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>

\[ \text{Difference} = \mu (\text{lengths (died)}) - \mu (\text{lengths (survived)}) \]

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) Create a hypothetical example of 5 quiz scores (integers from 0 to 5, with repeats allowed) with the property that 4 of them (i.e., 80%) are greater than the mean. Also report the value of the mean.

b) Suppose that I calculate the median score on an exam to be 78. Then I decide to add 2 points to every student’s score. How would this change affect the median, if at all? (Be as specific as possible.) Explain briefly.

c) Suppose that I calculate the median score on an exam to be 78. Then I decide to add 2 points to every student’s score. How would this change affect the inter-quartile range, if at all? (Be as specific as possible.) Explain briefly.

d) Which of the following sets of six quiz scores has the smallest standard deviation? (Circle your answer; do not bother to explain or perform any calculations.)

(A) 0, 0, 0, 5, 5, 5  (B) 3, 3, 4, 4, 5, 5  (C) 0, 1, 2, 3, 4, 5

e) Which of these (same) sets of six quiz scores has the largest standard deviation? (Circle your answer; do not bother to explain or perform any calculations.)

(A) 0, 0, 0, 5, 5, 5  (B) 3, 3, 4, 4, 5, 5  (C) 0, 1, 2, 3, 4, 5

7. Students in an introductory statistics class were asked how many states they have visited. The following output pertains to the sample results:
a) Determine a 90% confidence interval for the population mean number of states visited among all students at this university.

b) Check and comment on whether the technical conditions of this confidence interval are satisfied.

c) For what proportion of students in the sample is the number of states visited within the interval from a)?

d) Should you expect your answer to c) to be close to 90%? Explain why or why not.

e) Based on your interval, what can you say about the $p$-value if you were to conduct a two-sided significance test of whether the population mean differs from 10? Explain briefly, without conducting a test or doing new calculations.

8. I once collected data in class on how long (in seconds) it took for a chocolate chip to melt in your mouth and for a peanut butter chip to melt in your mouth. I took the differences in these times (chocolate minus peanut butter) for each person. The sorted data, and a dotplot, for the 31 differences appear below:

```
-41  -36  -35  -33  -31  -28  -25  -25  -20  -20
-17  -17  -16  -14  -11  -7   -6   -5   -5  -4
-4   -2   1   3   6   15  17   21   30   36
```

a) Explain what the value -41 means in terms of the student who produced that value and his/her melting times.

The mean of these 31 differences is -6.65 seconds, and the standard deviation is 23.61 seconds.
b) Conduct a test of whether the sample data provide strong evidence of a difference in melting times of chocolate and peanut butter chips on average. Report the hypotheses, test statistic, and p-value as accurately as you can.

c) Determine and interpret a 95% confidence interval based on the 31 differences.

d) Summarize your conclusion from this analysis.

e) Now suppose that you were to re-do this analysis after removing the outlier value of 67. Indicate how each of the following would change. Circle your answers. Do not bother to explain or perform any calculations.

<table>
<thead>
<tr>
<th>Mean:</th>
<th>Decrease</th>
<th>Increase</th>
<th>Remain the same</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation:</td>
<td>Decrease</td>
<td>Increase</td>
<td>Remain the same</td>
</tr>
<tr>
<td>Test statistic:</td>
<td>Decrease (more negative)</td>
<td>Increase (less negative)</td>
<td>Remain the same</td>
</tr>
<tr>
<td>p-value:</td>
<td>Decrease</td>
<td>Increase</td>
<td>Remain the same</td>
</tr>
</tbody>
</table>