I'm sad… Let's swim with some dolphins!

Swimming with dolphins can certainly be fun, but is it also therapeutic for patients suffering from clinical depression? To investigate this possibility, researchers recruited 30 subjects aged 18-65 with a clinical diagnosis of mild to moderate depression (Antonioli and Reveley, 2005). Subjects were required to discontinue use of any antidepressant drugs or psychotherapy for four weeks prior to the experiment, and throughout the experiment. These 30 subjects went to an island off the coast of Honduras, where they were randomly assigned to one of two treatment groups. Both groups engaged in the same amount of swimming and snorkeling each day, but one group did so in the presence of bottlenose dolphins and the other group did not. At the end of two weeks, each subjects' level of depression was evaluated, as it had been at the beginning of the study, and it was determined whether they showed "substantial improvement" (reducing their level of depression) by the end of the study. [From Rossman, Allan J. 2008. "Reasoning about Informal Statistical Inference: One Statistician's View." *Statistics Education Research Journal* 7 (2): 5–19.]

1. For this study:
   There were 30 participants, and 13 of the 30 participants showed substantial improvement. How many of these 13 improvers do you think were in the "Dolphin Group?" Briefly explain your reasoning.

2. Most students' answers to question 1 begin by making a guess about the situation under study. Perhaps something like, "Well if it's true that [some statement] then …."
   In statistical situations, we call this guess (or statement) a hypothesis. Any hypothesis might end up being correct or incorrect (true or false). Below, write out all of the possible hypotheses we could make for this study.

3. For this study:
    a. Identify the treatment and the response variable (HINT: What treatment are the researchers providing to some (but not all) of the participants in this study, and what response are the researchers measuring to determine if the treatment is effective?).

    b. In statistics, the hypothesis that says the treatment does not have an effect on the response variable is called the Null Hypothesis (symbolized by H_0). Using words, state what H_0 is in this study.

    c. Recall that 13 of the 30 participants in this study showed substantial improvement. If the null hypothesis is true, how many of the 13 improvers would you expect to be in the dolphin group? What would the difference between the number of improvers in the 2 groups be (dolphin minus control)?

4. In the actual study, Antonioli and Reveley found that _____ of 15 subjects in the dolphin therapy group showed substantial improvement, compared to _____ of 15 subjects in the non-dolphin (or the control) group. Complete the table based on these results.

	Dolphin therapy	Control group	Total
Showed substantial improvement			13
Did not show substantial improvement			17
Total	15	15	30

    a. If the null hypothesis (H_0) is true, we would expect the number of improvers in each group to be close to one another. We can measure how close these numbers are to one another by subtracting them (or finding their difference). For this study's results, what is the difference between the number of improvers (dolphin minus control) in the two different groups?

    b. What is your initial reaction to this result? Do you find the result surprising? What do you think this result says about H_0?

5. In this phase of the task, let's simulate the dolphin study assuming the null hypothesis (H_0) is true. Recall, the null says that improved depression levels does not depend on whether or not people swim with dolphins. Think of how you might use randomness to simulate the way that 13 improvers and 17 non-improvers are distributed among the dolphin and control groups. In this kind of simulation, randomness alone determines which group the improvers are in - not whether the participants swam with dolphins. We have dice, decks of cards, colored chips, spinners, etc. for you to use in your simulation.
    a. As a group, determine a plan for simulating the study. In your simulation plan, how are you representing the participants? How are you representing the improvers and the non-improvers? How will you decide who is in the dolphin and control groups? How are you recording the results? Describe your plan below.
    [Get the approval of your teacher ***before*** conducting your simulation.]

    b. After approval, each student in your group should use your group's simulation plan to fill out the table below:

	Dolphin therapy	Control group	Total
Showed substantial improvement			13
Did not show substantial improvement			17
Total	15	15	30

What is the difference in the number of improvers (dolphin minus control) for your simulation?

Make a dotplot on the whiteboard that reports the difference in the number of improvers (dolphin minus control) for every person's simulation in our class.

    c. Are the results of your simulations *as extreme* as the actual results that Antonioli and Reveley obtained (i.e. a 10 - 3 split)? How many times did results this extreme happen in our entire class? What can we conclude (if anything) from Antonioli and Reveley's dolphin study?

d. In our class, we only simulated a small number of times. We can use technology to simulate thousands of times. Let's do it now. We will go to http://lock5stat.com/statkey/ and access the Test for Difference in Proportions tool. Using this tool, we will simulate allowing randomness to place our improvers and non-improvers into the Dolphin and control groups. Based on this simulation, does it seem reasonable to conclude that randomness alone accounts for 10 - 3 split (a difference of 7) in the Antonioli and Reveley study? Briefly explain your reasoning.

6. Consider the following terminology and answer the questions below:
Terminology:
   - The ***null hypothesis*** is a statement about the value of a population measure that we are assuming to be true when we do our simulation.
   - The ***probability*** of an event is the long-run proportion of times an event of interest happens when a simulation of that event's experiment is repeated many, many times.
   - The ***p-value*** is the probability that randomness alone will produce results as extreme (or more extreme) as the results in an actual study, assuming that the null hypothesis is true.
   - A small p-value indicates that if the null hypothesis were true, the observed data in a study is unlikely (or surprising) if explained by randomness alone. Such a result is said to be ***statistically significant***, and provides evidence that the null is probably not true.

   a. Based on the simulations in question 5, what conclusion should the researchers draw regarding whether or not Dolphin therapy helps improve depression levels? Summarize your conclusions and use the above terminology in your explanation (use the back if necessary).

b. Are you willing to draw a cause-and-effect conclusion about dolphin therapy and depression based on these results? Justify your answer based on the design of the study.

c. Are you willing to generalize these conclusions to all people who suffer from depression? How about to all people with mild to moderate depression in this age range? Justify your answer based on the design of the study.

d. If Antonioli and Reveley had instead found that 8 of the 13 improvers were in the dolphin group, then what decision should they make? Justify your conclusions and use the above terminology in your justification.