are doing nothing in terms of affecting the memory test scores.

## Re-randomization: Inspecting Other Possible Random Assignments of the Subjects

A researcher can take advantage of the idea that each subject's score on the memory test would be identical whether she was assigned to treatment or control and examine other possible random assignments of the subjects to conditions that could have occurred. To do this, you will carry out a physical simulation (not using TinkerPlots™).

### *Physical Simulation of the Re-Randomization*

To aid you in creating these "new" random assignments of conditions, fill in the following:

> In the original experiment,_____subjects were randomly assigned to the experimental (chunking) condition and_____subjects were assigned to the control (no chunking) condition.

- You will be given several index cards. Each index card represents a single subject. On each card you will write an *E* (for experimental) or a *C* (for control). When you are done, you should have the same number of *E* cards as subjects originally assigned to the experimental condition and the same number of *C* cards as subjects originally assigned to the control condition. Set the *E* and *C* cards to the side.
- Now, record the first subject's name and score (number of letters correct) on new card. Continue with the other subjects' names and scores, recording each subject on a different card. At this point you should have *n* subject cards (with names and scores), and *n* condition cards (with an *E* or a *C*), where *n* is the total number of subjects in the combined control and experimental groups.
- Shuffle the *E* and *C* index cards together several times.
- Shuffle the index cards with the scores several times.
- Deal the shuffled *E* and *C* index cards out one at a time. Now deal the score cards out one at a time, placing each score card you deal on one of the *E* or *C* index cards.

This represents one possible randomization of subjects to either the experimental or control conditions. It is another possible way the subjects could have been assigned to conditions. This random assignment likely has different subjects in the control and experimental conditions than the observed data. Because of this, the mean memory score for the two conditions will also likely differ from the observed data. This, in turn, implies that the difference in means will also be different.

7. Record the subjects' scores based on this possible randomization below. Record subjects assigned to the *E* condition under *Experimental* heading, and those assigned to the *C* condition under the *Control* heading.

**Experimental** **Control**

8. Compute the means for the data from this random assignment for each condition and record them below.

9. Compute and record the **difference in means** for this random assignment of the data. Be sure that the order you use when subtracting is consistent with the order you subtracted to obtain the original observed result. (Note: You may obtain a negative number here.)

- Repeat the random assignment process four more times (five total). Each time, record the data, compute and record the mean score for each condition, and compute and record the difference between the means of the two groups. (Remember to subtract in the same order each time.)
- Record each of the five differences you obtained on the board.

## Examining the Distribution of the Difference in Means

10. Enter all the groups' mean differences into TinkerPlots™. Create a plot of the difference in means. Sketch the plot of the difference in means below.

11. Does it look like it centers around zero? Explain why the distribution should be centered at zero. (Hint: Think back to what the null hypothesis was.)

12. Use TinkerPlots™ to compute the standard deviation of the differences in means. Record that value below.

13. Use the mean and standard deviation to provide a range of likely results.

14. Now include a vertical line at the difference in means for the original (observed) data. How compatible is the observed difference in means with the results produced by the model specified in the null hypothesis?

15. Based on your response to the previous question, is the "no effect" model supported by the observed data or not? What does this suggest about the answer the research question? Explain.

**Overview of the Inferential Process for Comparing the Two Conditions in the Memorization Experiment**

If there really were no effect of the grouping of letters, is it possible that random chance alone could have resulted in such an extreme observed difference between the two conditions?
Once again, the answer is yes, this is indeed possible. Also once again, the key question is *how likely would it be for random chance alone to produce experimental data that favor the chunking condition by at least as much as the observational data do*. You will aim to answer that question using the following simulation analysis strategy:

- **Model:** Assume that there is no effect of the grouping of letters on the scores (the "no effect" model).

- **Simulate:** Replicate the random assignment of these subjects and their memory scores between the two conditions. You will repeat this random assignment a large number of times. Each time you will calculate a measure of how different the conditions are, in order to get a sense for what is expected and what is surprising.

- **Evaluate:** Using the observed result, evaluate how compatible the observed result is with the simulated results produced by the model specified in the null hypothesis.

# *Memorization Using TinkerPlots™*

In this activity, you will learn how to use TinkerPlots™ to carry out the randomization test.

## Examining the Observed Data

The first part of any analysis is to examine the observed data. These are the *data that are actually observed* in the research study.

- Enter the data collected in the study into a TinkerPlots™ case table.

**Setting up the Case Table for the Randomization Test**

- Drag a Table from the object toolbar into your document.
- Create a new attribute called *Score* in the first column of the case table.
- Create another new attribute called *Condition* in the second column of the case table.

Each row in the table will comprise a subject in the research study. Each column will comprise an attribute of the subject. For our purposes, you will need to enter data for two attributes. The first attribute will indicate the subject's score (i.e., the number of letters recalled) from the memory experiment. This is called the **response variable** since it contains data on the subjects' responses to the experiment. The second attribute will indicate the treatment condition that the subject was assigned to. This is called a **treatment variable**. In this research study the two treatment conditions are the **experimental condition** (chunking) and the **control condition** (no chunking).

- Enter the observed data from your class experiment into a TinkerPlots™ case table.
- Plot the observed data (see instructions on next page).

**Plotting Data to Compare Groups**

- Drag a Plot from the object toolbar into your document.
- Drag the response variable from the case table to the *x*- axis of the plot.
- Drag a case icon to the right until the cases in the plot are fully separated (e.g., no vertical bin lines). You can also double-click on one of the endpoints and change Bin width = 0 .
- Drag the treatment variable from the case table to the *y*-axis of the plot.
- Click the Vertical Stack button in the upper plot toolbar to organize the icons.

1. Sketch the plot below that you just created in TinkerPlots™.

## Model the Experimental Variation Due to Random Assignment

In order to carry out a randomization test using TinkerPlots™, you need to include multiple sampling devices in the sampler. The first sampling device will include the observed response data for all of the subjects. The second device will contain the experimental conditions.

### *Modeling a Set of Fixed Responses Under the "No Effect" Model*

Under the null hypothesis of no difference between the two experimental conditions, the response values for the subjects are **fixed**—they will always be the same for the subjects, regardless of which experimental condition the subject is assigned. To produce simulated data that are fixed, you can use a Mixer, but the values need to be *selected without replacement*.

- Set up a Mixer that will produce the fixed responses for the subjects under the "no effect" model (see instructions below).
- Run the model a couple times.

**Setting Up the Model: Fixed Responses**

- Drag a new Sampler from the object toolbar into your blank document.
- The default device in the sampler is a Mixer with three elements. Add elements to the mixer until you have the same number of elements as there are responses. (Each element represents a participant in the experiment.)
- Change the values of the elements so that they represent the response values ( i.e., the number of correctly memorized letters).
- Change the mixer to sample values without replacement. Do this by clicking on the Device Options button for the stacks device (upside-down triangle) and selecting Replacement.
- Change the name of the device from *Attr1* to *Responses* .
- Change the Draw value to *1*. Change the Repeat value to reflect the total number of participants in the experiment.

2. Do you get different response values in the outcomes when you run the model? Use your answer to explain why we refer to these as **fixed** outcomes.

## *Modeling the Random Assignment of the Treatment Condition Labels by Linking Multiple Devices*

3. Write a detailed explanation describing the process you used to physically re-randomized the notecards in the previous activity. Be specific enough in this description that another student could replicate what you did.

To model the random assignment of the treatment condition labels that might have occurred, you need to produce simulated data from another model that generates labels of *Experimental* and *Control*. To do this you will use the Stacks sampling device. We also need to include this sampling device in the same Sampler as the outcomes. To do this, you *link* multiple sampling devices in the same sampler the same way you did in the *Pet Factory* course activity.

You need to replicate the original experiment and have the same number of *Experimental* labels and *Control* labels as were in the original experiment.

- Link another sampling device that includes the **fixed group/condition labels** to the sampling device containing the outcomes. (See instructions on next page.)
- Run the model.

**Linking a Sampling Device to Model the Random Assignment of Conditions**

- Drag a Stacks sampling device from the device menu to the right-hand side of the existing *Responses* device. The sampler should now contain two devices linked by a grey line.
- Change the device name from *Attr2* to *Conditions*.
- Click the Add Element button (+) twice to add two elements to the stacks. These elements will indicate the condition/group labels.
- Change the label of the first bar from *a* to *E*. Change the label of the second bar from *b* to *C*.
- Click on the Device Options button for the stacks device (upside-down triangle) and select Show Count.
- Change the count value for the *E* label to reflect the number of participants originally assigned to the *E* condition. Change the count value for the *C* label to reflect the number of participants originally assigned to the *C* condition.
- Change the device to sample values without replacement.

When you add linked devices, remember that the value for Draw changes automatically to the number of devices included in the sampler. A TinkerPlots™ sampler showing two linked devices modeling the random assignment of responses to conditions is shown below.

The outcomes from both linked devices are recorded in the case table, each in their own attribute. In addition, an attribute called *Join* is also created that includes the outcomes of both linked devices separated by a comma.

Each trial represents w*hat might have occurred* under another random assignment of subjects to conditions *if there was no difference between the conditions*.

- Plot the trial data to obtain the difference in means. (Remember the response attribute from the trial's case table is dragged to the *x*-axis of the plot and the condition is dragged to the *y*-axis of the plot.)

4. Sketch the plot below.

## Simulate: Collect Results from Many Trials

As you have done in previous simulations, you will numerically summarize the trial results. We summarized the observed data by computing,

$$\bar{X}_E - \bar{X}_C$$

5. Compute the value ( by hand) for this difference. Be sure you are subtracting the *Control* mean from the *Experimental* mean.

To use TinkerPlots™ to collect the difference in means, we need to collect both the *Experimental* and the *Control* means. Fortunately, we can collect multiple summaries in TinkerPlots™.

- Use TinkerPlots™ to collect the mean score for the *Experimental* condition.
- Similarly, collect the mean score for the *Control* condition.

Now you should have a case table of results that includes the *Experimental* and the *Control* mean in the same row ( but in different columns).

### *Computing the Difference in Means*

To compute the difference in means we will use the Formula Editor. The Formula Editor allows us to compute new measures from existing information in a case table.

- Use the Formula Editor to compute the difference in the trial's means (see instructions below).
- Check that the difference in means is the same as the difference you computed in the previous question. (If the difference calculated by TinkerPlots™ is correct, but has a reversed sign, you need to re-open the Formula Editor and re-compute the mean difference.)

**Computing the Difference in Means**

- Create a third attribute (column) in the case table by clicking the column name, **<new>** . Rename this attribute *Difference*
- Select the *Difference* attribute to highlight it and then right-click the attribute and select **Edit Formula**.
- Select the *Attribute* triangle to display the names of the case table's attributes in the **Formula Editor** .
- Double-click the attribute for the *Experimental* mean value. Then click the subtraction key ( **–** ) in the **Formula Editor** calculator. Finally, double-click the attribute for the *Control* mean value.
- Click the **Apply** button and then click **OK** .

### *Collect Many Results*

- In the *History of Results* table, collect an additional 499 measures. (See instructions below for speeding up the simulation.)
- Plot the differences in means from the 500 simulated trials.

**Speeding Up the Simulation**

- Minimize all of the objects (sampler, results table, plot of the results) except for the collection window.
- Select the collection window and from the Objects menu select Inspect Collection.
- Uncheck the Animation On option.
- Close the inspector window.
- In the History of Results table, change the number of samples to collect to 499.
- Click the Collect button.

6. Sketch the plot of the results (i.e., mean differences) from the 500 simulated trials below.

7. What are the cases in the plot? ( Hint: Ask yourself what each individual dot represents.)

8. Where is the plot of the results centered (at which value)? Explain why this makes sense. (Hint: Think about what the hypothesis for the "no effect" model is.)

9. Use TinkerPlots™ to compute the standard deviation of the differences in means. Record that value below.

10. Using the mean and standard deviation, provide a range of likely values under the model that assumes the difference in means is due completely to random chance.

11. Now include a vertical line at the difference in means for the original (observed) data. How compatible is the observed difference in means with the results produced by the model specified in the null hypothesis?

12. Based on your response to the previous question, is the “no effect” model supported by the observed data or not?

14. How would you rate the level of internal validity evidence based on the study design? Explain.

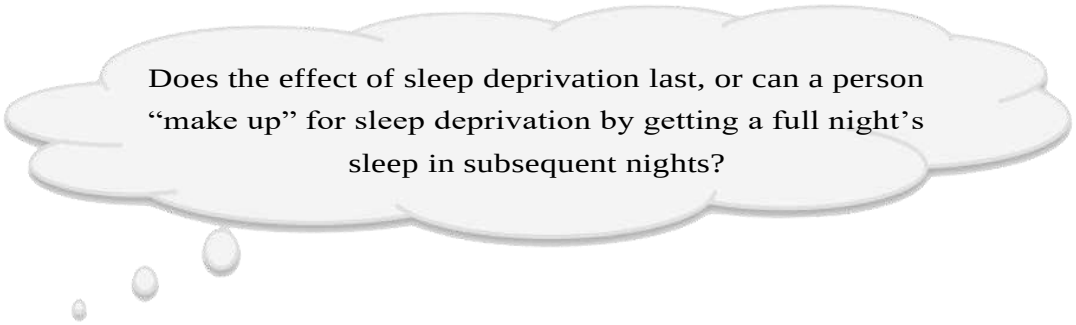
15. Based on your response to the previous question, are you willing to draw a causal association? If not, offer at least two other possible explanations.

# *Sleep Deprivation*

Sleep deprivation has been shown to have harmful effects such as fatigue, daytime sleepiness, clumsiness and weight loss or weight gain. Researchers have also established that sleep deprivation has a harmful effect on learning. But do these effects linger for several days, or can a person " make up" for sleep deprivation by getting a full night's sleep in subsequent nights?

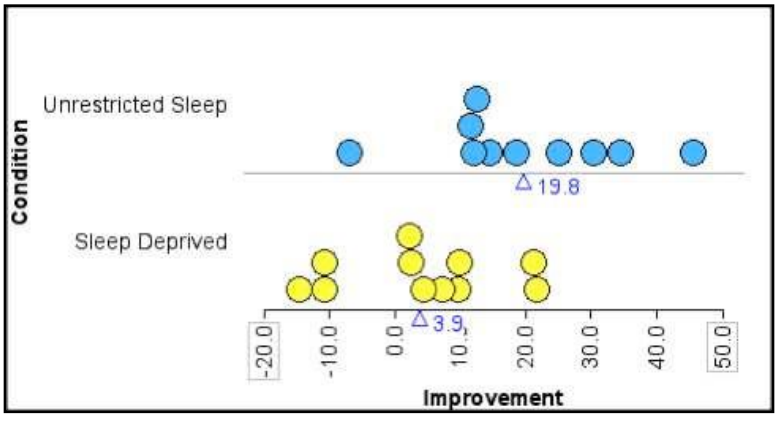
Stickgold, James, and Hobson ( 2000), in a recent study, investigated this question by randomly assigning 21 subjects (volunteers between the ages of 18 and 25) to one of two groups: One group was deprived of sleep on the night following training and pre-testing with a visual discrimination task, and the other group was permitted unrestricted sleep on that first night. Both groups were then allowed as much sleep as they wanted on the following two nights. All subjects were then re-tested on the third day.

In this activity, you will be exploring the following research question:



Subjects' performance on the test was recorded as the minimum time (in milliseconds) between stimuli appearing on a computer screen for which they could accurately report what they had seen on the screen. The sorted data and plots presented here are the improvements in those reporting times between the pre-test and post-test (a negative value indicates a decrease in performance):

| **Sleep Deprived** ($n$ = 11) | **Unrestricted Sleep** ($n$ = 10) |
|---|---|
| –14.7 | –7.0 |
| –10.7 | 11.6 |
| –10.7 | 12.1 |
| 2.2 | 12.6 |
| 2.4 | 14.5 |
| 4.5 | 18.6 |
| 7.2 | 25.2 |
| 9.6 | 30.5 |
| 10.0 | 34.5 |
| 21.3 | 45.6 |
| 21.8 | |




*Observed data and plot of the observed data for the sleep deprivation study. The triangle under each plot indicates the mean improvement score for the respective group.*

**Discuss the following questions.**

1. Does it appear that subjects who got unrestricted sleep on the first night tended to have higher improvement scores than subjects who were sleep deprived on the first night? Explain briefly.

2. Is the mean improvement higher for those who got unrestricted sleep? Calculate the difference in the means of the improvement scores.

3. Is it possible that there is really no harmful effect of sleep deprivation, and random chance alone produced the observed differences between these two groups?

## Model the Experimental Variation Due to Random Assignment

- Set up a sampling device that will produce the **fixed responses** for the subjects under the "no effect" model.
- Link another sampling device that includes the **fixed group/condition labels**.
- Run the model.

## Simulate and Evaluate the Results

- Use TinkerPlots™ to plot the randomized data and collect the mean for each condition. Then use the Formula Editor to compute the difference in means. Simulate an additional 499 randomizations (trials) of the data (500 total).
- Plot the results (difference in means) from the 500 randomizations.

4. Sketch the plot of the results (i.e., mean differences) from the 500 simulated trials below.

5. What are the cases in the plot? ( Hint: Ask yourself what each individual dot represents.)

6. Where is the plot of the results centered (at which value)? Explain why this makes sense. (Hint: Think about what the hypothesis for the “no effect” model is.)

7. Use TinkerPlots™ to compute the standard deviation of the differences in means. Record that value below.

8. Using the mean and standard deviation, provide a range of likely values under the model that assumes the difference in means is due completely to random chance.

9. Now include a vertical line at the difference in means for the original (observed) data. How compatible is the observed difference in means with the results produced by the model specified in the null hypothesis?

10. Based on your response to the previous question, is the "no effect" model supported by the observed data or not? What does this suggest about the answer the research question? Explain.

## *Quantifying the Likelihood of the Observed Result Given the Model: p-Value*

11. Compute and report the *p*-value for the observed difference in the Sleep study. (If you have forgotten how to do this, consult the *p*-value reading.)

12. Interpret the *p*-value you computed.

    The *p*-value of _____ is the probability of …

13. Based on the *p*-value you computed, how compatible is the observed difference in means with the results produced by the model specified in the null hypothesis? What does this suggest about the answer to the research question? Explain.

## Design and Inference

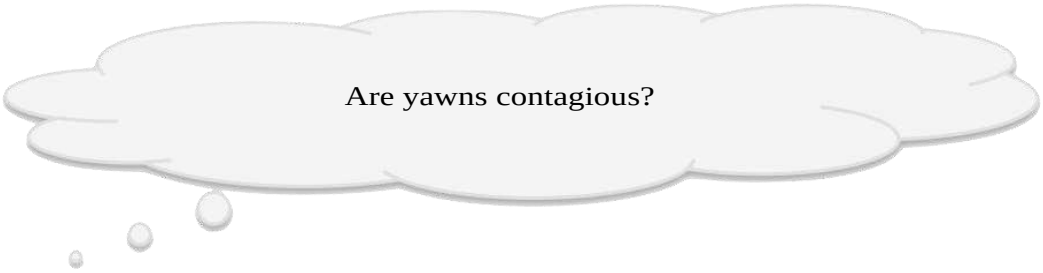
16. How would you rate the level of internal validity evidence based on the study design? Explain.

17. Based on your response to the previous question, are you willing to draw a causal association? If not, offer at least two other possible explanations.

# *Contagious Yawns*

Conventional wisdom says yawns are contagious; when you see someone else yawn, you are prone to feel sleepy and let out a yawn yourself. How many times have you caught yourself in this situation, or noticed it in someone else? But will this hypothesis withstand a scientific test? Will data support this claim?

The folks at MythBusters, a popular television program on the Discovery Channel, investigated this issue by using a two-way mirror and a hidden camera. Fifty subjects sat in a booth, accompanied only by an experimental attendee. For some of the subjects, the attendee yawned (planting a yawn "seed"), while for other subjects the attendee did not yawn. The researchers decided in advance, with a random mechanism, which subjects would receive the yawn seed and which would not. As time passed, the researchers watched to see which subjects yawned. In this activity, you will answer the following research question:



## Explore the Observed Data

1. Based on the research question, specify the treatment variable.

2. Based on the research question, specify the response variable. Also, identify whether it is a quantitative or categorical variable.

**Observed Data:** The researchers found that ten of 34 subjects who had been given a yawn seed actually yawned, compared with four of 16 subjects who had not been given a yawn seed.

3. Organize these data/results (i.e., frequencies) into a 2x2 table. This table is sometimes referred to as a **contingency table**.

| | **Yawn** | **No Yawn** | **Total** |
|---|---|---|---|
| **Yawn Seed** | | | |
| **No Yawn Seed** | | | |
| **Total** | | | |

4. Of the 34 subjects assigned to the yawn seed condition, what proportion yawned?

5. Of the 16 subjects assigned to the no yawn seed condition, what proportion yawned?

6. Find the difference between the proportion of subjects assigned to the yawn seed condition that yawned and the proportion of subjects assigned to the no yawn seed condition that yawned.

7. Write a few sentences summarizing the results in the sample. This should include a summary of what the data suggest about: (1) the *overall percentage of participants who yawned;* (2) the differences between the two treatment groups; and (3) whether or not the data appear to support the claim that yawns are contagious.

## Dummy Coding the Outcome

One trick that statisticians use to deal with categorical outcomes, is to *dummy code* the variable. Dummy coding is a way to turn the categories into numbers, so that the outcome becomes "quantitative". Once the outcome is quantitative, we can compute means, etc.

The idea of dummy coding is that each category of the outcome gets a numerical value of either "1" or "0". For example, consider the 16 subjects who did not receive the yawn seed. In the data, four of them yawned; give those folks a "1 ". Twelve of them did not yawn; assign them a "0". Their data looks like this:

| Condition | Outcome | Yawn? |
|---|---|---|
| No yawn seed | Yawn | 1 |
| No yawn seed | Yawn | 1 |
| No yawn seed | Yawn | 1 |
| No yawn seed | Yawn | 1 |
| No yawn seed | No yawn | 0 |
| No yawn seed | No yawn | 0 |
| No yawn seed | No yawn | 0 |
| No yawn seed | No yawn | 0 |
| No yawn seed | No yawn | 0 |
| No yawn seed | No yawn | 0 |
| No yawn seed | No yawn | 0 |
| No yawn seed | No yawn | 0 |
| No yawn seed | No yawn | 0 |
| No yawn seed | No yawn | 0 |
| No yawn seed | No yawn | 0 |
| No yawn seed | No yawn | 0 |

In this table the column "Yawn?" is the dummy coded outcome. You can think of the the variable name, "Yawn?", like the question, "Did the participant yawn?" The two responses are " 1 " ( Yes), or " 0" ( No).

Why choose " 0" and " 1 " as the numbers for our categories? Why not choose "5" and "10"; or "2" and "3"? The reason we choose "0" and "1" is that these values give meaning to interpretations of the mean and other summaries.

8. Compute the mean of the Yawn? dummy coded values for the 16 participants in the "No Yawn Seed" condition. Compare that value to your response to Question 5.

9. Re-code the outcome for the 16 participants to create a variable called Didn't yawn?. Use dummy coding to do this. This time participants who didn't yawn would get a "1" and the participants who waned would get a "0". Now compute the mean of the Didn't yawn? dummy coded values for the 16 participants in the "No Yawn Seed" condition. What does this value tell you?

10. Explain how to interpret the mean of a dummy coded variable generally.

## Modeling the Experimental Variation

You will answer the research question by using TinkerPlots™ to conduct a **randomization test** in order to account for experimental variation (variation in the difference of means just because of random assignment) *under the assumption that there is no effect of the yawn seed*.

- Set up a model that will produce the **fixed dummy coded responses** for all 50 of the subjects (use Yawn = 1 and No yawn = 0). You can use a Mixer, but since there are only two values for the response variable, it can be quicker to use a Stacks device. (If you have forgotten how to do this, refer back to the instructions in the *Sleep Deprivation* course activity.)
- Add a linked Stacks device that includes the **condition labels**. (If you have forgotten how to do this, refer back to the instructions in the *Sleep Deprivation* course activity.)
- Run the model.

11. Should the sampling device containing the outcome values be sampled *with* or *without* replacement? What about the sampling device containing the condition labels? Explain why.

### *Plotting and Collecting the Results*

- Use TinkerPlots™ to plot the results for the trial.
- Collect the results from the trial.

## Simulate and Evaluate the Results

- Carry out 500 randomized trials of the simulation in TinkerPlots™.
- Plot the differences in means for the 500 simulated trials.

12. Sketch the plot of the distribution of simulated differences.

13. Based on the hypothesized model, what is the *expected* difference in means? Explain.

14. Compute and report the standard deviation of the differences in means.

15. Using the expected difference in means and the standard deviation, provide a range of likely results under the hypothesized model.

16. How compatible is the observed difference in means with the results produced by the model specified in the null hypothesis? What does this suggest about the answer to the research question? Explain.

### *Quantifying the Likelihood of the Observed Result Given the Model: p-Value*

17. Compute and report the *p*-value for the observed difference in the *Yawn* study. (If you have forgotten how to do this, consult the *p*-value reading.)

18. Interpret the *p*-value you computed.

    The *p*-value of _____ is the probability of …

19. Based on the *p*-value you computed, how compatible is the observed difference in means with the results produced by the model specified in the null hypothesis? What does this suggest about the answer to the research question? Explain.

## Design and Inference

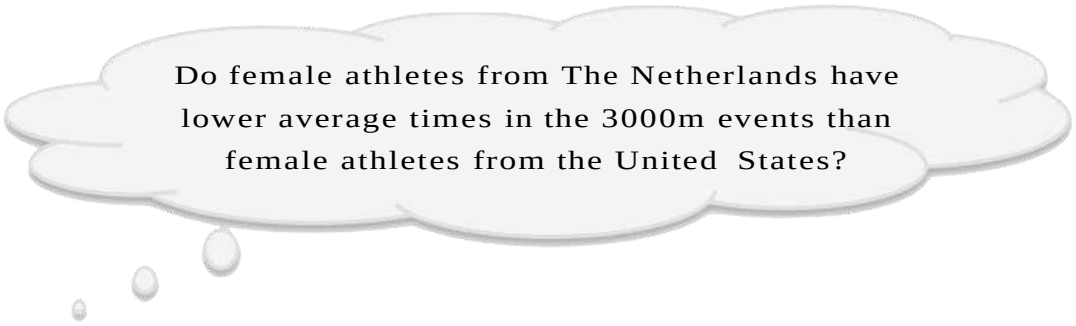
20. How would you rate the level of internal validity evidence based on the study design? Explain.

21. Based on your response to the previous question, are you willing to draw a causal association? If not, offer at least two other possible explanations.

# *Speed Skating*

Athletes from The Netherlands have faired rather well in speed skating events. In fact, Dutch competitors in the Winter Olympics have won 130 medals, 121 in speed skating events. But are the Dutch speed skaters that much better than those from the United States?

Using data from the *SpeedskatingResults. com* database, the times for 15 of the top 100 performances in the Ladies 3000 meter event in 2017 were randomly sampled from each country. You will use these data to answer the following research question:




## Discuss the following questions.

1. Based on the research question, identify each of the groups/conditions for the treatment variable.

2. Based on the research question, identify the response variable.

**Observed Data:** The data in *speed-skating.tp3* contains a random sample of 30 times (in seconds) for the Ladies 3000m event. Fifteen of these times were recorded by Dutch athletes and 15 from athletes from the United States.

3. Identify whether the response variable is categorical or quantitative in nature.

## Examine the Observed Data

- Plot the observed times for both groups of skaters in the same plot. Also compute and display the average time for both groups.

4. Sketch the plot of the observed data.

5. Calculate the difference in the mean times for the observed data (i.e., report the observed result). What does the sample result suggest about the answer to the research question.

## Modeling the Sampling Variation

You will now use TinkerPlots™ to conduct a **bootstrap test** in order to account for sampling variation (variation in the difference of means just because of random sampling) *under the assumption that there is no difference in mean times between Dutch and U.S. speed skaters*.

- Set up a sampling device that includes all 30 **observed responses**. You can copy-and-paste the responses from the observed data into a Mixer, but don't forget to first remove the default elements so that you paste the results into an empty Mixer. Set the Mixer to sample *with replacement*.
- Add a linked Stacks device that includes the **group/condition labels**.
- Run the model.

### *Plotting and Collecting the Results*

- Use TinkerPlots™ to plot the results for the trial.
- Collect the results (difference in means) from the trial.

## Simulate and Evaluate the Results

- Carry out 500 bootstrap trials of the simulation in TinkerPlots™.
- Plot the differences in means for the 500 bootstrap trials.

6. Sketch the plot of the distribution of simulated differences.

## *Quantifying the Likelihood of the Observed Result Given the Model:* p-*Value*

7. Compute and report the *p*-value for the observed difference.

8. Interpret the *p*-value you computed.

   The *p*-value of _____ is the probability of …

9. Based on the *p*-value you computed, how compatible is the observed difference in means with the results produced by the model specified in the null hypothesis? What does this suggest about the answer to the research question? Explain.

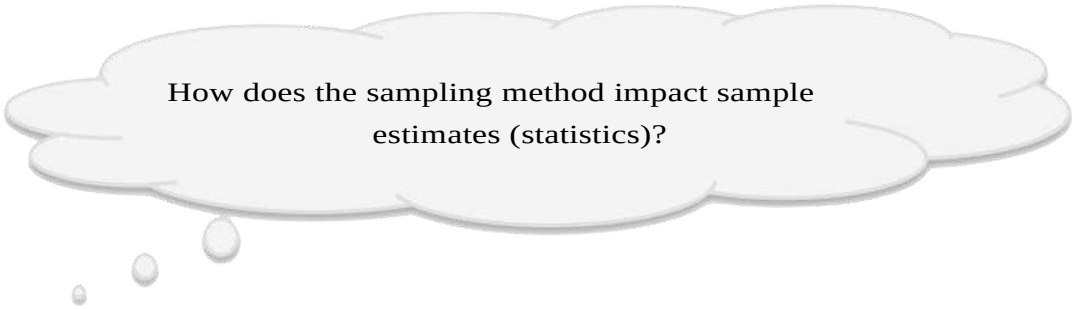
## Design and Inference

10. How would you rate the level of internal validity evidence based on the study design? Explain.

11. Based on your response to the previous question, are you willing to draw a causal association that the faster times posted by Dutch skaters is due to them being from The Netherlands? If not, offer at least two other possible explanations for the difference in performance.

# *Gettysburg Address*

In statistical inference, generalization refers to the process of using sample data to draw conclusions about the larger population from which the sample was drawn. Statisticians are typically concerned with making inferences about some population parameter using a sample statistic (*Remember:* Population summary measures are called **parameters.** Sample estimates of parameters are referred to as **statistics**.) Whether that sample statistic is a statistically good estimate of the population parameter depends on whether the sampling method used is biased. In this activity you will begin by exploring the following question:




To help answer this research question, you are going to compare two different sampling methods using the population of 268 words in the passage on the following page. The passage is, of course, Lincoln's *Gettysburg Address* , given November 19, 1863 on the battlefield near Gettysburg, PA.

Four score and seven years ago, our fathers brought forth upon this continent a new nation: conceived in liberty, and dedicated to the proposition that all men are created equal.

Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battlefield of that war.

We have come to dedicate a portion of that field as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.

But, in a larger sense, we cannot dedicate, we cannot consecrate, we cannot hallow this ground. The brave men, living and dead, who struggled here have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember, what we say here, but it can never forget what they did here.

It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us, that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion, that we here highly resolve that these dead shall not have died in vain, that this nation, under God, shall have a new birth of freedom, and that government of the people, by the people, for the people, shall not perish from the earth.

The goal in many studies is to provide information about some characteristic of a population. For example, you may want to say something about the percentage of Americans who would support a particular piece of legislation. Or, you may want to provide information about the average amount of time University of Minnesota students take to graduate. One potential solution to obtain such information would be to collect the necessary data from every member of the target population.

In many studies, however, it may not be feasible given time and money constraints to collect data from each member of the population. In these cases it is only possible to consider data collected for a smaller subset, or **sample** from that population. In these cases, the characteristic of the population would be estimated from the sample data and inferences would be drawn about the population. The key is then to carefully select the sample so that the results estimated from the sample are representative of the characteristic in the larger population.

> The **population** is the entire collection of who or what (e.g., the observational units) that you would like to draw inferences about. A **sample** is a subset of observational units from the population.

*Circle a sample of ten words in the text of the Gettysburg Address (the population) such that the sample you select is representative (i.e., has the same characteristics) of the population.*

1. Describe how the ten words in your sample are representative of the 268 words in the population.

2. Record the ten sampled words and their lengths:

| Word | Length |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

3. Determine the average (mean) word length for your sample. This sample average (a statistic) is an estimate of the average word length in the population.

*Add your sample estimate to the case table on the instructor's computer.*

4. Sketch the plot of all of the sample estimates. Make sure to label the axis appropriately.

5. The actual population average word length based on all 268 words is 4.3 letters. This is a parameter. Where does this value fall in the above plot? Were most of the sample estimates around the population mean? Explain.

6. For how many groups in your class did the sample estimate **exceed the population average** ? What proportion of the class is this?

7. Based on your answer to the previous question, are the sample estimates just as likely to be above the population average as they are to be below the population average?

When the sampling method produces characteristics of the sample that systematically differ from those characteristics of the population, you say that the **sampling method is biased**. To try to eliminate potential biases, it is better to take a random sample. This should create a representative sample, no matter what variable is focused on. Humans are not very good "random samplers", so it is important to use other techniques to do the sampling for us.

## Simple Random Sampling

A **simple random sample** (SRS) is a specific type of random sample. It gives every observational unit in the population the same chance of being selected. In fact, it gives every sample of size *n* the same chance of being selected. In this example you want every possible subset of ten words that could be sampled to have the same probability of being selected.

The first step in drawing a simple random sample is to obtain a **sampling frame** or a list of each member of the population. Then, you can use software to randomly select a sample from the sampling frame. We have already prepared a sampling frame of the words in the Gettysburg Address for you and saved it in a Tinkerplots™ file.

### *Use TinkerPlots™ to Draw a SRS*

- Open the file *gettysburg.tp3*.
- Draw a simple random sample of ten words from the sampler.

8. Record the ten randomly sampled words and their lengths:

| Word | Length |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

Use TinkerPlots™ to automatically compute the length of each word in your sample. To do this,

- Create a new attribute in the case table called *wordLength*.
- Right-click the attribute name *wordLength* and select the Formula Editor.
- Select stringLength() from the Text functions, and add the sampled words attribute between the parentheses.

9. Use TinkerPlots™ to plot and compute the mean word length for your ten randomly sampled words. Record the mean below.

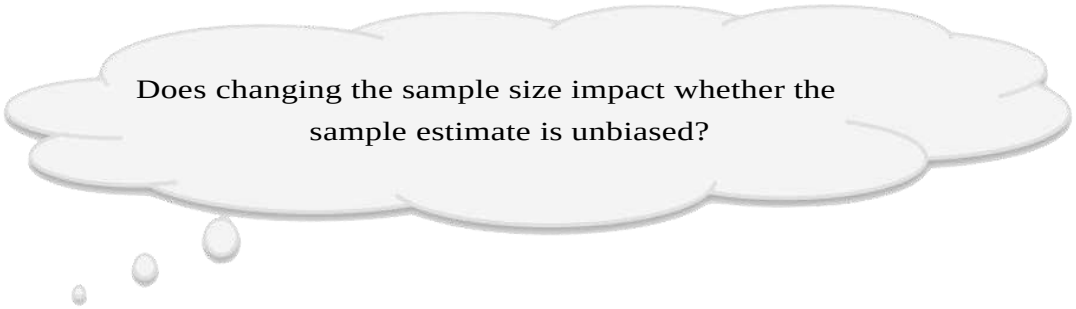
10. Use Collect to carry out 500 trials of this simulation in which you randomly sample ten words and compute their mean length. Sketch the plot of these means. Make sure to label the axis appropriately.

11. If the **sampling method is unbiased** the sample statistics should be centered at the population average word length of 4.3. Does simple random sampling produce an unbiased estimate of the population average?
Explain.

## Sample Size

Even when an unbiased sampling method, such as simple random sampling, is used to select a sample, you do not expect the estimate from each individual sample drawn to match the population average exactly. You should see,
however, that the estimates are just as likely to over- or underestimate the population parameter. Because of this predictability to the variation in the possible sample estimates, inferences drawn about the population are said to be valid.

On the other hand, if the sampling method is biased, any inferences made about the population based on a sample estimate may not be valid. In such cases the estimate of the parameter is more likely to be too large or too small compared to the parameter. It is therefore very important to determine how a sample was selected before believing inferences drawn from sample results.




- Change the sample size from 10 to 25.
- Use TinkerPlots™ to draw 500 random samples of 25 words, and collect the average word length for each sample.

12. Sketch the plot of the sample estimates based on the 500 samples drawn. Make sure to label the axis appropriately.

13. Record the average value for the estimate of the average word length.

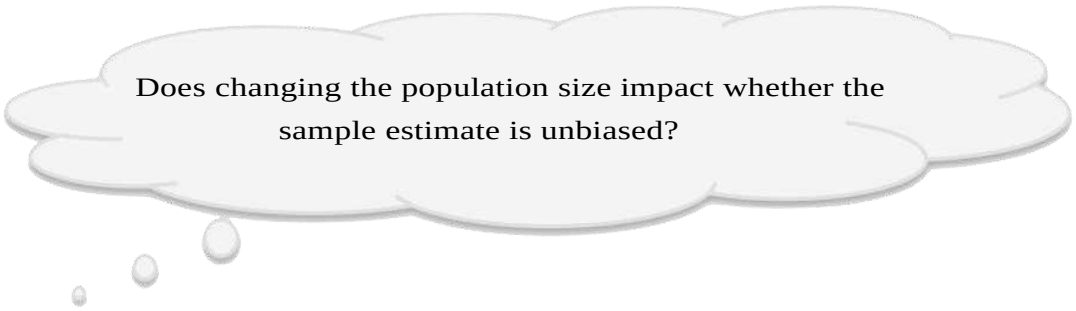
14. Does the sampling method still appear to be unbiased? Explain.

15. Compare and contrast the distribution of sample estimates for $n = 10$ and the distribution of sample estimates for $n = 25$. How are they the same? How are they different?

16. Using the evidence from your simulations, answer the research question: Does changing the sample size impact whether the sample estimates are unbiased?

### *Population Size*

It is clear that changing the size of the sample does not affect whether or not an unbiased estimate is produced. Now we examine another question:




To examine this we will now sample from a population that is quadruple the size of the original population (size = 1072) while keeping the population characteristics the same (e.g., mean word length is still 4.3 letters).

- Open the file *gettysburg-larger-population.tp3*.
- Draw a simple random sample of ten words from the sampler.
- Compute the word length for each randomly sampled word.
- Plot and compute the mean word length for the ten randomly sampled words.
- Collect the mean word length for 500 random samples.

17. Sketch the plot of the sample estimates based on the 500 samples drawn. Make sure to label the axis appropriately.

18. Record the average value for the estimate of the average word length.

19. Does the sampling method still appear to be unbiased? Explain.

20. Compare and contrast the distribution of sample estimates for $n = 10$ now that you are sampling from a larger population to the distribution of sample estimates for $n = 10$ from before. How are they the same? How are they different?

21. Use the evidence collected from the simulation to answer the research question: Does changing the size of the population impact whether the sample estimates are unbiased?

A rather counterintuitive, but very crucial, fact is that when determining whether or not a sample estimate produced is unbiased **the size of the population does not matter!** Even more counterintuitive might be that the precision of the sample estimate is unaffected by the size of the population! (You will learn about the precision of a sample estimate in Unit 5.) This is why organizations like Gallup can state poll results about the entire country based on samples of just 1,000–2,000 respondents as long as those respondents are randomly selected.

In summary, it is important to note some caveats about random sampling:

- One still gets the occasional "unlucky" sample whose results are not close to the population even with large sample sizes.
- Second, the sample size means little if the sampling method is biased. As an example, in 1936 the *Literary Digest* magazine had a huge sample of 2.4 million people, yet their predictions for the Presidential election did not come close to the truth about the population.
- The size of the population does not affect the bias of the estimate, even if a small sample size is used.

# *Murderous Nurse*

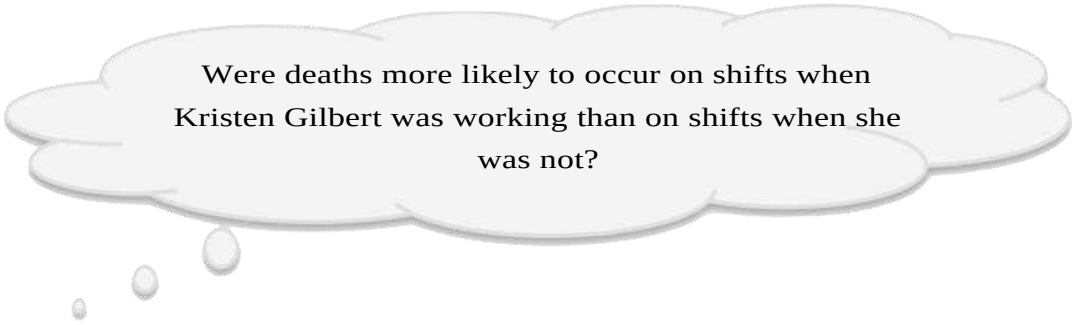
For several years in the 1990s, Kristen Gilbert worked as a nurse in the intensive care unit (ICU) of the Veteran's Administration hospital in Northampton, Massachusetts. Over the course of her time there, other nurses came to suspect that she was killing patients by injecting them with the heart stimulant epinephrine.

Part of the evidence against Gilbert was a statistical analysis of more than one thousand 8-hour shifts during the time Gilbert worked in the ICU[1]. Here are the data presented during her trial:

| | **Gilbert working on shift** | **Gilbert not working on Shift** | **Total** |
|---|---|---|---|
| **Death occurred on Shift** | 40 | 34 | 74 |
| **No death occurred on shift** | 217 | 1350 | 1567 |
| **Total** | 257 | 1384 | 1641 |

[1] Cobb, G. W., & Gehlbach, S. (2006). Statistics in the courtroom: United States vs. Kristen Gilbert. In R. Peck, G. Casella, G. Cobb, R. Hoerl, D. Nolan, R. Starbuck and H. Stern (Eds.), *Statistics: A guide to the unknown* (4th Edition), pp. 3–18. Duxbury: Belmont, CA.

You will use these data to answer the following research question:




## Discuss the Following Questions

1. Among all 1,641 shifts, what percentage of shifts had a death occur?

2. Among the 257 shifts when Gilbert was working, what percentage of shifts had a death occur?

3. Among the 1,384 shifts when Gilbert was not working, what percentage of shifts had a death occur?

4. Compute the difference between the percentage of shifts in which a death occurred when Gilbert was working and the percentage of shifts in which a death occurred when Gilbert was not working.

5. Based on the research question, specify the treatment variable.

6. Based on the research question, specify the response variable. Also, identify whether it is a quantitative or categorical variable.

7. The sample data indicates that shifts that Gilbert worked had a higher percentage of deaths occur than shifts when that she didn't work. Does the difference in percentages convince you that Gilbert was giving lethal injections of epinephrine to patients? Why or why not?

- Read the section *Observational Studies and the Bootstrap Test* from the online book.

## Modeling Sampling Variation

You will conduct a bootstrap test using TinkerPlots™ to find out how likely it would be, assuming there is no difference between the percent of shifts in which a death occurred when Gilbert was working and those in which she was not working.

- Set up a sampling device that includes all 1,641 **observed (dummy) responses** . Code the responses so that 1 = death occurred and 0 = no death occurred.
- Link another sampling device to produce the **fixed group/condition labels**.

8. Should the sampling device containing the outcome values be sampled *with* or *without* replacement? What about the sampling device containing the condition labels? Explain why.

## Simulate and Evaluate the Results

- Use TinkerPlots™ to bootstrap 500 resamples (trials) of the data.
- Collect and plot the results from these trials.

9. Sketch the plot below.

10. Compute and report the *p*-value.

11. Based on the *p*-value you computed, how compatible is the observed difference in means with the results produced by the model specified in the null hypothesis? What does this suggest about the answer to the research question? Explain.

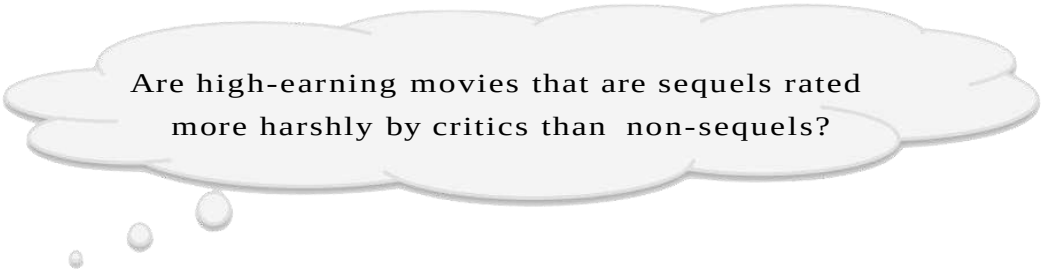
## Design and Inference

12. How would you rate the level of internal validity evidence based on the study design? Explain.

13. Based on your response to the previous question, are you willing to draw a causal association between shifts Gilbert worked and increased death rates? If not, offer at least two other possible explanations for the difference in percentage in the data.

# *Movie Sequels*

As of February 1, 2017, there were 28 major motion pictures in history that earned over $1 billion at the box office, worldwide. Many of these box office winners were sequels to other movies. While these sequels clearly earned a lot of money, many movie sequels are widely panned by critics. In this activity you will examine the following research question:




*Rotten Tomatoes* is a popular service that aggregates critics' reviews of movies into a "Rotten Tomatoes Score". Each of the critics' reviews are rated as positive or negative. The Rotten Tomatoes Score indicates the percentage of positive reviews that the movie received. On the *Rotten Tomatoes* website, movies are given a *Tomatometer* rating based on their score.

Another rating system used by some movie watchers is the *Meg Classification System* (MCS). Any movie having a Rotten Tomatoes Score at or above 80 is classified as "Ripe", while those films having a score below 80 are classified as "Moldy".

**Observed Data:** Of the 28 movies that earned over $1 billion, 17 were given a MCS rating of " Ripe". Seven of the 11 non-sequels received an MCS rating of "Ripe".

## Explore the Observed Data

1. Organize the observed data (i.e., frequencies) into a 2x2 contingency table.

2. Compute and report: (a) the percentage of movie sequels that were rated as "Ripe"; ( b) the percentage of non-sequels that were rated as "Ripe", and (c) the observed difference in percentages.

3. Write a few sentences summarizing the results in the sample. This should include a summary of what the data suggest about: (1) the overall percentage of movies rated as "Ripe"; (2) the differences between the two groups of movies; and (3) whether or not the data appear to support the claim that sequels are rated more harshly by critics than non-sequels.

## Model the Chance Variation

Consider a sampler that you could simulate from in order to explore the chance variation that would be expected in the difference of percentages if there really was no difference in how harshly critics reviewed sequels and non-sequels.

4. Based on the study design, should the sampler model experimental variation or sampling variation? Explain.

5. In the space below draw a picture of your sampler that you will use to generate outcomes. Be sure to clearly indicate whether each device in your sampler is sampling with or without replacement.

## Simulate and Evaluate the Results

- Use TinkerPlots™ to carry out 500 trials of the simulation.
- Collect and plot the results from these trials.

6. Sketch the plot of the distribution of simulated differences into your word-processed document. Also give the *expected mean* based on the model and compute and report the standard deviation.

7. Compute and report the *p*-value based on the observed result.

8. Based on the *p*-value you computed, how compatible is the observed difference in means with the results produced by the model specified in the null hypothesis? What does this suggest about the answer to the research question? Explain.

## Design and Inference

9. How would you rate the level of internal validity evidence based on the study design? Explain.

10. Based on your response to the previous question, are you willing to draw a causal association between whether a movie is a sequel or not and whether it receives a positive review? If not, offer at least two other possible explanations for the difference in percentage in the data.

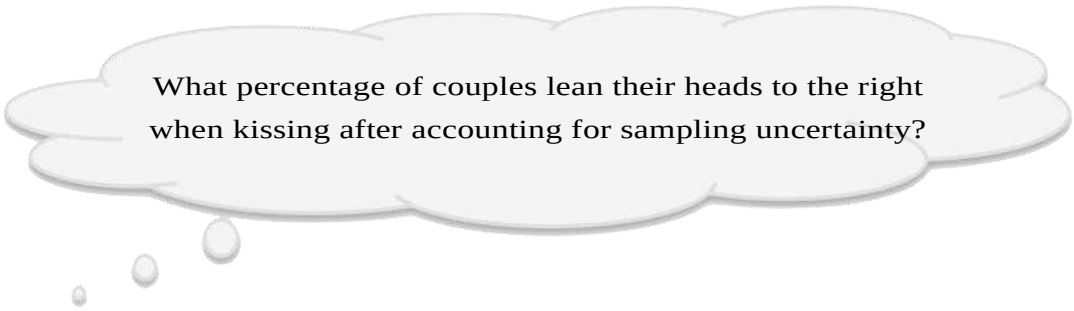
11. How would you rate the level of external validity evidence based on the study design? Explain.

12. Based on your response to the previous question, are you willing to draw a generalization about all movies, even those that didn't earn $1 billion? Explain.

# *Kissing the 'Right' Way*

A German bio-psychologist, Onur Güntürkün, was curious whether the human tendency for right-sightedness (e.g., right-handed, right-footed, right-eyed), manifested itself in other situations as well. In trying to understand why human brains function asymmetrically, with each side controlling different abilities, he investigated whether kissing couples were more likely to lean their heads to the right than to the left[1] . He and his researchers observed 124 couples (estimated ages 13 to 70 years, not holding any other objects like luggage that might influence their behavior) in public places such as airports, train stations, beaches, and parks in the United States, Germany, and Turkey.

In this activity, you will be exploring the following research question:




**Observed Data:** Of the 124 couples observed, 80 leaned their heads to the right when kissing.

[1] Güntürkün, O. (2003). Human behaviour: Adult persistence of head-turning asymmetry. *Nature, 421*, 711.

## Discuss the following questions.

1. Based on only the observed data, answer the research question.

2. Consider if Güntürkün had observed a **different sample of 124 couples**. Would the data for these couples provide the same estimate of the percentage of couples who lean their heads to the right when kissing? Explain.

## Modeling Sampling Variation

if Güntürkün had observed a different sample of 124 couples the one-number best guess for the percentage of couples lean their heads to the right when kissing would differ (i.e., there is sampling variation). Because of sampling
variation, when answering research questions like Güntürkün's that ask for an estimate, it is important that we acknowledge that there is uncertainty in the estimate we provide because we know that sample estimates will vary from sample to sample.

## *Bootstrapping*

To model the sampling uncertainty, you will bootstrap from the observed data. Bootstrapping to estimate the sampling uncertainty works whether you have one sample of data or two (as in Unit 4). To carry out a nonparametric bootstrap analysis using TinkerPlots™, you:

- Set up a sampling device to bootstrap the observed data. You can either dummy code these data, or leave them as categorical labels.
- Be sure that you set the device to sample with replacement. The Repeat value should be set to the same sample size as the observed data. ( You want to replicate the sampling of 124 couples.)

3. Carry out 500 bootstrap trials, each time collecting the percentage of people who turn their heads to the right when kissing. Plot the results from the 500 trials and sketch the plot below. Make sure to label the axis. This distribution is referred to as the *bootstrap distribution*.

## *The Bootstrap Distribution*

4. Find the mean of the bootstrap distribution. Explain why you could expect the bootstrap distribution to be centered at this value by referring to the model from your TinkerPlots™ sampler.

5. Compute the standard deviation of the bootstrapped percentages.

> The standard deviation of a plot of results (e.g., means or proportions) is referred to as the **standard error (SE)** . To compute the SE using TinkerPlots™ continue to use the stdDev() function that you have been using. When you report that value from now on, you just will refer to it using its technical name, *standard error*

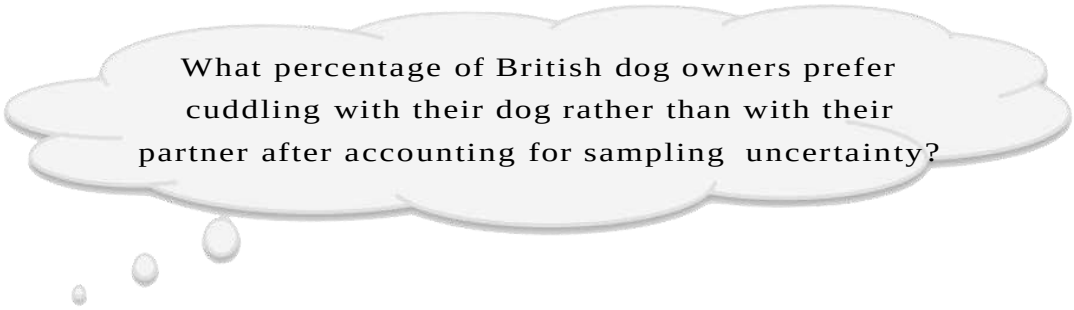
## Margin of Error and Compatibility Interval

6. Use the standard error from the bootstrap simulation to compute the margin of error.

7. Compute ( by hand) the compatibility interval for the percentage of couples that lean to the right when kissing. Use the compatibility interval to provide an answer to the research question.

# *Cuddling Preferences*

A recent poll of British pet owners shed light on the extent to which pet owners like to cuddle with their pets. In many cases, pet owners appear to prefer cuddling with their pets more than with their partners. In this activity, you will be exploring the following research question:




## Examine the Observed Data

1. Use the data in the file *british-cuddle-20.tp3* to provide an answer to the research question based on the observed data.

## Bootstrapping a Compatibility Interval

2. Carry out 500 bootstrap trials. Plot the results from the 500 trials and sketch the plot below. Make sure to label the axis.

## *Evaluating the Bootstrap Distribution*

3. Find the mean of the bootstrap distribution. Explain why you could expect the bootstrap distribution to be centered at this value by referring to the model from your TinkerPlots™ sampler.

4. Compute the standard error (use the `stdDev()` function) based on this simulation.

5. Using the standard error, compute the margin of error.

6. Compute the compatibility interval for the percentage of British dog owners who prefer to cuddle with their dog rather than their partner.

## Exploring the Effect of Sample Size

The compatibility intervals you computed were both based on a sample size of 20 British dog owners. What happens to the uncertainty in the compatibility interval if you have a different sample size? You will explore this by computing compatibility intervals using two other sample sizes.

7. Fill in the first row of the table below with the information from the compatibility interval for British dog owners based on the sample size of 20.

| Sample Size | Observed Percentage | Standard Error | Compatibility Interval |
|---|---|---|---|
| | | | |
| | | | |
| | | | |

8. Open the file *british-cuddle-100.tp3*. Use these data to provide an answer to the research question: What percentage of British dog owners prefer cuddling with their dog rather than with their partner after accounting for sampling uncertainty? To do this, estimate the percentage from the observed data, and then carry out 500 bootstrap trials to estimate the uncertainty in the estimate. Fill in the information from this analysis in the second row of the table.

9. Open the file *british-cuddle-500.tp3.* Use these data to provide an answer to the research question: What percentage of British pet owners prefer cuddling with their dog rather than with their partner after accounting for sampling uncertainty? To do this, estimate the percentage from the observed data, and then carry out 500 bootstrap trials to estimate the uncertainty in the estimate. Fill in the information from this analysis in the second row of the table.

10. Use the information in the table to explain the relationship between sample size and the uncertainty expressed in the compatibility interval.

11. Why do you think that sample size and uncertainty are related in this way? Explain.

# *Minnesota College Debt*

The Department of Education recently released its College Scorecard data. One of the variables in this data is the typical amount of loan debt accumulated at the institution by student borrowers. In this activity, you will be exploring the following research question:

> What is the average amount of loan debt accumulated by student borrowers who attend public colleges/universities in Minnesota after accounting for sampling uncertainty?

To answer this question, you will use the data in the file *mn-colleges.tp3*. This data set contains a sample of 25 colleges/universities randomly selected from the larger population of all public colleges/universities in Minnesota. The variable *debt* provides the average student loan debt for students who attend that college/university.

## Examine the Observed Data

1. Plot the sample debt data using TinkerPlots™. Describe the sample distribution. Be sure to describe the shape of the distribution and give a measure of center and variability.

## Bootstrapping a Compatibility Interval

You can also carry out a bootstrap simulation to estimate the standard error when you have quantitative data.

2. Carry out 500 bootstrap trials. Plot the results from the 500 trials and sketch the plot below. Make sure to label the axis.

### *Evaluating the Bootstrap Distribution*

3. Compute the standard error (use the `stdDev()` function) based on this simulation.

4. Using the standard error, compute the margin of error.

5. Compute the compatibility interval for the average amount of loan debt for students who attend public college/university in Minnesota.

### Design and Inference

6. Based on the validity evidence for this study, is the compatibility interval an unbiased estimate for the average amount of loan debt for ALL students who attend college/university in Minnesota? Explain.

## Exploring the Effect of Sample Size

The compatibility intervals you computed were both based on a sample size of 25 colleges and universities. What happens to the uncertainty in the compatibility interval if you have a different sample size? You will explore this by computing compatibility intervals using one other sample size.

7. Fill in the first row of the table below with the information from the compatibility interval based on the sample size of 25.

| Sample Size | Observed Percentage | Standard Error | Compatibility Interval |
|---|---|---|---|
| | | | |
| | | | |
| | | | |

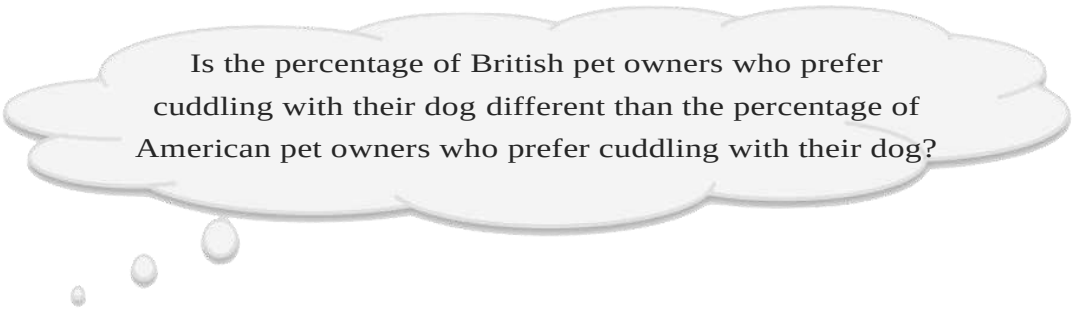
8. Consider a second sample of Minnesota colleges/universities that had the same sample mean debt as the data in the *mn-colleges.tp3* file. However, this second sample is twice as large; it includes 50 observations. Use these data to provide an answer to the research question: What is the average amount of loan debt accumulated by student borrowers who attend public colleges/universities in Minnesota after accounting for sampling uncertainty? To do this, estimate the percentage from the observed data, and then carry out 500 bootstrap trials to estimate the uncertainty in the estimate. Fill in the information from this analysis in the second row of the table.

9. How does the uncertainty in the compatibility interval from this second sample compare to the uncertainty in the compatibility interval you computed in Question #5? Explain.

10. ### Some question about sample size not fixing shitty validity. ### Based on the validity evidence for this study, is the compatibility interval an unbiased estimate for the average amount of loan debt for ALL students who attend college/university in Minnesota? Explain.

# *Comparing Cuddling Preferences*

A recent poll of British and American pet owners shed light on the extent to which pet owners like to cuddle with their pets. In many cases, pet owners appear to prefer cuddling with their pets more than with their partners. In this activity, you will be exploring the following research question:



## Bootstrap Test

To answer this question, you will carry out a bootstrap test using the data from the files *american-cuddle-20.tp3* and *british-cuddle-20.tp3*. Test the statistical hypothesis that there is no difference between the percentage of American dog owners who prefer to cuddle with their dogs rather than their partners and the percentage for British dog owners.

1. Carry out 500 trials of the bootstrap test assuming no differences between the percentage of British and American dog owners who prefer to cuddle with their dog rather than their partner. Sketch the distribution of bootstrapped differences.

2. Compute the sample estimates for: (1) the percentage of American dog owners who prefer to cuddle with their dog rather than their partner, (2) the percentage of British dog owners who prefer to cuddle with their dog rather than their partner, and (3) the difference between the two percentages.

3. Compute and report the $p$-value based on the observed result.

4. Based on the $p$-value you computed, how compatible is the observed difference in means with the results produced by the model specified by the statistical hypothesis? What does this suggest about the answer to the research question? Explain.

## Compatibility Intervals

Compatibility intervals can also be used to evaluate whether there are **statistical differences** between two groups.

5. Use the data in the file *british-cuddle-20.tp3* to provide an answer to the research question: What percentage of American dog owners prefer cuddling with their dog rather than with their partner after accounting for sampling uncertainty? To do this, estimate the percentage from the observed data, and then carry out 500 bootstrap trials to estimate the uncertainty in the estimate. Plot the results from the 500 trials and sketch the plot below. Make sure to label the axis.

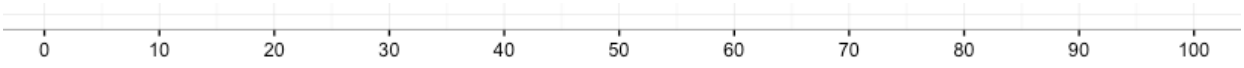
6. Find the mean of the bootstrap distribution. Explain why you could expect the bootstrap distribution to be centered at this value by referring to the model from your TinkerPlots™ sampler.

7. Compute the standard error (use the `stdDev()` function) based on this simulation.

8. Using the standard error, compute the margin of error.

9. Compute the compatibility interval for the percentage of British dog owners who prefer to cuddle with their dog rather than their partner.

10. Draw the range of the compatibility interval for British dog owners who prefer cuddling with their dog using the axis below.



11. Use the data from the file *american-cuddle-20.tp3* to provide an answer to the research question: What percentage of American dog owners prefer cuddling with their dog rather than with their partner after accounting for sampling uncertainty? To do this, estimate the percentage from the observed data, and then carry out 500 bootstrap trials to estimate the uncertainty in the estimate. Fill in the information from this analysis in the table.

| Sample Size | Observed Percentage | Standard Error | Compatibility Interval |
|---|---|---|---|
| | | | |

12. Draw the range of the compatibility interval for American dog owners on the axis in Question #10.

Remember that the compatibility interval for British dog owners gives the percentage of British dog owners who prefer cuddling with their dogs after accounting for sampling uncertainty. Similarly the compatibility interval for American dog owners gives the percentage of American dog owners who prefer cuddling with their dogs after accounting for sampling uncertainty. If both intervals include some of the same values (the intervals overlap), it provides evidence that the two groups could have the same level of preference (i.e., evidence of no difference)...or at least the uncertainty is too great for us to differentiate which group has the higher percentage.

13. Do the two intervals you drew in Question #10 overlap each other?

14. Explain using your drawing whether there is evidence that the percentage of British dog owners who prefer cuddling with their dog is higher than the percentage of American dog owners who prefer cuddling with their dog.

## Effect of Sample Size: n = 200 (100 per Group)

How would our results change if we had used a different sample size?

### *Bootstrap Test*

Carry out a bootstrap test using the data from the files *american- cuddle-100.tp3* and *british-cuddle-100.tp3*.

15. Carry out 500 trials of the bootstrap test assuming no differences between the percentage of American and British dog owners who prefer to cuddle with their dog rather than their partner. Sketch the distribution of bootstrapped differences.

16. Compute the sample estimates for: (1) the percentage of American dog owners who prefer to cuddle with their dog rather than their partner, (2) the percentage of British dog owners who prefer to cuddle with their dog rather than their partner, and (3) the difference between the two percentages.

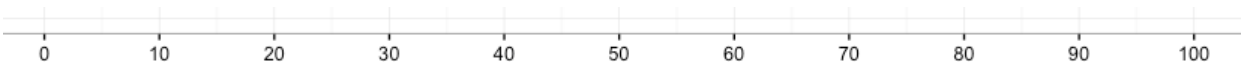
17. Compute and report the $p$-value based on the observed result.

18. Based on the *p*-value you computed, how compatible is the observed difference in means with the results produced by the model specified in the null hypothesis? What does this suggest about the answer to the research question? Explain.

19. How does the *p*-value for the bootstrap test based on a sample size of 40 (20 in each group) compare to the *p*-value for the bootstrap test based on a sample size of 200 (100 in each group)?

## *Compatibility Intervals*

20. Use the data from the file *british-cuddle-100.tp3* to provide an answer to the research question: What percentage of British dog owners prefer cuddling with their dog rather than with their partner? To do this, estimate the percentage from the observed data, and then carry out 500 bootstrap trials to estimate the uncertainty in the estimate. Draw the range of this compatibility compatibility interval using the axis below.



21. Use the data from the file *american-cuddle-100.tp3* to provide an answer to the research question: What percentage of American dog owners prefer cuddling with their dog rather than with their partner? To do this, estimate the percentage from the observed data, and then carry out 500 bootstrap trials to estimate the uncertainty in the estimate. Fill in the information from this analysis in the table.

| Sample Size | Observed Percentage | Standard Error | Compatibility Interval |
|---|---|---|---|
| | | | |

22. Draw the range of the compatibility interval for American dog owners on the axis in Question #20.

23. Based on whether or not the two intervals you drew in Question #20 overlap each other, explain whether there is evidence that the percentage of British dog owners who prefer cuddling with their dog is higher than the percentage of American dog owners who prefer cuddling with their dog.

24. How is the result of the bootstrap test (i.e., the *p*-value) effected by sample size? Explain by comparing the *p*-values for the three bootstrap tests.

25. How does sample size affect statistical uncertainty (i.e., the range of the compatibility interval)?

26. Based on your answer to the previous question, how is statistical uncertainty related to *p*-value?