1. Some of the statistical inference techniques we have studied include:
   A. One-sample $z$-procedures for a proportion
   B. Two-sample $z$-procedures for comparing proportions
   C. One-sample $t$-procedures for a mean
   D. Two-sample $t$-procedures for comparing means
   E. Paired-sample $t$-procedures

For each of the following questions, identify (by letter) the procedure that you would use to investigate that question. Also indicate (either in symbols or in words) the null and alternative hypothesis to be tested in each case.

a) Do cows tend to produce more milk if their handler speaks to them by name every day than if the handler does not speak to them by name? A farmer randomly assigned half of her cows to each group and then compared how much milk they produced after one month.

   Procedure: Null hypothesis: \[ H_0: \mu_A = \mu_B \] Alternative hypothesis: \[ H_A: \mu_A > \mu_B \]

b) I want to investigate whether a typical full-time Cal Poly student studies for less than 25 hours per week.

   Procedure: Null hypothesis: \[ H_0: \mu < 25 \] Alternative hypothesis: \[ H_A: \mu > 25 \]

c) You want to investigate whether teenagers in the United Kingdom (UK) have read more Harry Potter books, on average, than teenagers in the United States (US). You take a random sample of 500 teenagers in each country and ask how many Harry Potter books they have read.

   Procedure: Null hypothesis: \[ H_0: \mu_{UK} = \mu_{US} \] Alternative hypothesis: \[ H_A: \mu_{UK} > \mu_{US} \]

d) I want to investigate a claim that I read, stating that 70% of new restaurants go out of business within the first year.

   Procedure: Null hypothesis: \[ H_0: p = 0.70 \] Alternative hypothesis: \[ H_A: p < 0.70 \]

2. I’m curious about the number of Facebook friends that Cal Poly students have. I asked one of my students (Emily) to gather some data, so she recorded the number of Facebook friends for 20 Cal Poly students who are majoring in Statistics. Her sample data are displayed in the following dotplot and summarized with the following statistics:

<table>
<thead>
<tr>
<th>Number of Friends</th>
<th>120</th>
<th>240</th>
<th>360</th>
<th>480</th>
<th>600</th>
<th>720</th>
<th>840</th>
<th>960</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>N</td>
<td>Mean</td>
<td>StDev</td>
<td>Minimum</td>
<td>Q1</td>
<td>Median</td>
<td>Q3</td>
<td>Maximum</td>
</tr>
<tr>
<td>Number of Friends</td>
<td>20</td>
<td>391.8</td>
<td>273.8</td>
<td>39.0</td>
<td>176.0</td>
<td>327.0</td>
<td>639.0</td>
<td>947.0</td>
</tr>
</tbody>
</table>
a) (8 pts) Determine a 90% confidence interval for the population mean \( \mu \).

b) (6 pts) Write a sentence (or two) interpreting this interval in context. Be sure to include a clear description of what \( \mu \) represents here.

c) (4 pts) Would you expect 90% of Cal Poly student Facebook users’ values (for number of Facebook friends) to fall within this interval? Explain.

d) (6 pts) State the technical conditions required for this confidence interval procedure to be valid, and comment on whether you consider them to be satisfied in this case.

3. Researchers randomly assigned 14 male volunteers with high blood pressure to one of two diets for four weeks: a fish oil diet and a regular oil diet. The subjects’ diastolic blood pressure was measured at the beginning and end of the study, and the reduction was recorded for each subject (based on a study published in the New England Journal of Medicine). Prior to conducting the study, researchers conjectured that those on the fish oil diet would tend to experience greater reductions in blood pressure than those on the regular oil diet. The results were:

- Fish oil diet: 8 12 10 14 2 0 0
- Regular oil diet: -6 0 1 2 -3 -4 2

a) Identify the explanatory and response variables. Also classify each as being categorical or quantitative.

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

c) Calculate the mean and median blood pressure reductions in each group. Does the “fish oil” group have a higher mean and median reduction than the “regular oil” group? Is this consistent with the researchers’ conjecture?

d) How many ways are there to randomly assign 14 subjects into two groups of 7 each?

e) The following graph was obtained by simulating a randomization test on these data. Shade the region of this graph corresponding to the approximate p-value from the simulation. Also indicate whether the p-value appears to be less than or greater than .05.
f) Interpret what the p-value means in the context of this study.

g) Summarize your conclusions from this study. Include an explanation of the reasoning process behind your conclusion. Be sure to address the issues of causation (i.e., is a cause-and-effect conclusion warranted?) and generalizability (i.e., how broadly can you legitimately generalize your conclusion?), as well as the issue of statistical significance.

4. Researchers are often looking for physiological differences that can help to predict susceptibility to diseases. They collected data on the volume (in cubic centimeters) of the hippocampus region of the brain for 15 pairs of identical twins, one of whom was affected by schizophrenia and one of whom was not. Summary statistics are presented here:

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>unaffected</td>
<td>15</td>
<td>1.7587</td>
<td>0.2424</td>
<td>0.0626</td>
</tr>
<tr>
<td>affected</td>
<td>15</td>
<td>1.5600</td>
<td>0.3013</td>
<td>0.0778</td>
</tr>
<tr>
<td>difference</td>
<td>15</td>
<td>0.1987</td>
<td>0.2383</td>
<td>0.0615</td>
</tr>
</tbody>
</table>

a) Explain why a paired analysis is appropriate here.

b) Report the appropriate hypotheses for a paired \( t \)-test of whether the data suggest that the mean volumes of the hippocampus region differ between those with schizophrenia and those without. Be sure to define the relevant parameter(s) in words as well as symbol(s).

c) Perform the appropriate test. Show how to calculate the test statistic by hand, and also report the p-value. State your test decision at the .05 and .01 significance levels. Finally, summarize your conclusion about the research question.

d) Determine and interpret a 99% confidence interval for the mean difference in hippocampus volumes between the two kinds of twins.

e) Is the confidence interval consistent with the test decision? Explain.

5. Suppose that you record body temperatures (in degrees Fahrenheit) for a random sample of 130 healthy adults. Also suppose that the population mean body temperature is 98.6 and that the population standard deviation of body temperatures is 0.7. Let \( \bar{X} \) represent the sample mean of the 130 body temperature measurements.

a) Describe and produce a well-labeled sketch of the (approximate) sampling distribution of \( \bar{X} \).

b) Determine the value of \( k \) such that \( \Pr(98.6 - k < \bar{X} < 98.6 + k) = .95 \). Show how you determine the value \( k \).

c) What would happen to the value of \( k \) in part c) if the sample size were larger?

d) Determine the sample size needed for \( \Pr(98.58 < \bar{X} < 98.62) = .9997 \).