Activity 6

kissing the “right” way

# directions

Have one group member upload this docx file to your Google Drive and share it with your groupmates and your instructor. Name your document “**GroupX\_Activity6”** (where X is your group number). Work together to type up your responses to each question. Download your document as a PDF and submit this to Canvas individually.

# introduction

A Germen bio-psychologist, Onur Gunturkun, 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. He hypothesized couples were more likely to lean their heads to the right.

**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, **of which 80 couples leaned their heads to the right when kissing**. In this activity, you will be exploring the following research question:***What percentage of all couples lean their heads to the right when kissing?***

1. What percentage of couples **from this** **study** leaned their heads to the right?
2. Is the value you computed above the (highlight the best choice):

* Population proportion/percentage (A population parameter)

* Observed proportion/percentage (A sample statistic)

1. Based only on the observed data, give your best answer to the research question. Explain your reasoning.

1. Do you think that the answer you gave above is 100% accurate? Explain.

1. Suppose Gunturken had **observed a different sample of 124 couples**. What would you expect the data for this new sample of couples to be like? Highlight the best response.

* The ***exact same*** estimate (i.e., same percent) of observed couples who lean their heads to the right when kissing.

* An ***extremely*** ***different*** estimate (i.e., extremely different percent) of observed couples who lean their heads to the right when kissing.

* A ***different*** ***estimate but likely somewhat close to the original sample estimate*** of couples who lean their heads to the right when kissing.

* Any of the above is equally likely.

1. Explain the reason for the choice you made above.
2. To understand how much our sample percentage might vary from the **population parameter**, suppose this study was replicated 100 times. That is, under the same conditions as the original study, researchers observed 124 couples kissing and recorded whether the couples leaned their heads to the right or to the left. After each replication/trial of the study, the **percentage of couples that leaned their heads to the right** from that trial was recorded. Below is a plot of the 100 trials.

A screen shot of a graph

Description automatically generated

* Interpret a dot in this plot.

* Discuss the center, shape and spread (or variability) of the plot.

* Based on **a single sample** of 124 couples, do you think it is reasonable to give a single value estimate for the percentage of couples that lean their heads to the right? If so, justify why you think a single value is reasonable. If not, what might you give instead of a single value as an estimate? Explain.

1. The sampling distribution given previously was generated under the **assumption** that the **true percentage (parameter) of all couples** who lean their head to the right is **68%**. (IMPORTANT: the key word in the previous sentence is **assumption**. This is not a true figure, but we assumed it for the sake of this exercise.) How would you set up a TinkerPlots sampler to get such a sampling distribution? That is, how would you set up the sampler to model the act of observing new samples of 124 couples kissing and recording which way they leaned their heads? Be sure to do the following:

* Copy and paste your sampler below.

* Describe repeat and draw in the context of the problem.

* Are you setting the sampler to with or without replacement? Explain.

* What are you calling your attribute(s) and why?

* What are the elements in your sampler? Explain.

* Describe what happens when the sampler is run.

# model evaluation

In constructing the sampler for problem (8) above, you had to make an assumption that the population parameter (that is, the true percentage of all couples that lean their head to the right when kissing) was 68%. But the reality in most statistical situations is that **we do not know the population parameter** (and, in most cases, the population parameter is an unknowable value). All we typically have available to use is data from a single sample, as researchers use all their resources (time, money, etc.) collecting the appropriate amount of data just once; it is not feasible to collect repeated samples in the real world. (And if you could take repeated samples, it makes more sense to just combine all those samples into one large sample!) Statisticians need to be able to create models to simulate feasible/plausible data so that they can make accurate inferences and know how samples might vary depending on the people they picked in the sample. And they need to do this using only the sample data that was available to them!

In the spirit of constructing statistical models to answer this research question, using only the data available, below are 5 possible TinkerPlots models that other statistics 243 students have created to model Gunturken’s study (some are valid and others are not). **For each model below, describe what is being modeled, whether it is a valid model for this situation, and why you think it is/isn’t**. To aid you in answering this question for each model, you may consider some/all of the following ideas in your discussion:

* Discuss the values of **repeat** and **draw** and how they relate to the context of the study.

* Discuss whether the model is using **with or without replacement** and explain what impact that has on the model and simulation.

* **Discuss what happens when the model is run and describe what you see in the outputs shown** (each model shows three plots that were produced from running the model three times). Be as specific as possible.

* **Discuss assumptions implicit in the model**. **Be specific and as detailed as possible**.

Model A   
A screenshot of a computer

Description automatically generated   
A screenshot of a computer

Description automatically generated  
A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Describe what is being modeled in Model A, whether it is a valid model for this situation, and why you think it is/isn’t. See the bullets previously mentioned for ideas on what to discuss.

Model B

A screenshot of a computer

Description automatically generated  
A screenshot of a computer

Description automatically generated  
A screenshot of a computer

Description automatically generated  
A screenshot of a computer

Description automatically generated

1. Describe what is being modeled in Model B, whether it is a valid model for this situation, and why you think it is/isn’t. See the bullets previously mentioned for ideas on what to discuss.

Model C

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated  
A screenshot of a computer

Description automatically generated  
  
A screenshot of a computer

Description automatically generated

1. Describe what is being modeled in Model C, whether it is a valid model for this situation, and why you think it is/isn’t. See the bullets previously mentioned for ideas on what to discuss.

Model D   
  
A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated  
A screenshot of a computer

Description automatically generated  
A screenshot of a computer

Description automatically generated

[Note: the sampler above has 44 “left” and 80 “right” balls inside the mixer.]

1. Describe what is being modeled in Model D, whether it is a valid model for this situation, and why you think it is/isn’t. See the bullets previously mentioned for ideas on what to discuss.

Model E

A screenshot of a computer

Description automatically generated  
A screenshot of a computer

Description automatically generated  
A screenshot of a computer

Description automatically generated  
A screenshot of a computer

Description automatically generated

1. Describe what is being modeled in Model E, whether it is a valid model for this situation, and why you think it is/isn’t. See the bullets previously mentioned for ideas on what to discuss.
2. If you were going to select one of the six models above to answer the research question in the Gunturken study, which model would you select and why?

# BOOTSTRAPPING

In statistics, bootstrapping is a technique for developing valid and statistically sound inferences from sample data. Without any other evidence as to what the true model is (and we almost never really know the population parameter nor do we usually get to replicate the study many times), **the most informed choice is to** ***use the observed data as a stand-in****,* ***or proxy, to model the unknown population*** (assuming the data was collected using appropriate sampling techniques). Without any other additional information (or help) you use the sample data to pull yourself up by the bootstraps and simulate additional data to help make a statistical inference.

In the Gunturken study, the stand-in model (that is, the model created using the observed data of 80 couples leaning to the right out of 124 couples) can be used to generate simulated data that represents many samples of 124 couples. ***We can use the observed sample data as representing the population on a smaller scale, thus creating a model from which to sample from***. As you have witnessed throughout the course, a TinkerPlots sampler generates many different samples of data.

1. Use the model you selected from Question 14 above. Construct that model in TinkerPlots. Then answer the questions below:

* Are you sampling with or without replacement? Explain.

* Describe what will happen when you click run on the sampler. How does the actions of the sampler relate to the context?

* Take one sample. Plot your results (copy and paste the plot below). What did you find for the percentage of couples who leaned right when kissing from that sample?

* What will you collect statistics on? Explain.

1. Carry out 500 trials of the simulation. Create a dot plot of your results and paste it below. **Interpret what a dot represents in the plot.**

1. Where is the distribution centered? Explain why that makes sense.

1. Based on your 500 trials, how might you estimate the percentage of *all* couples who lean their heads to the right to be?

1. Use the divider tool in TinkerPlots to create an **interval of likely values that estimates this percentage**, and copy that plot below. Explain your choice of interval.