1. Suppose that 15% of all Cal Poly students ride a bicycle to campus.

a) Is the 15% number a parameter or a statistic? Explain briefly.

Suppose that Josiah takes a random sample of 80 Cal Poly students, and Kellen take a random sample of 160 Cal Poly students.

b) Who is more likely to find that more than 20% of his/her sample rides a bicycle to campus? Explain briefly.

c) Suppose that you take a random sample of Cal Poly students, and you want to use the Central Limit Theorem to conduct probability calculations involving the proportion of your sample who ride a bicycle to campus. How large would your sample have to be, in order for using the Central Limit Theorem to be valid?

2. Have you ever pretended to be talking on your cell phone in order to avoid interacting with people around you? A recent survey conducted by the Pew Research Center during April 26 – May 22, 2011 asked cell phone users about this issue. The survey involved selecting a random sample of 1858 American cell phone users, 13% of whom admitted to faking cell phone call in the past 30 days.

a) Check whether the technical conditions for using a one-proportion z-interval (i.e., a confidence interval based on the normal approximation) are satisfied here.

b) Determine a 99% confidence interval.

c) Write a sentence or two interpreting what this confidence interval reveals.

d) Based only on your confidence interval, would you expect to reject the null hypothesis that 10% of all American cell phone users have faked a cell phone call in the past 30 days, using the .01 significance level? Explain briefly.

e) If you were to reduce the confidence level, how (if at all) would the midpoint and width of this confidence interval change? (Circle your answers; do not bother to explain.)

<table>
<thead>
<tr>
<th></th>
<th>Larger</th>
<th>No change</th>
<th>Smaller</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midpoint:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Width:</td>
<td>Wider</td>
<td>No change</td>
<td>Narrower</td>
</tr>
</tbody>
</table>

3. The news website CNN.com regularly posts a poll question that people who view the website can respond to. The following results were posted on January 10, 2012:
a) What is the sample size in this survey?

b) Are the numbers 62%, 25%, and 13% statistics or parameters? (Circle your answer; do not bother to explain.)

statistics  parameters

c) Using these data to produce a 99.9% confidence interval for the population proportion of all employed Americans who surf the Web often while on the job produces the interval (.615, .625). Why is this interval so narrow? (Use no more than 6 words in your answer.)

d) Are you very confident that between 61.5% and 62.5% of all employed Americans surf the Web often while on the job? Explain your answer. (Do not perform any calculations.)

4. A teacher conducted a study to see whether students could distinguish between the tastes of two brands of cola. Each of 30 students was presented with three cups: Two cups contained the same brand of cola, and the third cup contained a different brand of cola. Each student was asked to identify which cup contained the different brand. Let the random variable X represent the number of students who correctly identify the different cup.

a) If students are really not able to distinguish, and so they randomly choose among the three cups, what would be the probability of correctly identifying the different one?

b) It turned out that 16 of 30 students correctly identified the different cup. The p-value turns out to equal .019. Finish the following sentence to interpret this p-value:

The probability is .019 that ____________________________________________

assuming that ________________________________________________.

c) Based on this p-value, would you conclude that the sample data provide fairly strong evidence that students really can identify the different cola more often than by random chance? Also explain the reasoning process behind your answer.

5. 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)?

6. The following Minitab 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.

7. 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

8. Students in an introductory statistics class were asked how many states they have visited. The following output pertains to the sample results:

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SE Mean</th>
<th>StDev</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>states</td>
<td>50</td>
<td>13.36</td>
<td>1.03</td>
<td>7.27</td>
<td>2.00</td>
<td>7.00</td>
<td>12.00</td>
<td>20.00</td>
<td>31.00</td>
</tr>
</tbody>
</table>

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.