1. I recently read an article in Science magazine that described an experiment concerning people’s willingness to be organ donors. All of the subjects in the experiment were told that to imagine that they have moved to a new state and have applied for a driver’s license, and they must make a decision about whether to become an organ donor. Some of the subjects were randomly assigned to be told that the default option is not to be a donor, and the rest of the subjects were told that the default option is to be a donor. All subjects were then given the choice of whether or not to become an organ donor. The researchers suspected that a higher proportion of people are willing to be donors when the default option is to be a donor.

   a) Identify the explanatory and response variables in this study.

   b) State (in symbols) the appropriate null and alternative hypotheses (in symbols) for testing the researchers’ suspicion.

   c) The article reported that 42% of the subjects in the “default is not to be a donor” group decided to become a donor, compared to 82% in the “default is to be a donor” group. What further information would you need in order to conduct the test (i.e., to calculate a test statistic and p-value)?

2. The following output comes from analyzing data collected by a naturalist named Bumpus on the lengths (in millimeters) of sparrows, some of which had survived a severe winter storm and some of which had perished.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>lengths (died)</td>
<td>24</td>
<td>162.00</td>
<td>2.41</td>
<td>0.49</td>
</tr>
<tr>
<td>lengths (survived)</td>
<td>35</td>
<td>159.06</td>
<td>2.81</td>
<td>0.47</td>
</tr>
</tbody>
</table>

   Difference = \( \mu \) (lengths (died)) - \( \mu \) (lengths (survived))

   95% CI for difference: (1.57150, 4.31421)

   a) Write out the formula (with numbers plugged in) that would have produced the confidence interval reported here. (Do not bother to work out the calculations, though; just show the formula with the numbers plugged in.)

   b) Interpret the confidence interval given here. Be sure to address whether the sample data provide evidence that the population mean length of sparrows who survived is different from that of sparrows who perished. If there is evidence of a difference, also address which type (those that survived or those that perished) tends to be longer.

   c) Is it reasonable to conclude from this study that the sparrow’s length caused it to die or survive? Explain briefly.

3. In April of 2006, the National Cancer Institute released a report about a study that compared two drugs (raloxifene and tamoxifen) intended to reduce a woman’s risk of developing breast cancer. Women who participated in the study were randomly assigned to take one of these drugs on a daily basis. It turned out that 167 of the 9745 women who took raloxifene developed breast cancer, compared to 163 of the 9726 women who took tamoxifen. (For simplicity, refer to the
drugs as R and T.) Conduct a hypothesis test of whether these proportions differ significantly. Include the following components:

a) null and alternative hypotheses

b) test statistic

c) p-value

d) test decision at $\alpha = .05$ level

e) check of technical conditions

f) summary of conclusion in context

4. I read an article about a study conducted at Ohio State University, which found that students who have a facebook account have significantly lower grade point averages than students who do not have a facebook account.

a) Identify the explanatory and response variables in this study. Also classify each variable as categorical or quantitative.

b) Is this an observational study or an experiment? Explain briefly.

c) Is it reasonable to conclude that having a facebook account causes students to have lower grade point averages? Explain why or why not.

Now suppose that you want to collect data on Cal Poly students to investigate whether the same phenomenon holds true here.

d) Which statistical test would you use to determine whether the difference between the groups is statistically significant? Circle your answer; do not bother to explain.

5. Do people lie more with email than with pencil-and-paper? A study reported at the August 2008 meeting of the Academy of Management involved 48 graduate students in business who participated in a “bargaining” game. The response variable of interest was whether the person misrepresented (lied about) the size of the pot when negotiating with another player. Some of the participants were randomly assigned to use email for their communication, while others used paper-and-pencil. It turned out that 24 of 26 who used email were guilty of lying about the pot size, compared to 14 of 22 who used paper-and-pencil.

A simulation analysis was used to investigate whether the difference between these groups is statistically significant. The simulation results are displayed in the following graph:
a) Use the simulation results to determine the approximate p-value of the test.

b) Interpret this p-value. (In other words, this is the probability of what, assuming what?)

c) Summarize the conclusion that you draw from this study, as related to the research question in the first sentence. Be sure to address issues of statistical significance, causation, and generalizability. Also explain the reasoning process and justification for your conclusions.

6. A chimpanzee named Sarah, who had been raised in captivity since age one, was the subject in a study of whether chimpanzees can solve problems. Sarah was shown 30-second videotapes of a human actor struggling with one of several problems (for example, not able to reach bananas hanging from the ceiling, a record player not playing). Then Sarah was shown two photographs, one that depicted a solution to the problem (like stepping onto a box, plugging in the record player) and one that did not match that scenario. Researchers watched as Sarah selected one of the photos, and they kept track of whether Sarah chose the correct photo depicting a solution to the problem. They found that Sarah chose the correct photo in 7 of 8 scenarios that she was presented.

a) Describe how you could use a coin to conduct a simulation analysis for testing whether Sarah genuinely does tend to choose the correct photo more than would be expected by random chance. Be sure to indicate how many times you would toss the coin and what variable you would keep track of.

The graph below depicts the results of using the Coin Tossing applet to simulate Sarah’s choices 1000 times:

(Notice that the numbers at the top of the graphs give counts of dots at each number. The number above 0 is hard to read: 4 dots are at the value 0.)
b) Use this graph to determine the approximate p-value.

c) Describe what this p-value means.  *[Hint: The p-value is the probability of what, assuming what?]*

d) What would you conclude from this simulation analysis about whether Sarah is able to do better than guessing? Explain the reasoning process behind your conclusion.

7. Suppose that the observational units in a study are patients who visit a particular hospital’s emergency room in a particular week.

a) Provide an example of a **quantitative** variable that could be recorded on these observational units.

b) Provide an example of a **categorical** variable that could be recorded on these observational units.